GENOTYPIC AND PHENOTYPIC CHARACTERIZATION OF PURDUE SOFT RED WINTER WHEAT BREEDING POPULATION

by

Rupesh Gaire

A Dissertation

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Doctor of Philosophy

Department of Agronomy

West Lafayette, Indiana

May 2020

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Mohsen Mohammadi, Chair

Department of Agronomy

Dr. Mitchell Tunistra

Department of Agronomy

Dr. Steven Scofield

Department of Agronomy

Dr. Gina Brown-Guedira

Department of Crop and Soil Science, North Carolina State University

Approved by:

Dr. Ronald Turco

To my late grandfather Mr. Nara Bhupal Gaire who always encouraged me.

Thank you for your love, support, and blessings.

ACKNOWLEDGMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Comprehensive information of breeding germplasm is a necessity to develop effective strategies for accelerated breeding. I characterized Purdue University soft red winter wheat breeding population that was subject of intensive germplasm introduction and introgression from exotic germplasm. Using genotyping-by-sequences (GBS) approach, I developed ~15,000 single nucleotide polymorphisms (SNPs) and studied extent of linkage disequilibrium (LD) and hidden population structure in the population. The extent of LD and its decay varied among chromosomes with chromosomes 2B and 7D showing the most extended islands of high-LD and slow rates of decay. Four sub-populations, two with North American origin and two with Australian and Chinese origins, were identified. Genome-wide scans for signatures of selection using F_{ST} and hapFLK identified 13 genomic regions under selection, of which six loci (*LT*, *Ppd-B1*, *Fr-A2*, *Vrn-A1*, *Vrn-B1, Vrn3*) were associated with environmental adaptation and two loci were associated with disease resistance genes (*Sr36* and *Fhb1*).

The population was evaluated for agronomic performance in field conditions across two years in two locations. Genome-wide association studies identified major loci controlling yield and yield related traits. For days to heading and plant height, large effects loci were identified on chromosome 6A and 7B. For test weight, number of spikes per square meter, and number of kernels per square meter, large effect loci were identified on chromosomes 1A, 4B, and 5A, respectively. However, for grain yield *per se*, no major loci were detected. A combination of selection for other large effect loci for yield components and genomic prediction could be a promising approach for yield improvement.

In addition, the population was evaluated for FHB resistance under misted FHB nurseries inoculated with scabby corn across 2017-18 (Y1) and 2018-19 (Y2) seasons at Purdue Agronomy Farm, West Lafayette, in randomized incomplete block designs. Phenotypic data included disease incidence (INC), disease severity (SEV), *Fusarium* damaged kernels (FDK), FHB index (FHBdx), and deoxynivalenol concentration (DON). Twenty-five loci were identified at *-logP* ≥ 4.0 to be associated with five FHB-related traits. Of these 25, eighteen explained more than 1% of the

phenotypic variations. A major QTL on chromosome 2B i.e., $Q_{2B.1}$ that explained 36% of variation in FDK was also associated with INC, FHBdx, and DON. The marker-trait associations that explained more than 5% phenotypic variation were identified on chromosomes 1A, 2B, 3B, 5A, 7A, 7B, and 7D. To investigate the applicability of other QTL with less signal intensity, the threshold criterion was lowered to $-logP \geq 3.0$, which resulted in the identification of 67 unique regions for all traits. This study showed that the FHB-related traits have significant correlations with the number of favorable alleles at these loci, suggesting their utility in improving FHB resistance in this population by marker-assisted selection. The genotype and phenotype data produced in this study will be valuable to train genomic prediction models and study the optimal design of genomic selection training sets. This study laid foundation for the design and breeding decisions to increase the efficiency of pyramiding strategies and achieving transgressive segregation for economically important traits such as yield and FHB resistance.

CHAPTER 1.   LITERATURE REVIEW

1.1   Origin and domestication

Common hexaploid wheat (*Triticum aestivum*) belongs to the *Triticeae* tribe in the grass family *Poaceae* (Feldman and Levy, 2015). The *Triticeae* tribe diverged from rest of the *Poaceae* tribe around 25 million years ago and gave rise to domesticated barley, rye, and wheat which consists of many diploid and polyploid species. The basic haploid chromosome number in wheat is x = 7 (Huang et al., 2002). The bread wheat is an allohexaploid (2n=6x =42) and chromosomes pair in a diploid-like fashion. Hexaploid wheat evolved around ~10,000 years ago in the Fertile Crescent of Mesopotamia through two separate hybridization events followed by chromosome doubling between three different diploid (2n=2x = 14) progenitor species (reviewd by Salamini et al., 2002 and Matsuoka 2011). Evidence based on morphological, cytological and genetic studies suggests that, around 0.5 million years ago, first interspecific hybridization and chromosome doubling occurred between wild diploid wheat *T. urartu* (AA: 2n =2x =14) and another species that belonged to the lineage of the current wild wheat species, *Aegilops speltoides* Tausch (SS: 2n=2x=14) (Miki et al., 2019) that gave rise to wild tetraploid wheat *T. turgidum* (AABB: 2n=4x= 28) (Dvorak et al., 1988). The interspecific hybridization of domesticated *T. turgidum* (AABB) with wild, diploid wheat *Aegilops tauschii* (DD) followed by spontaneous chromosome doubling gave rise to hexapliod *T. aestivum* (AABBDD).

Crop domestication refers to the genetic alterations in wild plant species caused by inadvertent human selection that lead to new crop types to meet human need. Einkorn (*T. monococcum*, 2n = 14, AA), a diploid wheat, was the first wheat to be cultivated, and it was domesticated from its wild progenitor *T. boeoticum* (Salamini et al., 2002). Domesticated Einkorn had larger seeds in comparison to wild Einkorn. In addition, domesticated Einkorn did not have brittle rachis that break spontaneously at maturity (Matusoka 2011; Salamini 2002). The presence of non-brittle rachis allowed domesticated forms to hold seeds together at maturity making it amenable to harvesting and threshing. Therefore, non-brittle rachis is considered as one of the first and most essential domestication traits acquired by cultivated wheat (Peng et al., 2002; Faris 2014).

Another important event in the evolution of cultivated common wheat was the domestication of tetraploid wheat commonly known as emmer wheat – progenitor of present-day durum wheat (*T. turgidum subsp. durum*) (Matusoka 2011; Salamini 2002). The two important domestication syndrome traits associated with emmer wheat included inability to scatter seeds (non-brittle rachis) and free-threshing character. In domesticated tetraploid wheat, the presence of recessive alleles of two genes *Br-A1* and *Br-B1* located on short arm of chromosome 3A and 3B governs the non-brittleness (Wantanabe et al., 2002). The domestication of free-threshing tetraploid wheat is considered the final step in evolution of present-day polyploid wheat cultivars (Salamini 2002). Early wild and domesticated emmer wheat consisted of tough glumes and hulled seeds which were difficult to thresh. The free-threshing phenotype is a quantitative trait and seems to be controlled by at least four quantitative trait loci (QTL) (Simonneti et al., 1999). Evidence showed that two of the major QTL, *Tg* (tenacious glume) (on 2B) and the *Q* locus (on 5A), accounted for ~25% variation in a segregating population derived from a cross between durum and wild emmer wheat (Simonneti et al., 1999). Epistatic interaction between the *Tg* and *Q* loci control free-threshing phenotype. Individuals with the dominant allele at *Q* locus (*QQ*) and recessive allele at *Tg* locus (*tgtg*) are free threshing whereas *QQTgTg* phenotypes are non-free threshing (Simonneti et al., 1999: Faris 2014). In common hexaploid wheat, *Tg* gene on chromosome 2D derived from *Ae. tauschii* has the major effect on the expression of the free-threshing phenotype (Jantasuriyarat et al. 2004).

1.2   Wheat in the United States

Wheat occupies almost one-fifth of the world's cultivated land and is a major staple food for almost two-thirds of world population (~ 2.5 billion people from 89 different countries, Figure 1.1; FAO 2020). In 2019, 731.6 terragram (Tg) of wheat was produced worldwide on 215 million hectares of land accounting for 27.5% of total cereal production (FAO, 2020). Wheat is grown in every habitable continent and nearly 50 billion dollars-worth of wheat is traded globally each year. China, India, and Russia are the top three producers of wheat producing 134.3, 98.5, and 85.6 Tg in 2017 (FAO, 2020).
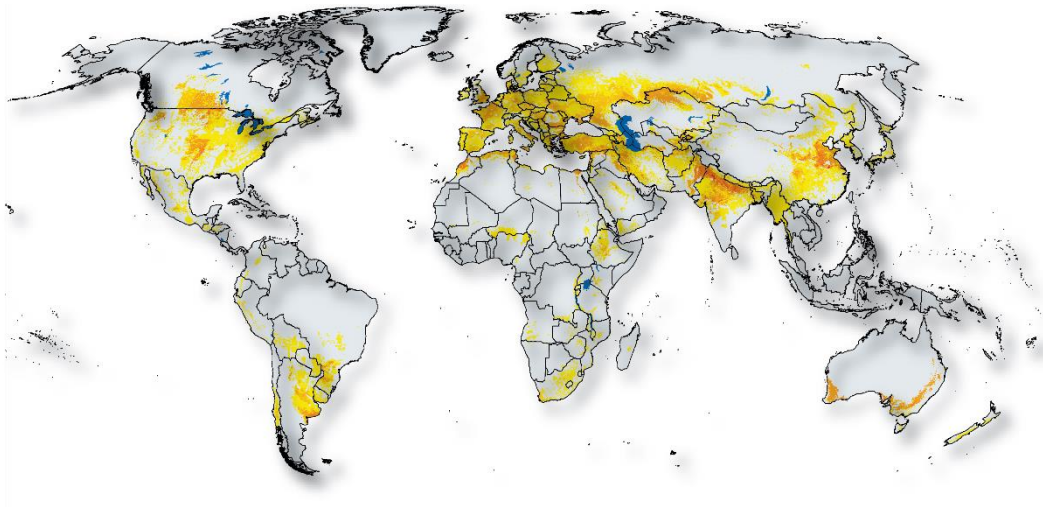
Figure 1.1. Wheat growing regions in the world. The picture was downloaded from https://wheat.org/wheat-in-the-world/ (CGIAR, 2020)

1.2.1 History of wheat introduction and improvement in the US

The history of wheat introduction and improvement in America is well described by Carleton R. Ball in "The History of American Wheat Improvement" that was published in 1930. Wheat was introduced in the United States from Spain in the 16th century in the pre-colonial period. Wheat was first introduced to West Indies in Caribbean and then promptly introduced to Mexico and South America. From there, wheat adapted to hot and dry climates was introduced to California. On the North, during the early colonial period in the 17th century, the European colonist who migrated to New England introduced wheat from several countries including England, France Sweden, and Netherlands. The various kinds of wheat brought by the colonists provided a foundation of genetic variation and adaption of wheat in the US. The cultivation of wheat increased steadily for a century following the introduction.

During the 19th century, the territory of the wheat expanded enormously with the adoption of 'modern' communication (i.e. newspaper), improvement of seeding and tilling machineries, development of reaping machinery for harvest, and better milling machineries and processes. During the first half of the 19th century, wheat cultivation was primarily in the Eastern states. The most commonly grown wheat was soft white wheat because of high demand for white flour and lack of milling methods that separated red bran from the flour. Superior cultivars were introduced from central Europe, Mediterranean, and Australia for adaptability, disease resistance, and

agronomic performance. In the second half of the 19th century, around the 1870s, when mills with purifiers could satisfactorily separate the bran from the flour, hard red spring and winter wheat that were adapted to dry conditions were introduced in the Upper and Central Mississippi, respectively, for commercial production.

In the 1862, the Morrill Act created the land grant system of agricultural colleges which later intensified research on the genetic improvement of wheat. In 1880, the first project for testing wheat varieties was approved by Purdue University's Board of Trustees. During the early 20th century, wheat was improved through introduction, hybridization, and selection of pure lines. In 1914, Ball and Clark assembled, described, and classified all of the commercial varieties based on the growth habit and their commercial uses into soft red winter wheat, soft white wheat, hard red winter wheat, and hard red spring wheat (Figure 1.2). In 1915, Martin L. Fisher was first appointed as the Purdue Agriculture Experiment Station's plant breeder (Ohm and McFee 2012). Throughout the 19th century, improvement in wheat was achieved through generation of genetic variation through hybridization (or crossing) and selection of lines based on visual performance.

Significant genetic gains in wheat was achieved through conventional breeding approaches in the United States. A positive linear yield increase has been achieved in wheat from 1930-2012 with the highest annual gain of 280 Kgha^{-1} (Xu et al., 2017). In uniform nurseries from 1958 to 1980, an annual rate of gain for yield of about 0.74% has been reported Schmidt (1984). While conventional breeding has been effective, yet, the rate of genetic gains through visual observation was low especially for highly quantitative traits such as yield and FHB resistance. For quantitative traits, allele segregation and recombination of many genes with small effects make selection ineffective. Breeders prefer to delay the selection process to later generations when the homozygosity is higher and thus, prolonging the breeding cycle and reducing the amount of genetic gains. The use of molecular markers that are linked with the genes or loci governing the traits facilitates effective selection especially in the early generations and thus can reduce time required for completing a breeding cycle. The use of DNA-based markers in wheat started around late 1990 (Chao et al., 1989) and has been continued to this date.

1.2.2 Market classes in the US

Currently, USDA classifies wheat into six different classes based on the endosperm color, growth and texture of the grain including durum, hard red winter (HRW), hard red spring (HRS), hard white wheat, soft red winter (SRW), and soft white wheat (USDA-ERS, 2019; Figure 1.2). The HRW wheat is predominately grown east of the Rocky Mountains and west of the Mississippi River. Hard red spring (HRS) and durum wheat varieties are concentrated mainly in South Dakota, North Dakota, Western Minnesota, and Montana (USDA-ERS, 2019; Figure 1.2). Hard red wheat is mainly used in hearth breads, hard rolls, croissants, bagels, and pizza crust. Durum wheat, being very hard, is suitable for semolina, premium pasta products, couscous, and Mediterranean breads. Among the white wheat, soft white wheat dominates the area of cultivation where it is mainly grown in the Pacific northwest region and is used in making brighter and whiter products of noodles, pastries, and cakes. The SRW wheat (SRWW), which is targeted in this PhD dissertation, is mainly grown in the states along the Mississippi River and the eastern states including Alabama, Arkansas, Georgia, Illinois, Indiana, Kentucky, Maryland, Michigan, Missouri, North Carolina, New York, Ohio, Pennsylvania, Tennessee, Virginia, and Wisconsin (USDA-ERS, 2020; Figure 1.2). Soft red wheat is used to produce a wide range of confectionary products such as cookies, crackers, and cakes.
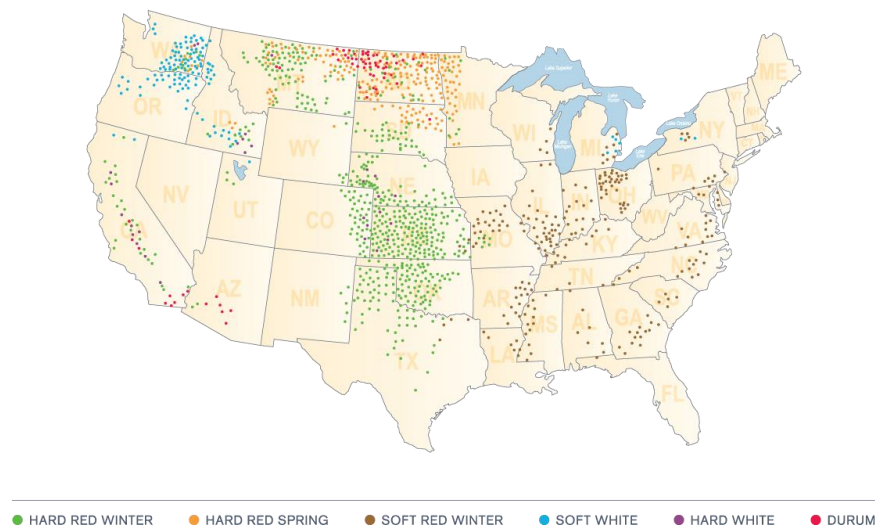
Figure 1.2. Different market classes and their growing regions in the United States. The picture was downloaded from https://www.uswheat.org/working-with-buyers/wheat-classes/.

1.2.3 The genetic mechanisms underlying different market classes

The genetics of growth habit in wheat is rather complex owing to the multi-genic nature of trait and the hexaploid nature of wheat genome. Based on the growth habit, wheat can be classified as spring or winter wheat. Spring wheat is planted after winter has passed whereas winter wheat is planted in the fall before the winter sets in. Winter wheat requires exposure to chilling temperatures. This requirement which is called vernalization is needed for winter wheat to transit from vegetative to reproductive growth phase while spring wheat lacks vernalization requirement (Law and Worland, 1997). In winter wheat, vernalization requirement can protect the vital organs from winter damage since flowering is delayed until spring (Law and Worland, 1997). Vernalization is regulated by three groups of genes: *Vrn1*, *Vrn2*, and *Vrn3* (Guedira et al., 2016). *Vrn1* gene is located in long arm of group 5 homeologous chromosomes and is orthologous to the *Arabidopsis* meristem identity gene *APETALA1* (Yan et al., 2003). The expression of *Vrn1* gene is induced by cold temperatures. Similarly, *Vrn2* (*ZCCT2*) is located on chromosome 5 homeologous groups that consists of a putative zinc finger in the first exon and the CCT domain in the second exon. (Yan etal., 2004a). The *Vrn2* gene represses flowering and is downregulated by vernalization and short days. *Vrn3* gene was mapped on chromosome 7B and is orthologous to Arabidopsis *FLOWRING LOCUS T* (*FT*) gene (Yan et al., 2006). The *FT* gene is a positive regulator of flowering. During the early winter, when the vernalization requirement has not been reached, *Vrn2* gene suppresses flowering. As the winter progresses and the days get shorter, *Vrn1* is induced and the expression of *Vrn2* is suppressed. In the spring, as the length of day increases, *Vrn3* is upregulated which promotes flowering (Yan et al., 2006). Mutations in *Vrn1* gene had been reported to confer spring growth habit and the alleles of *Vrn1* conferring spring growth habit are dominant to the wild-type winter-type allele (Yan et al., 2003; Yan et al., 2004b).

The genetics of endosperm texture in wheat is much simpler than that of growth habit. Wheat is classified as hard or soft based on the texture of the endosperm and has pivotal role in determining end-use quality. *Hardness* locus (*Ha*) locus was identified in the short arm of the chromosome 5D (Mattern et al 1973, Law et al 1978) and later found to be associated with friabilin protein, a 15kD protein found abundant in endosperm of soft wheat, in small amounts in endosperm of hard wheat, and absent in hard endosperm of durum wheat (Greenwell and Schofield 1986). The friabilin protein is composed of 3 main components: Puroindoline A (PINA),

16

Puroindoline B (PINB), encoded by *Pina* and *Pinb* genes, and the grain softness protein (GSP) (Sourdille et al 1996, Giroux and Morris 1998). Mutations that disrupt the function of *Pin-a* and *Pin-b* genes convert soft wheat grains to hard (Morris 2002). Giroux and Morris (1998) identified a single nucleotide polymorphism (SNP) in *Pinb* substituting the codon for Gly-46 to Ser-46, leading to the loss of function PINB and expression of hard endosperm phenotype. Similarly, a null allele in *pina* also led to hardness in wheat (Giroux and Morris 1998). Later, several other mutations in Pin-b were identified that led to hard endosperm in wheat (Lillemo and Morris, 2000; Morris et al., 2001;). Durum wheat, which lacks the D genome, also lacks all the functional copies of *Pina* and *Pinb* genes and therefore expresses a very hard endosperm texture (Bhave and Morris 2008). The genes on other homeologous groups 5A and 5B have not shown deterministic effects on endosperm texture.

There are three copies of *R* genes for each sub-genome of wheat located on the long arm of homeologous group 3 chromosomes (Metzger and Silbaugh, 1970) that together control the red coat color in wheat. Red pigment in grain coat tissue are derivatives of catechin-tannin, probably, phlobaphene or proanthocyanidin synthesized through flavonoid biosynthesis pathway. In white wheat, the expression of genes involved in flavonoid biosynthesis was shown to be suppressed, and the pigment conferring red color is absent in the developing grain (Himi and Noda 2005). In red wheat, the *R* loci encode a *myb-type* transcription factor that activates the flavonoid biosynthesis genes and produces red color (Himi and Noda 2005).

1.2.4 Economic significance of wheat in the US

The United States ranked fourth and produced 47.4 Tg of wheat on 16.03 million hectares producing a revenue of 9.7 billion dollars in the US (USDA-NASS, 2019). Wheat ranked third after corn and soybean for area planted and revenue in the U.S (USDA-NASS, 2019). Of the total supply of wheat in 2018 (production + 2017 ending stocks), 35.3% were used for domestic consumption, 30% were exported, and the rest of wheat were included as ending stocks for 2018.

From different market classes, HRW wheat is the most produced wheat class in the U.S followed by HRS wheat and SRWW. In 2018, a total of approximately 18 and 16 Tg of HRW and HRS wheat were produced on 9.1 and 5.3 million hectares of land, respectively. Similarly, 7.4 and

1.4 million Tg of white and durum wheat were produced in 4.7 and 0.8 million hectares of land. The United States produced 7.8 Tg of SRWW on 2.3 million ha (5.6 million acres) of land in 2019 (USDA-NASS, 2020). About 54% of the total production of SRWW (4.2 Tg) is used for domestic consumption as food, seed, and feed (USDA-ERS, 2020), whereas ∼41% (23.2 Tg) of the US SRWW was exported as grain, flour, and processed products. In 2019, the state of Indiana produced 0.44 million tons of SRWW planted in 133,546 hectares that was equivalent to 79.8 million dollars. In the last 10 years, on average, SRWW provided an estimated annual economic value of over US$2.2 billion (USDA-NASS, 2020).

Despite the significant economic contribution, wheat acreages in the United States has significantly declined over the last three decades from ~28 million hectares in 1990 to ~15 million hectares in 2017. Similar trend was observed in SRWW where 14.8 million tons of SRWW was produced in 1990 compared to 4.38 million tons in 2019. Indiana produced 1.4 million tons on 438,713 hectares land in 1990. This figure dropped to a total of 0.44 million tons of that was produced on 133,546 hectares in 2019. The rapid decline is due to the policy changes related to the support of agricultural commodities and biofuel production, introduction and adaptation of genetically modified (GM) corn and soybean, and epidemic outbreaks of fusarium head blight (FHB) disease which caused reduced wheat yield and market prices (McMullen et al., 2012).

1.3 The need for rapid genetic gains in wheat

Wheat is one of the most important staple crops for low and middle-income nations in North Africa and West and Central Asia. It is the primary source of protein and provides up to half of calories consumed (CGIAR, 2020). Globally, wheat yield is increasing in 61.2% of the cultivated areas while in 38.8% of the areas the yield has stagnated (Ray et al., 2012). Although, conventional breeding methodologies successfully exploited genetic variation in crops and led to 0.8–1.2% annual genetic gains in five major food crops including wheat (Ray et al., 2013), the current growth rate of cereal production is not enough. By the year 2050, global population is expected to reach 9.1 billion people adding 1.4 billion people to current population (FAO, 2020). To meet the growing demands of food, annual cereal production should increase by 2.4% adding almost one billion tons of wheat to total production each year (Ray et al., 2013). Climate change brings uncertainty in future crop production with an estimated decrease in wheat production over the next

decade by 6% for every increase of 1°C on global temperature (Asseng et al., 2015). To ensure food security in the world, significant progress is needed to produce more food per unit area. Plant breeders must leverage and integrate recent advancement in next-generation sequencing and phenotyping technologies to accelerate genetic gains and develop high-yielding, climate-resilient wheat varieties.

1.4   Past and present of genetic mapping in wheat

Genetic mapping is a helpful tool to identify genomic regions that control traits of interest. Traditionally, mapping has been conducted by using bi-parental mapping populations but recently a newer method called genome-wide association studies has become popular. Bi-parental QTL mapping follows four basic steps. First step is crossing two parental lines that show contrasting phenotypes for the trait of interest and creating a segregating population. The second is genotyping the population with polymorphic markers between the two parents. The next step is to create the linkage maps based on the recombination frequency observed between the polymorphic markers in the segregating population. The last step is to perform statistical analyses to identify markers that show significant differences for trait of interest between allelic groups at the marker loci (Collard et al., 2005). In rare cases, the identified QTL are validated through introgression to different genetic backgrounds, and when validated, the markers tightly linked to the QTL are used for marker-assisted selection (Collard and Mackill, 2008).

QTL mapping in wheat can be traced back to 1989 when Chao et al., (1989) reported the use of Restriction Fragment Length Polymorphism (RFLP) for constructing genetic map of 42 markers on group 7 homeologous chromosome of wheat. RFLP markers were really time consuming and labor-intensive nature (Collard et al., 2005) and resulted in low-resolution genetic maps (Gupta et al., 1999). Nevertheless, QTL for grain protein content, pre-harvest sprouting tolerance, vernalization response, kernel hardness, and dwarfing genes (reviewed by Gupta et al., 1999) were identified using RFLP markers.

The availability of polymerase chain reaction (PCR) based markers such as randomly amplified polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP), and simple sequence repeats (SSR) has created a momentum in QTL mapping in wheat (reviewed by Gupta et al., 1999). In particular, SSRs, DNA sequences consisting of 1-6 bp tandem repeats,

became the most commonly used marker systems due to its codominant nature, ubiquitous presence in eukaryotic genome, and transferability between populations (Röder et al., 1998). Many QTL mapping studies in wheat have been performed using SSR markers for important traits. For example, these traits include yield and yield-related traits (Kumar et al., 2007; Kirigwi et al., 2007) and disease and insect resistance (Anderson et al., 2001; Tsilo et al., 2009, Mago et al., 2005).

Identification of markers that are tightly linked with QTL governing trait of interest opened the avenue to perform marker-assisted selection that enabled plant breeders to make rapid genetic gains. There are several reviews that have described established markers available to perform MAS in wheat (https://maswheat.ucdavis.edu/, Miednar and Korzun 2012, Vagndorf et al., 2018). Similarly, the established markers for major loci have been integrated in the USDA-ARS Small Grains Genotyping Laboratories that enables wheat breeder in the US to genotype their breeding lines for important agronomic and disease resistance traits such as vernalization requirements, photoperiod sensitivity, plant height, Barley Yellow Dwarf Virus resistance, FHB resistance, and stem rust resistance.

Although significant progress has been made in the past using bi-parental mapping approaches in wheat, there are several limitations with its application. First, it cannot capture the allelic diversity available in the germplasm collection, rather only capturing allelic diversity available in parents used for population development (Borevitz and Nordborg 2003). Second, the mapping resolution is low due to the limited amount of recombination after the initial crosses.

Genome-wide association studies (GWAS) overcomes both problems prevalent in bi-parental mapping approaches. The strength of association between markers and functional variation governing the trait rests on the extent of the linkage disequilibrium (LD) between them (Risch and Merikangas 1996). As GWAS can take advantage of genetically diverse populations, it eliminates the necessity of developing time and resource consuming artificial mapping populations. In addition, confidence intervals for significant loci narrows with greater probability of pinpointing shorter list of candidate genes due to greater resolution through reduced haplotype size (Cockram et al 2010, Weng et al 2011, and Ehrenreich et al 2009). The tradeoff of GWAS however is the inflated risk of a false association between non-causal markers and a trait. When there is population structure, there is possibility that two individuals in the same sub-population

share both causal and non-causal alleles, and LD between these sites can lead to false associations (Korte et al 2013).

Breseghello and Sorrels (2006) reported the first association mapping study in wheat for kernel size and milling quality. After Breseghello and Sorrels (2006), several studies reported association analysis in different populations of wheat for different traits. For example, Roy et al., (2006) used 519 markers (SSR and AFLP) for association studies of 14 agronomic traits in a population of 55 individuals. Similarly, Crossa et al., (2007) performed association studies using 1148 (DaRT, SSR, AFLP, and RFLP) for rust resistance and yield in population of 170 CIMMYT Elite spring wheat lines. These early studies, however, were limited by small sample sizes and low marker density.

The advancement in next generation sequencing technologies enabled genotyping across the whole wheat genome. However, high-throughput sequencing in wheat was still challenging due to the large and complex genome of wheat, sequence homologies in the homeologus chromosome, and the presence of paralogus genes (Cavannagh et al., 2013; Wang et al., 2014). Elshire et al., (2011) developed a genotyping-by-sequencing (GBS) method in maize. They chopped the genome small fragments by the use of a restriction enzyme and only those fragments with restriction enzyme overhangs were amplified and sequenced. Poland et al., (2012) modified the original GBS approach by using two methylation-sensitive restriction enzymes- one rare cutter and one frequent cutter, followed by amplification and sequencing of only those reads that have overhangs of both of the enzymes. In the absence of reference genome of wheat, de novo marker discovery and genetic map construction were used in the early studies (Poland et al., 2012).

Despite the challenges posed by large (17.2 Gbp) and complex nature of wheat genome, a chromosome-based draft genome of hexplaoid wheat became available in 2014 by the International Wheat Genome Sequencing Consortium. To avoid the complexity, each arm of wheat chromosomes except 3B (whole chromosome was sequenced together) were sequenced separately and aligned de novo. However, the genome coverage was only around 61% (IWGSC 2014). This genomic resource provided a platform to align short read sequences generated through GBS method and facilitated identification of genome wide SNP markers. Four years later, the complete wheat genome assembly (IWGSC RefSeq v1.0) with 94% coverage and 107,891 high confidence

genes became publicly available (Appels et al., 2018). This resource facilitated GBS marker discovery for GWAS analysis and characterizing the physical location of previously known QTL. In addition, it is possible to search the candidate genes in the vicinity of the marker-trait associations (Gaire et al., 2019).

GWAS is conducted on diversity panels, population of landraces, and elite breeding populations. The incentive behind the use of diversity panel or population of landraces is that they are not resulting from selection pressure and therefore, identification of large-effect QTL is more likely. In addition, the higher resolution of populations enables pinpointing candidate genes with more confidence. In contrast, in elite population, large-effect QTL will have likely already become fixed and thus is limited to identification of small effects QTL (Salvi and Tuberosa, 2015). Yet, GWAS in elite populations is more relevance to the process of cultivar development (Spindel et al., 2015) because it enable breeders to understand the genetic architecture of traits, identify QTL present in breeding populations, and estimate frequency of desirable alleles in the population (Hoffstetter et al., 2016). The major caveat of using elite and breeding population is that extended LD blocks exists within a chromosome due to selection and relatedness between individuals that makes identifying candidate genes difficult.

1.5   Evolution of mixed model GWAS methodologies

Several statistical methods have been developed to control the false association in GWAS (Figure 1.3). Yu et al., (2006) presented a Q + K mixed linear model (MLM) to control the confounding effect of population structure (Q) and relatedness (K). The major issue associated with the original Q + K model was requirement of large computational resources to estimate variance components of the population in every iteration. Kang et al., (2008) proposed a computationally superior method, efficient mixed-model association (EMMA), that simplified the relationship matrix compared to method described by Yu et al (2006) and was able to control false-positives. However, the computational efficiency was still lower in the studies with thousands of individuals due to requirement of estimating large matrices in every iteration while estimating variance components. Kang et al., (2010) developed EMMA eXpedited (EMMAX) that markedly reduced the computational cost compared to the original EMMA by avoiding the repetitive variance component estimation procedure for every single marker. Later, Zhang et al., (2010)

reported a method that reduced the computational cost in two ways. First, the individuals in the population were clustered into groups and summary statistics of the kinship between and within groups was used as the kinship matrix. Second, association tests were performed in two steps: first, the MLM is optimized to estimated population parameter such as additive genetic variance and residual variance without the marker effects, second, using the population parameter determined in the first step, the non-population parameters, including the marker effect and the random genetic effect, are estimated for each marker. This method markedly reduced the computational time required to perform GWAS with thousands of markers and individuals while improving the statistical power compared to earlier MLM.
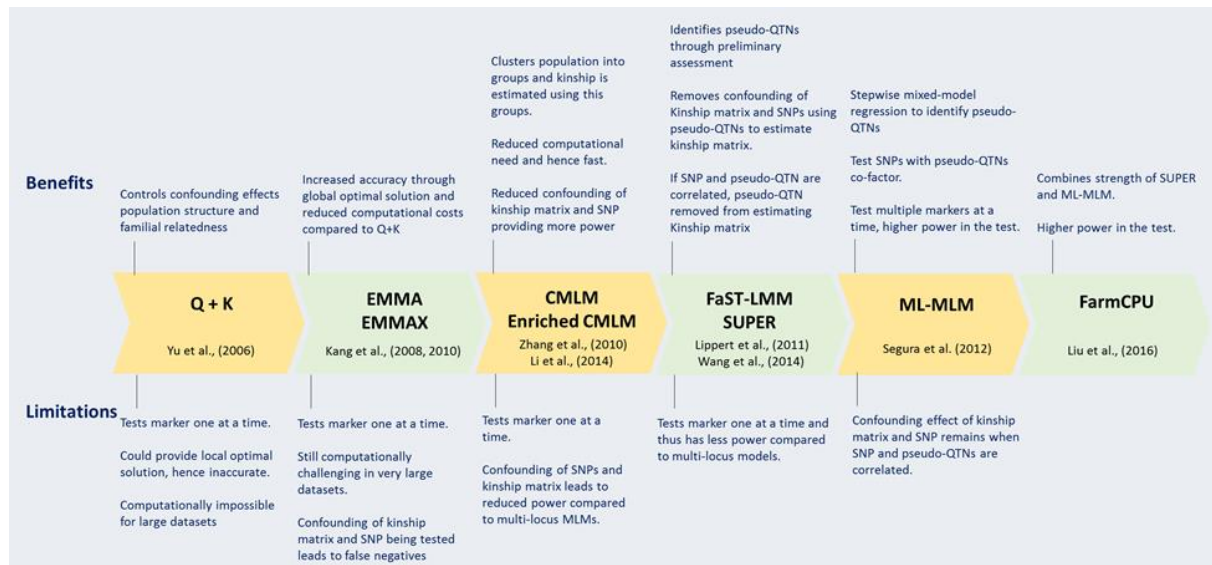
Figure 1.3 Evolution of mixed model methodologies with benefits and limitations

Although MLM can control the false positives in the test, true association could not be detected due to the confounding nature of the SNP being tested and the kinship matrix used to account for relatedness. To improve the power of the GWAS, FaST-LMM (Factored Spectrally Transformed Linear Mixed Model) (Lippert et al., 2011) and SUPER (Settlement of MLM Under Progressively Exclusive Relationship) (Wang et al., 2014b) further simplified the definition of kinship matrix. At first, a preliminary test for each marker is performed to obtain a set of markers that are influential to the trait, known as pseudo quantitative trait nucleotides (pseudo-QTN). In the second step, the identified pseudo-QTN are used to derive a kinship matrix. Furthermore, whenever a pseudo-QTN is correlated with SNP under the test of

significance, it is removed from estimating the kinship matrix, thus eliminating the confounding effects. In the FaST-LMM-Select method, a pseudo QTN is considered correlated if it is within a 2Mb interval on either side of the testing marker whereas in SUPER a threshold on Linkage Disequilibrium (LD) between the pseudo QTN and the testing marker is applied. Both of the methodologies test markers one at a time.

Segura et al. (2012) presented another methodology, MLMM (Multi-Locus Mixed-Model) that tests multiple markers simultaneously and is more powerful than the methods that test a single marker at a time. In this method, stepwise mixed-model regression is performed with forward inclusion and backward elimination. Initially, genetic and residual variances and p-value for each marker are estimated using a mixed model framework. For forward inclusion, the most significant marker (pseudo QTN) is then added as cofactor in the model and the genetic and residual variances are re-estimated. Using the re-estimated variances and pseudo QTN as cofactor, F-test is performed to obtain p-value for the rest of the markers. Again, the most significant SNP is added as co-factor and the variances are re-estimated. The forward inclusion is performed until all the genetic variance is used up. The backward elimination is performed by removing the least significant marker form the model. The forward inclusion and backward elimination are performed until pseudo QTN have converged. The test of significance is performed for each pseudo QTN using the rest as cofactor in the model. For the rest of the markers, all the pseudo QTN are used as cofactors. Although MLMM is powerful compared to mixed model that tests single marker at a time, it does not eliminate the confounding effects between testing markers and kinship that leads to false negatives, in other words, reduces power.

Liu et al., (2016) developed FarmCPU (Fixed and random model Circulating Probability Unification) that combines the strength of MLMM and SUPER to remove the confounding effects between testing markers and kinship and increased the power. In FarmCPU, MLMM is divided into two separate parts: fixed effect model (FEM) and a random effect model (REM) and used them iteratively. The REM is optimized to identify pseudo QTN using combination of p-value and LD between the pseudo QTN. At first, whole genome is divided into user defined bins and the most significant markers are identified for each bin based on p-value. Then, pairwise LD (r^2) between the pseudo QTN are estimated. If the pseudo QTN are correlated above a user-defined threshold, the most significant SNP is used. Using the pseudo QTN identified as covariates, FEM

test each SNP using them as fixed effects. Since, all pseudo QTN are being tested while testing non-pseudo QTN markers, the most significant p-value from these tests is used for each pseudo QTN. Using flowering time as a phenotype, Liu et al., (2016) compared the performance of seven different models including t-test, generalized linear model (GLM), MLM, CMLM, FAST-LMM, MLMM, and FarmCPU using a dataset consisting of 199 *Arabidopsis thaliana* individuals and 250,000 SNPs. T-test and GLM failed to control false positives in the association study. While FAST-LMM, MLM, and CMLM controlled the false positives well, they failed to identify any significant SNPs at threshold of 1% Bonferroni correction. Meanwhile, both MLMM and FarmCPU controlled false positives while successfully identifying significant QTN at the threshold. Furthermore, FarmCPU identified novel region associated with genes involved in controlling flowering in Arabidopsis. These results suggest that FarmCPU has the potential to identify novel genomic regions governing complex traits such as yield and FHB resistance that are governed by many small effects QTL.

1.6 Theoretical framework and literature for improving grain yield and FHB resistance

Grain yield (YLD) is a quantitative trait influenced by genotype, environment, and management. YLD is a cumulative effect of spike number per unit area (NS), number of kernels per unit area (NK), test weight (TW), and thousand kernel weight (TKW) (Slafer et al., 1996). Due to quantitative nature of inheritance, it is very difficult to find stable major QTL that holds true in different environment and genetic backgrounds. In addition to components of yield, QTL that govern the agronomic features such as plant height and days to heading pleiotropically affect yield. Plant height (PH) is one of the most important traits in wheat that played a major role in the Green Revolution by reducing lodging and increasing yield globally (Peng et al., 1999). Three genes-*Rht-B1*, *Rht-D1*, and *Rht8* located on chromosomes 4B, 4D, and 2D are the most widely used genes across the world. *Rht-B1* and *Rht-D1* are the major genes conferring semi-dwarfism in the Green Revolution, and *Rht8* do not seem to have a negative effect on yield.

Days to heading is an important trait. Early flowering and maturity facilitate early harvest of wheat enabling early planting of soybean in double crop systems. Earlier planting for double crop soybeans can translate into a 1-bushel gain in yield per day. In common wheat, vernalization and photoperiod are the major regulators of flowering. The genetic control of flowering through

vernalization is well described in section 1.2.1. Briefly, *Vrn1*, *Vrn2*, and *Vrn3* epistatically interact with each other to regulate flowering in winter wheat. Mutation disrupting *Vrn1* gene leads to spring type growth. In addition, mutation in *Vrn-A1* gene of winter wheat has been shown to affect the length of vernalization required for transition from vegetative phase to flowering. A SNP occurring in exon 4 of *Vrn-A1* and copy number variation in *Vrn-A1* gene has been associated with shortening the vernalization requirement in short and long vernalization requirement in winter wheat (Chen et al., 2009; Guedira et al., 2016). Photoperiod response is also critical to flowering in wheat. In photoperiod sensitive wheat, long days are required to accelerate flowering after the vernalization requirement has been met. In photoperiod insensitive wheat, the transition from vegetative to reproductive phase takes place after the spring temperature rises. The *PHOTOPERIOD1* (*Ppd1*) gene located on chromosomes 2A, 2B, and 2D control response to photoperiod in wheat where the dominant allele (*Ppd1a*) confers insensitivity. Among the dominant alleles of the three *Ppd1* genes, *Ppd-D1a* showed the most intense effect followed by *Ppd-B1a* and *Ppd-A1a*.

FHB or Scab is one of the most devastating fungal disease of wheat caused mainly by *Fusarium graminearum* Schwabe (telemorph *Gibberella zeae* Schwein). Being a necrotrophic funcgus, effector triggered immunity, known as "gene for gene interaction resistance" is absent for *F. graminearum*. The unavailability of complete resistance against *F. graminearum* makes it highly quantitative trait. Initially, Schroeder and Christensen (1963) dissected FHB resistance into two components: resistance to initial infection in the spikes (Type I) and resistance to spread of the disease within the spike (Type II). In the developing kernels, colonization of fungus leads to shriveled kernels with colors ranging from pink, soft-gray, to light-brown known as *Fusarium* Damaged Kernels (FDK). Severe infestation leads to lower test weights and damaged kernels resulting in loss of yield and quality. Resistance to FDK is known as Type III resistance (Mesterhazy et al., 1995). The fungus produces the mycotoxin known as Deoxynivalenol (DON) that affects animal and human health causing acute temporary nausea, vomiting, and fever (Sobrova et al., 2010). Food and Drugs Administration (FDA) has established guidelines of 1 ppm for DON levels in human food and 10 ppm in animal feed. The presence of higher level of DON in wheat grains leads to discounts in wheat price which further results in financial loss for wheat growers (McMullen et al., 1997). The Type IV resistance is when a plant shows symptoms of

infection and FDK but expresses lower accumulation of DON in the kernels (Mesterhazy et al., 1999).

The locus *Fhb1* is one of the most widely used QTL that provides Type II resistance (McMullen et al., 2012). *Fhb1* in on chromosome 3B and originally derived from the Chinese cultivar Sumai 3 (Waldron et al., 1999) and its derivatives such as Ning 7840 (Bai et al., 1999), and CM-82306 (Buerstmayr et al. 2002). Several attempts had been made to fine map and identify the candidate gene for *Fhb1* but were contradictory. Rawat et al., (2016) reported a pore-foring toxin like (*PFT*) gene as causal gene, however, Su et al., (2019) and Li et al., (2019) reported a histidine-rich calcium binding protein (*TaHRC*) as the candidate gene for the *Fhb1* locus, although they disagree in its mode of action.

In addition to *Fhb1*, two other QTL on chromosome 5A and 6B had been mapped for FHB resistance in a Sumai 3 derivative line CM-82036 (Buerstmayr et al., 2002) and BW278 (Cuthbert et al., 2007). QTL on chromosome 5A, *Qfhs.ifa-5A*, provides Type I resistance while QTL on chromosome 6B, known as *Fhb2*, provides Type II resistance. In population derived from Chinese line Wangshuibai x Nanda 2419, three QTL on chromosome 2B (*QFhb.nau-2B*), 4B (*Fhb4*) and 5A (*Fhb5*) were reported (Lin et al., 2004; Lin et al., 2006). Evidence shows that the genomic positions of *Qfhs.ifa-5A* and *Fhb5* overlap (Jia et al., 2018; Steiner et al., 2019) suggesting a probability of common gene for the two QTL (Buerstmayr et al., 2019). Steiner et al., (2019) showed that *Qfhs.ifa-5A* consists of two QTL *Qfhs.ifa-5Ac* and *Qfhs.ifa-5As* associated with resistance against initial infection and anther extrusion (Steiner et al., 2019).

Exotic introgression from wild relatives of wheat also provided FHB resistance traits. The *Fhb3*, *Fhb6*, and *Fhb7* were introgressed into chromosome 7A, 1A, and 7D from alien species *Leymus racemosus* (Qi et al., 2008), *Elymus tsukushiensis* (Cainong et al., 2015), and *Thinopyrum ponticum* (Guo et al., 2015) respectively. The *Fhb7* locus on distal end of 7E chromosome introgressed in 7D chromosome (Shen and Ohm 2007) and validated by several studies (Zhang et al., 2011; Guo et al., 2015). One of the major caveats of using exotic sources for FHB resistance is the linkage drag associated during introgression in elite lines that impact agronomic and end-use quality performance (Balut et al., 2013).

Genetic studies had been performed to identify QTL in cultivars native to the US. Several soft red winter wheat cultivars including Ernie, Freedom, Massey, NC-Neuse, Jamestown, and Bess have been identified as source of resistance with QTL mapped on chromosomes 3BL from Massey (Liu et al., 2013); 3B and 5A from Ernie (Liu et al., 2007) ; 2B and 3B from Bess (Petersen et al., 2017); 1A, 4A, and 6A from NC-Neuse (Petersen et al., 2016); and 1B from Jamestown (Wright 2014). Similarly, hard red winter wheat cultivars that carry native resistance QTL include Everest, Overland, Lyman, Heyne, and Hondo. QTL mapping identified QTL in chromosomes 2D from Everest (Clinesmith et al., 2019); 2A, 5A, and 6A from Overland (Eckard et al., 2019); and 3A, 4D, and 4A from Heyne (Zhang et al., 2012). Identification of QTL from different sources provide a diverse genetic base to FHB resistance and pyramiding these QTL into cultivars is necessary to provide protection against the devastating disease. The existing body of literature on large effect agronomic, disease resistance, and end-use quality traits has been compiled at MASWheat Server (maswheat.ucdavis.edu) and is practiced in applied wheat breeding programs in the US by a partnership between public wheat breeding programs and the USDA-ARS Small Grains Genotyping Laboratories (wheat.pw.usda.gov/GenotypingLabs/).

The main goal of this PhD dissertation is to characterize phenotypic and genotypic diversity prevalent in the SRWW breeding lines developed at Purdue University over the last few decades. This goal was accomplished by three specific objectives. The first objective was to genotype the population using GBS methodology to generate genome-wide variants and identify genomic regions under selection in the population. The second objective was to evaluate the SRWW breeding population in field environment for two years in standard breeding plot size for agronomic performance for the analysis of agronomic and yield-related traits, and to identify loci that control these traits. The third objective was to evaluate the population's FHB resistance by using row-plots under artificially inoculation and misted system and perform GWAS studies to identify genomic regions associated with FHB traits. This study provides valuable information about hidden population structure in the population, extent of LD decay in each chromosome, distribution of phenotypic data for agronomic performance and scab resistance of specific lines, genomic regions under selection that are important for adaptation, and major QTL and their frequencies in the population.

1.7   References

Anderson, J. A., Stack, R. W., Liu, S., Waldron, B. L., Fjeld, A. D., Coyne, C., ... & Frohberg, R. C. (2001). DNA markers for *Fusarium* head blight resistance QTLs in two wheat populations. *Theoretical and Applied Genetics*, *102*(8), 1164-1168.

Appels, R., Eversole, K., Feuillet, C., Keller, B., Rogers, J., Stein, N., ... & Ronen, G. (2018). Shifting the limits in wheat research and breeding using a fully annotated reference genome. Science, 361(6403), eaar7191.

Asseng, S., Ewert, F., Martre, P., Rötter, R.P., Lobell, D.B., Cammarano, D. et al. (2015). Rising temperatures reduce global wheat production. Nature climate change, 5(2), p.143.

Bai, G., Kolb, F. L., Shaner, G., & Domier, L. L. (1999). Amplified fragment length polymorphism markers linked to a major quantitative trait locus controlling scab resistance in wheat. Phytopathology, 89(4), 343-348.

Balut, A. L., Clark, A. J., Brown-Guedira, G., Souza, E., & Van Sanford, D. A. (2013). Validation of *Fhb1* and *QFhs. nau-2DL* in several soft red winter wheat populations. Crop science, 53(3), 934-945.

Bhave, M., & Morris, C. F. (2008). Molecular genetics of puroindolines and related genes: allelic diversity in wheat and other grasses. Plant molecular biology, 66(3), 205-219.

Borevitz, J. O., & Nordborg, M. (2003). The impact of genomics on the study of natural variation in *Arabidopsis*. *Plant physiology*, *132*(2), 718-725.

Breseghello, F., & Sorrells, M. E. (2006). Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars. *Genetics*, *172*(2), 1165-1177.

Buerstmayr, H., Lemmens, M., Hartl, L., Doldi, L., Steiner, B., Stierschneider, M., & Ruckenbauer, P. (2002). Molecular mapping of QTLs for *Fusarium* head blight resistance in spring wheat. I. Resistance to fungal spread (Type II resistance). Theoretical and Applied Genetics, 104(1), 84-91.

Buerstmayr, M., Steiner, B., & Buerstmayr, H. (2019). Breeding for *Fusarium* head blight resistance in wheat—Progress and challenges. Plant Breeding.

Cavanagh, C. R., Chao, S., Wang, S., Huang, B. E., Stephen, S., Kiani, S., ... & See, D. (2013). Genome-wide comparative diversity uncovers multiple targets of selection for improvement in hexaploid wheat landraces and cultivars. Proceedings of the national academy of sciences, 110(20), 8057-8062.

Chao, S., Sharp, P. J., Worland, A. J., Warham, E. J., Koebner, R. M. D., & Gale, M. D. (1989). RFLP-based genetic maps of wheat homoeologous group 7 chromosomes. *Theoretical and Applied Genetics*, *78*(4), 495-504.

Chen, Y., Carver, B. F., Wang, S., Zhang, F., & Yan, L. (2009). Genetic loci associated with stem elongation and winter dormancy release in wheat. Theoretical and Applied Genetics, 118(5), 881-889.

Clinesmith, M. A., Fritz, A. K., Lemes da Silva, C., Bockus, W. W., Poland, J. A., Dowell, F. E., & Peiris, K. H. (2019). QTL Mapping of *Fusarium* Head Blight Resistance in Winter Wheat Cultivars 'Art'and 'Everest'. Crop Science, 59(3), 911-924.

Cockram, J., White, J., Zuluaga, D. L., Smith, D., Comadran, J., Macaulay, M., ... & Tapsell, C. (2010). Genome-wide association mapping to candidate polymorphism resolution in the unsequenced barley genome. *Proceedings of the National Academy of Sciences*, *107*(50), 21611-21616.

Collard, B. C., & Mackill, D. J. (2008). Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1491), 557-572.

Collard, B. C., Jahufer, M. Z. Z., Brouwer, J. B., & Pang, E. C. K. (2005). An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: the basic concepts. *Euphytica*, *142*(1-2), 169-196.

Crossa, J., Burgueno, J., Dreisigacker, S., Vargas, M., Herrera-Foessel, S. A., Lillemo, M., ... & Reynolds, M. (2007). Association analysis of historical bread wheat germplasm using additive genetic covariance of relatives and population structure. *Genetics*, *177*(3), 1889-1913.

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.). Theoretical and Applied Genetics, 114(3), 429-437.

Dvorak, J., McGuire, P. E., & Cassidy, B. (1988). Apparent sources of the A genomes of wheats inferred from polymorphism in abundance and restriction fragment length of repeated nucleotide sequences. *Genome*, *30*(5), 680-689.

Eckard, J. T., Gonzalez-Hernandez, J. L., Caffe, M., Berzonsky, W., Bockus, W. W., Marais, G. F., & Baenziger, P. S. (2015). Native *Fusarium* head blight resistance from winter wheat cultivars 'Lyman,''Overland,''Ernie,'and 'Freedom'mapped and pyramided onto 'Wesley'-Fhb1 backgrounds. Molecular breeding, 35(1), 6.

Ehrenreich, I. M., Hanzawa, Y., Chou, L., Roe, J. L., Kover, P. X., & Purugganan, M. D. (2009). Candidate gene association mapping of Arabidopsis flowering time. *Genetics*, *183*(1), 325-335.

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PloS one*, *6*(5).

Faris, J. D. (2014). Wheat domestication: Key to agricultural revolutions past and future. In *Genomics of plant genetic resources* (pp. 439-464). Springer, Dordrecht.

Feldman, M., & Levy, A. A. (2015). Origin and evolution of wheat and related *Triticeae* species. In *Alien introgression in Wheat* (pp. 21-76). Springer, Cham.

Food and Agriculture Organization of the United Nations. (2020). FAOSTAT statistical database.

Gaire, R., Huang, M., Sneller, C., Griffey, C., Brown-Guedira, G., & Mohammadi, M. (2019). Association Analysis of Baking and Milling Quality Traits in an Elite Soft Red Winter Wheat Population. *Crop Science*, *59*(3), 1085-1094.

Giroux, M. J., & Morris, C. F. (1998). Wheat grain hardness results from highly conserved mutations in the friabilin components puroindoline a and b. *Proceedings of the National Academy of Sciences*, *95*(11), 6262-6266.

Greenwell, P. and Schofield, J. D. Cereal Chern. 63 (1986) 379-380.

Guedira, M., Xiong, M., Hao, Y. F., Johnson, J., Harrison, S., Marshall, D., & Brown-Guedira, G. (2016). Heading date QTL in winter wheat (*Triticum aestivum* L.) coincide with major developmental genes *VERNALIZATION1* and *PHOTOPERIOD1*. *PloS one*, *11*(5).

Guo, J., Zhang, X., Hou, Y., Cai, J., Shen, X., Zhou, T., ... & Han, F. (2015). High-density mapping of the major FHB resistance gene *Fhb7* derived from *Thinopyrum ponticum* and its pyramiding with *Fhb1* by marker-assisted selection. Theoretical and applied genetics, 128(11), 2301-2316.

Gupta, P. K., Varshney, R. K., Sharma, P. C., & Ramesh, B. (1999). Molecular markers and their applications in wheat breeding. *Plant breeding*, *118*(5), 369-390.

Himi, E., & Noda, K. (2005). Red grain colour gene (*R*) of wheat is a *Myb-type* transcription factor. *Euphytica*, *143*(3), 239-242.

Hoffstetter, A., Cabrera, A., & Sneller, C. (2016). Identifying quantitative trait loci for economic traits in an elite soft red winter wheat population. *Crop Science*, *56*(2), 547-558.

Huang, S., Sirikhachornkit, A., Su, X., Faris, J., Gill, B., Haselkorn, R., & Gornicki, P. (2002). Genes encoding plastid acetyl-CoA carboxylase and 3-phosphoglycerate kinase of the *Triticum/Aegilops* complex and the evolutionary history of polyploid wheat. *Proceedings of the National Academy of Sciences*, *99*(12), 8133-8138.

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

Jantasuriyarat, C., Vales, M. I., Watson, C. J. W., & Riera-Lizarazu, O. (2004). Identification and mapping of genetic loci affecting the free-threshing habit and spike compactness in wheat (*Triticum aestivum* L.). Theoretical and Applied Genetics, 108(2), 261-273.

Jia, H., Zhou, J., Xue, S., Li, G., Yan, H., Ran, C., ... & Luo, J. (2018). A journey to understand wheat *Fusarium* head blight resistance in the Chinese wheat landrace Wangshuibai. The Crop Journal, 6(1), 48-59.

Kang, H. M., Sul, J. H., Service, S. K., Zaitlen, N. A., Kong, S. Y., Freimer, N. B., ... & Eskin, E. (2010). Variance component model to account for sample structure in genome-wide association studies. Nature genetics, 42(4), 348.

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.

Kirigwi, F. M., Van Ginkel, M., Brown-Guedira, G., Gill, B. S., Paulsen, G. M., & Fritz, A. K. (2007). Markers associated with a QTL for grain yield in wheat under drought. *Molecular Breeding*, *20*(4), 401-413.

Korte, A., & Farlow, A. (2013). The advantages and limitations of trait analysis with GWAS: a review. *Plant methods*, *9*(1), 29.

Kumar, N., Kulwal, P. L., Balyan, H. S., & Gupta, P. K. (2007). QTL mapping for yield and yield contributing traits in two mapping populations of bread wheat. *Molecular Breeding*, *19*(2), 163-177.

Law, C. N., & Worland, A. J. (1997). Genetic analysis of some flowering time and adaptive traits in wheat. *The New Phytologist*, *137*(1), 19-28.

Law, C. N., Young, C. F., Brown, J. W. S., Snape, J. W. & Worland, J. W. (1978) in Seed Protein Improvement by Nuclear Techniques (International Atomic Energy Agency, Vienna, Austria), pp. 483–502.

Li, G., Zhou, J., Jia, H., Gao, Z., Fan, M., Luo, Y., ... & Ma, S. (2019). Mutation of a histidine-rich calcium-binding-protein gene in wheat confers resistance to *Fusarium* head blight. Nature genetics, 51(7), 1106-1112.

Lillemo, M., & Morris, C. F. (2000). A leucine to proline mutation in puroindoline b is frequently present in hard wheats from Northern Europe. *Theoretical and Applied Genetics*, *100*(7), 1100-1107.

Lin, F., Kong, Z. X., Zhu, H. L., Xue, S. L., Wu, J. Z., Tian, D. G., ... & Ma, Z. Q. (2004). Mapping QTL associated with resistance to *Fusarium* head blight in the Nanda2419× Wangshuibai population. I. Type II resistance. Theoretical and Applied Genetics, 109(7), 1504-1511.

Lin, F., Xue, S. L., Zhang, Z. Z., Zhang, C. Q., Kong, Z. X., Yao, G. Q., ... & Wei, J. B. (2006). Mapping QTL associated with resistance to *Fusarium* head blight in the Nanda2419× Wangshuibai population. II: Type I resistance. Theoretical and Applied Genetics, 112(3), 528-535.

Lippert, C., Listgarten, J., Liu, Y., Kadie, C. M., Davidson, R. I., & Heckerman, D. (2011). FaST linear mixed models for genome-wide association studies. Nature methods, 8(10), 833.

Liu, S., Abate, Z. A., Lu, H., Musket, T., Davis, G. L., & McKendry, A. L. (2007). QTL associated with *Fusarium* head blight resistance in the soft red winter wheat Ernie. Theoretical and Applied Genetics, 115(3), 417-427.

Liu, S., Griffey, C. A., Hall, M. D., McKendry, A. L., Chen, J., Brooks, W. S., ... & Schmale, D. G. (2013). Molecular characterization of field resistance to *Fusarium* head blight in two US soft red winter wheat cultivars. Theoretical and applied genetics, 126(10), 2485-2498.

Liu, X., Huang, M., Fan, B., Buckler, E. S., & Zhang, Z. (2016). Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies. PLoS genetics, 12(2).

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. *Theoretical and Applied Genetics*, *111*(3), 496-504.

MASwheat (2020). Marker-assisted Selection in Wheat. https://maswheat.ucdavis.edu/. Accesseed online 14 Feb. 2020.

Matsuoka, Y. (2011). Evolution of polyploid *Triticum* wheats under cultivation: the role of domestication, natural hybridization and allopolyploid speciation in their diversification. Plant and cell physiology, 52(5), 750-764.

Mattern, P. J., Morris, R., Schmidt, J. W.&Johnson, V. A. (1973). in Proceedings of the 4th International Wheat Genetics Symposium (Univ. Missouri, Columbia, MO), pp. 703–707.

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.

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). Types and components of resistance to *Fusarium* head blight of wheat. Plant breeding, 114(5), 377-386.

Mesterházy, Á., Bartók, T., Mirocha, C. G., & Komoroczy, R. (1999). Nature of wheat resistance to *Fusarium* head blight and the role of deoxynivalenol for breeding. Plant breeding, 118(2), 97-110.

Metzger, R. J., & Silbaugh, B. A. (1970). Location of genes for seed coat color in hexaploid wheat, triticum aestivum l. 1. *Crop science*, *10*(5), 495-496.

Miedaner, T., & Korzun, V. (2012). Marker-assisted selection for disease resistance in wheat and barley breeding. *Phytopathology*, *102*(6), 560-566.

Miki, Y., Yoshida, K., Mizuno, N., Nasuda, S., Sato, K., & Takumi, S. (2019). Origin of wheat B-genome chromosomes inferred from RNA sequencing analysis of leaf transcripts from section Sitopsis species of Aegilops. *DNA Research*, *26*(2), 171-182.

Morris, C. F. (2002). Puroindolines: the molecular genetic basis of wheat grain hardness. *Plant molecular biology*, *48*(5-6), 633-647.

Morris, C. F., Lillemo, M., Simeone, M. C., Giroux, M. J., Babb, S. L., & Kidwell, K. K. (2001). Prevalence of puroindoline grain hardness genotypes among historically significant North American spring and winter wheats. *Crop Science*, *41*(1), 218-228.

Ohm H., McFee W.W. (2012). Small Grains Breeding at Purdue University. Agronomy, Purdue University. Retrieved from https://ag.purdue.edu/agry/Documents/Small%20Grains%20Breeeding%20at%20Purdue KLS.pdf

Peng, J., Richards, D. E., Hartley, N. M., Murphy, G. P., Devos, K. M., Flintham, J. E., ... & Sudhakar, D. (1999). 'Green revolution' genes encode mutant gibberellin response modulators. Nature, 400(6741), 256-261.

Peng, J., Ronin, Y., Fahima, T., Röder, M. S., Li, Y., Nevo, E., & Korol, A. (2003). Domestication quantitative trait loci in Triticum dicoccoides, the progenitor of wheat. *Proceedings of the National Academy of Sciences*, *100*(5), 2489-2494.

Petersen, S., Lyerly, J. H., Maloney, P. V., Brown-Guedira, G., Cowger, C., Costa, J. M., ... & Murphy, J. P. (2016). Mapping of *Fusarium* head blight resistance quantitative trait loci in winter wheat cultivar NC-Neuse. Crop Science, 56(4), 1473-1483.

Petersen, S., Lyerly, J. H., McKendry, A. L., Islam, M. S., Brown-Guedira, G., Cowger, C., ... & Murphy, J. P. (2017). Validation of *Fusarium* head blight resistance QTL in US winter wheat. Crop Science, 57(1), 1-12.

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).

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. Theoretical and Applied Genetics, 117(7), 1155-1166.

Rawat, N., Pumphrey, M. O., Liu, S., Zhang, X., Tiwari, V. K., Ando, K., ... & 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.

Ray, D.K., Mueller, N.D., West, P.C. and Foley, J.A. (2013). Yield trends are insufficient to double global crop production by 2050. PloS one, 8(6), p.e66428.

Risch, N., & Merikangas, K. (1996). The future of genetic studies of complex human diseases. *Science*, *273*(5281), 1516-1517.

Röder, M. S., Korzun, V., Wendehake, K., Plaschke, J., Tixier, M. H., Leroy, P., & Ganal, M. W. (1998). A microsatellite map of wheat. *Genetics*, *149*(4), 2007-2023.

Roy, J. K., Bandopadhyay, R., Rustgi, S., Balyan, H. S., & Gupta, P. K. (2006). Association analysis of agronomically important traits using SSR, SAMPL and. *Current science*, *90*(5).

Salamini, F., Özkan, H., Brandolini, A., Schäfer-Pregl, R., & Martin, W. (2002). Genetics and geography of wild cereal domestication in the near east. Nature Reviews Genetics, 3(6), 429-441.

Salvi, S., & Tuberosa, R. (2015). The crop QTLome comes of age. Current opinion in biotechnology, 32, 179-185.

Schmidt, J. W. (1984). Genetic contributions to yield gains in wheat. *Genetic contributions to yield gains of five major crop plants*, *7*, 89-101.

Schroeder, H. W., & Christensen, J. J. (1963). Factors affecting resistance of wheat to scab caused by Gibberella zeae. Phytopathology, 53(7, 1), 831-838.

Segura, V., Vilhjálmsson, B. J., Platt, A., Korte, A., Seren, Ü., Long, Q., & Nordborg, M. (2012). An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. Nature genetics, 44(7), 825.

Shen, X., & Ohm, H. (2007). Molecular mapping of Thinopyrum-derived *Fusarium* head blight resistance in common wheat. Molecular Breeding, 20(2), 131-140.

Simonetti, M. C., Bellomo, M. P., Laghetti, G., Perrino, P., Simeone, R., & Blanco, A. (1999). Quantitative trait loci influencing free-threshing habit in tetraploid wheats. Genetic Resources and Crop Evolution, 46(3), 267-271.

Slafer, G. A., Calderini, D. F., & Miralles, D. J. (1996). Yield components and compensation in wheat: opportunities for further increasing yield potential. Increasing yield potential in wheat: breaking the barriers, 101-133.

Sobrova, P., Adam, V., Vasatkova, A., Beklova, M., Zeman, L., & Kizek, R. (2010). Deoxynivalenol and its toxicity. Interdisciplinary toxicology, 3(3), 94-99.

Sourdille, P., Perretant, M. R., Charmet, G., Leroy, P., Gautier, M. F., Joudrier, P., ... & Bernard, M. (1996). Linkage between RFLP markers and genes affecting kernel hardness in wheat. *Theoretical and Applied Genetics*, *93*(4), 580-586.

Spindel, J., Begum, H., Akdemir, D., Virk, P., Collard, B., Redona, E., ... & McCouch, S. R. (2015). Genomic selection and association mapping in rice (*Oryza sativa*): effect of trait genetic architecture, training population composition, marker number and statistical model on accuracy of rice genomic selection in elite, tropical rice breeding lines. PLoS Genet, 11(2), e1004982.

Steiner, B., Buerstmayr, M., Wagner, C., Danler, A., Eshonkulov, B., Ehn, M., & Buerstmayr, H. (2019). Fine-mapping of the *Fusarium* head blight resistance QTL *Qfhs. ifa-5A* identifies two resistance QTL associated with anther extrusion. Theoretical and applied genetics, 132(7), 2039-2053.

Su, Z., Bernardo, A., Tian, B., Chen, H., Wang, S., Ma, H., ... & Trick, H. (2019). A deletion mutation in *TaHRC* confers *Fhb1* resistance to *Fusarium* head blight in wheat. Nature genetics, 51(7), 1099-1105.

Tsilo, T. J., Chao, S., Jin, Y., & Anderson, J. A. (2009). Identification and validation of SSR markers linked to the stem rust resistance gene *Sr6* on the short arm of chromosome 2D in wheat. *Theoretical and applied genetics*, *118*(3), 515-524.

USDA-ERS. 2020. Wheat data. USDA Econ. Res. Serv. https://www.ers.usda.gov/data-products/wheat-data/ (accessed 15 Feb. 2020).

USDA-NASS. 2019. Quick stats. USDA Natl. Agric. Stat. Serv. https://quickstats.nass.usda.gov/results/9140FA3B-C859-3EA7-89BB-6E7FEC709F7E?pivot=short_desc (accessed 10 Nov. 2019).

Vagndorf, N., Kristensen, P. S., Andersen, J. R., Jahoor, A., & Orabi, J. (2018). Marker-Assisted Breeding in Wheat. *Next Generation Plant Breeding*, 1.

Waldron, B. L., Moreno-Sevilla, B., Anderson, J. A., Stack, R. W., & Frohberg, R. C. (1999). RFLP mapping of QTL for *Fusarium* head blight resistance in wheat. Crop Science, 39(3), 805-811.

Wang, Q., Tian, F., Pan, Y., Buckler, E. S., & Zhang, Z. (2014b). A SUPER powerful method for genome wide association study. PloS one, 9(9).

Wang, S., Wong, D., Forrest, K., Allen, A., Chao, S., Huang, B. E., ... & Mastrangelo, A. M. (2014a). Characterization of polyploid wheat genomic diversity using a high-density 90 000 single nucleotide polymorphism array. *Plant biotechnology journal*, *12*(6), 787-796.

Watanabe, N., Sugiyama, K., Yamagishi, Y. and Sakata, Y., 2002. Comparative telosomic mapping of homoeologous genes for brittle rachis in tetraploid and hexaploid wheats. *Hereditas*, *137*(3), pp.180-185.

Weng, J., Xie, C., Hao, Z., Wang, J., Liu, C., Li, M., ... & Li, X. (2011). Genome-wide association study identifies candidate genes that affect plant height in Chinese elite maize (Zea mays L.) inbred lines. *PLoS One*, *6*(12).

Wright, E. E. (2014). Identification of Native FHB Resistance QTL in the SRW Wheat Cultivar Jamestown (Doctoral dissertation, Virginia Tech).

Xu, Y., Li, P., Zou, C., Lu, Y., Xie, C., Zhang, X., ... & Olsen, M. S. (2017). Enhancing genetic gain in the era of molecular breeding. *Journal of Experimental Botany*, *68*(11), 2641-2666.

Yan, L., Fu, D., Li, C., Blechl, A., Tranquilli, G., Bonafede, M., ... & Dubcovsky, J. (2006). The wheat and barley vernalization gene *VRN3* is an orthologue of FT. *Proceedings of the National Academy of Sciences*, *103*(51), 19581-19586.

Yan, L., Helguera, M., Kato, K., Fukuyama, S., Sherman, J., & Dubcovsky, J. (2004b). Allelic variation at the *VRN-1* promoter region in polyploid wheat. *Theoretical and applied genetics*, *109*(8), 1677-1686.

Yan, L., Loukoianov, A., Blechl, A., Tranquilli, G., Ramakrishna, W., SanMiguel, P., ... & Dubcovsky, J. (2004a). The wheat *VRN2* gene is a flowering repressor down-regulated by vernalization. *Science*, *303*(5664), 1640-1644.

Yan, L., Loukoianov, A., Tranquilli, G., Helguera, M., Fahima, T., & Dubcovsky, J. (2003). Positional cloning of the wheat vernalization gene *VRN1*. *Proceedings of the National Academy of Sciences*, *100*(10), 6263-6268.

Yu, J., Pressoir, G., Briggs, W. H., Bi, I. V., Yamasaki, M., Doebley, J. F., ... & Kresovich, S. (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nature genetics, 38(2), 203-208.

Zhang, X., Bai, G., Bockus, W., Ji, X., & Pan, H. (2012). Quantitative trait loci for *Fusarium* head blight resistance in US hard winter wheat cultivar Heyne. Crop science, 52(3), 1187-1194.

Zhang, X., Shen, X., Hao, Y., Cai, J., Ohm, H. W., & Kong, L. (2011). A genetic map of *Lophopyrum ponticum* chromosome 7E, harboring resistance genes to *Fusarium* head blight and leaf rust. Theoretical and Applied Genetics, 122(2), 263-270.

Zhang, Z., Ersoz, E., Lai, C. Q., Todhunter, R. J., Tiwari, H. K., Gore, M. A., ... & Buckler, E. S. (2010). Mixed linear model approach adapted for genome-wide association studies. Nature genetics, 42(4), 355.

CHAPTER 2.   IDENTIFICATION OF GENOMIC REGIONS UNDER SELECTION AND LOCI CONTROLLING AGRONOMIC TRAITS IN A SOFT RED WINTER WHEAT POPULATION

2.1   Abstract

Comprehensive information of a breeding population is a necessity to design promising crosses. This study was conducted to characterize a soft red winter wheat breeding population that was subject of intensive germplasm introductions and introgression from exotic germplasm. We used genome-wide markers and phenotypic assessment to identify signatures of selection and loci controlling agronomic traits in a soft red winter wheat population. The study of linkage disequilibrium (LD) revealed that the extent of LD and its decay varied among chromosomes with chromosomes 2B and 7D showing the most extended islands of high-LD slow rates of decay. Four sub-populations, two with North American origin and two with Australian and Chinese origins, were identified. Genome-wide scans for selection signatures using F_{ST} and hapFLK identified 13 genomic regions under selection, of which five loci (*LT*, *Fr-A2*, *Vrn-A1*, *Vrn-B1*, *Vrn3*) were associated with environmental adaptation and two loci were associated with disease resistance genes (*Sr36* and *Fhb1*). Genome-wide association studies identified major loci controlling yield and yield related traits. For days to heading and plant height, large effects loci were identified in chromosome 6A and 7B. For test weight, number of spikes per square meter, and number of kernels per square meter, large effect loci were identified on chromosomes 1A, 4B, and 5A, respectively. However, for yield alone, no major loci were detected. A combination of selection

for large effect loci for yield components and genomic selection could be a promising approach for yield improvement.

2.2 Introduction

Wheat occupies almost one-fifth of world's cultivated land and is a major staple food crop for almost two-thirds of world population. In the United States, wheat ranked third in terms of revenue and acres planted after corn and soybean (USDA-NASS, 2019). In 2018, wheat was planted in approximately 16 million hectares producing 48 million metric tons that valued approximately 9.7 billion dollars in the US (USDA-NASS, 2019). Over the last century, conventional breeding methodologies were able to exploit genetic variation in food crops that led to nearly 1% annual genetic gains in five major food crops including wheat (Ray et al., 2013). However, by the year 2050, global population is expected to reach 9.1 billion people, adding 1.4 billion to the current population (FAO, 2009). To meet the growing demands of food, annual cereal production should increase by 2.4% and should add one billion tons of wheat to total production every year (Ray et al., 2013). In addition, climate change brings uncertainty in crop production, anticipated to decrease wheat production over the next decade by 6% for every increase of 1°C on global temperature (Asseng et al., 2015). To address these challenges, plant breeders must leverage information from recent advancement in next-generation sequencing technologies, bioinformatics, genomics, and phenotyping to better characterize and exploit genetic diversity for increasing genetic gains.

Existing germplasm in breeding populations are sources of native genes and desirable phenotypes. Each breeding program have several germplasms that harbor the desired adaptive genes e.g., for disease resistance or abiotic stress tolerance for breeding. When native genes are not enough to address environmental challenges or end-use marketability, germplasm introduction and exotic introgression is a necessity. to bring novel genes. For example, *Fusarium* head blight (FHB) is one of the most devastating disease of wheat and barley in the US. FHB or Scab is one of the most devastating fungal disease of wheat caused mainly by *Fusarium graminearum* Schwabe (telemorph *Gibberella zeae* Schwein). Since *Fusarium graminearum* is a necrotrophic fungus, effector triggered immunity (alternatively referred to as gene for gene resistance) is not available for FHB, and only small effect loci are available to control the resistance. The existing

native sources of resistance (Ernie, Freedom, Lyman, Overland, Jin et al., 2013) have not been sufficient to improve disease resistance. The *Fhb1* allele (Anderson et al., (2001), from the Chinese cultivar 'Sumai 3' has been widely introduced and incorporated in many breeding germplasms in North America (Buerstmayr et al., 2009; Jin et al., 2013).

In addition to introduction from the primary genepool, introgression of alien chromatin from wild relatives harboring useful exogenous genes has been successfully exercised to enrich the genetic bases of disease resistance. For example, barley yellow dwarf resistance gene *Bdv3* from Thinopyrum intermedium (Sharma et al., 1995) and the exotic *Fusarium* head blight *Fhb7* from Th. elongatum (Shen and Ohm, 2007; Guo et al.,2015) were successfully transferred into the hexaploid wheat (Triticum aestivum L.). The introgressed alien chromatin region is not able to recombine when crossed with wild type chromatin due to lack of homologous pairing (Gill et al., 2011), and therefore creates an extended non-recombinant chromatin in the chromosomal region where it is introgressed. This phenomenon will result in extended islands of high linkage disequilibrium (Nagy et al., 2010).

Trait-based analyses such as bi-parental mapping or genome-wide association studies (GWAS) have been extensively used to identify quantitative trait loci (QTLs) and genomic regions that control traits. QTL mapping is time- and resource-consuming, has low mapping resolution, and does not capture all allelic diversity in germplasm (Myles et al., 2009). In contrast, GWAS identifies loci and candidate genes controlling traits in natural or breeding populations (Gupta et al., 2014). GWAS has been used in all major crops including barley (Adhikari et al., 2020), corn (Xiao et al., 2016), rice (Muers 2011), soybean (Patil et al., 2017), and wheat (Gaire et al., 2019). While popular, GWAS requires phenotypic evaluation of germplasm preferentially in multiple environments and has better power to detect causative regions common (and not rare) variants within the population (Myles et al., 2009).

A complementary approach to delineate genomic regions and genes associated with adaptive responses is to identify signatures of selection or selective sweeps (SS) (Evan et al., 2014; Fariello et al., 2014; Xie et al., 2015). In plants, genome-wide SNP markers had been used to detect SS in maize (Beissinger et al 2014; Gage et al., 2014), rapeseed (Wei et al., 2017), rice (Xu et al., 2011; Xie et al., 2015), soybean (Zhou et al., 2015), tomato (Lin et al., 2014), and wheat (Cavanagh et

al 2013; Gao et al., 2017). In rice, Xie et al., (2015) evaluated breeding signatures by using ~1500 lines, revealing that SS are enriched for adaptive traits such as semi-dwarfism. In wheat, Cavanagh et al (2013) used genome-wide markers to identify signatures of domestication and post-domestication. They estimated the extent of genetic differentiation (F_{ST}) between spring and winter wheat and reported that *Vrn-A1*, *Vrn-B1*, and *Vrn-B3* have been significantly selected. Similarly, *Ppd-B1*, *Rht-B1*, *Rht-D1* and *Vrn-A1* loci were found to be selected during domestication in Gao et al. (2017) study.

The population differentiation-based method is a simple and yet powerful approach to identify SS (Vitti et al 2013). When a locus has been under selection in one population but not in the other related populations, then the allele frequencies at that locus differ significantly among the populations (Holsinger and Weir 2009). A commonly used and well-known method in this class is the Wright's Fixation index (F_{ST}). The F_{ST} approach compares the variances of allele frequency within and between subpopulations (Vitti et al., 2013). Lower F_{ST} values indicate neutral selection whereas higher F_{ST} values are suggestive of considerable selection (Holsinger and Weir, 2009). In other words, higher values of F_{ST} mean that the variations among subpopulations out-yields the variations within subpopulations.

While commonly used, F_{ST} method does not account for heterogeneity of sub-population sizes, shared ancestry, and hierarchical structure of subpopulations (Bonhomme et al., 2010; Fariello et al., 2013; Holsinger and Weir, 2015). That is, regions may be identified that are inherent to these underlying factors rather than differentiation, which results in the identification of false positive regions. Bonhomme et al (2010) suggested to account for hierarchical structure and heterogeneity of effective population size. They modified the F_{ST} approach to FLK test where allele frequencies are first rescaled by using a population kinship matrix based on the Reynolds' distance (Reynolds et al., 1983). Later, Fariello et al., (2013) modified FLK test such that instead of SNP frequencies, haplotype frequencies are used in the estimating F_{ST}. Therefore, the F_{ST} statistics which is after consideration of hierarchical structure and heterogeneity of effective population size, and uses haplotype frequency rather than SNP frequency, is called hapFLK. The hapFLK approach has been commonly used in breeding of animals such as sheep (Fariello et al 2014), goat (Brito et al., 2017), cattle (Yurchenko et al., 2018), and chickens (Gholami et al., 2015); Brito et al., 2017; Gholami et al., 2015). In sheep, Fariello et al., 2014 identified SS that were

selected for morphology, color, and adaptation. In goat, Brito et al., (2017) identified SS for fertility, reproduction and adult body mass. This approach can be extended to evaluate selection history of crop plants when phenotyping large experimental populations without the need to phenotype the population in large experimental settings.

In this study, we examine genetic diversity in a US SRW wheat breeding population developed at small grain breeding program at Purdue University, which released cultivars as early as 1920s. During the 1970s, cultivars release from this program covered more than 80% of SRW wheat acreages in the United States (Ohm and McFee, 2012). Given such breeding history which involved germplasm introduction and alien introgression for various biotic stress resistance genes, the objectives of this study were to characterize the genotypic and phenotypic diversity of the germplasm, as related to environmental adaptation and agronomic traits.

2.3 Materials and Methods

2.3.1 Genotyping

We genotyped 436 SRW wheat germplasm from Purdue's small grain breeding population. The materials were developed from 123 crosses conducted from 1988 to 2010. Leaf samples from single plants grown in the greenhouse were used for genomic DNA extraction and preparing reduced representation libraries following *Pst1-Msp1* digestion (Poland et al. 2012). The libraries were 144-plex sequenced on three lanes of Illumina Hi-Seq 2500. The NGS data was processed using TASSEL 5 (Bradbury et al., 2007) GBS v2 Pipeline. The reads were aligned to IWGSC RefSeq assembly v1.0 (Appels et al., 2018). We removed single nucleotide polymorphisms (SNPs) that had more than 20% missing data and those with less than 5% of minor allele frequency. Then, the remaining missing genotypic data was imputed using LDKNNimp method (Money et al., 2015) implemented in TASSELv 5.2 (Bradbury et al., 2007).

The population was genotyped for major genes at USDA-ARS Eastern Regional Small Grains Genotyping Lab, with the primer sequences indicated by Ward et al., (2019). Briefly, these large effect loci included the population was genotyped for the 1RS:1BL translocations from rye (*Secale cereale* L.), stem rust resistance *Sr36*, photoperiod loci located on the group 2

homoeologous chromosomes *Ppd-1*, reduced height locus *Rht-B1* and *Rht-D1* located on 4B and 4D respectively, vernalization *Vrn1* loci for spring and winter alleles located on the group 5 homeologous chromosomes, and a SNP located on the 4th exon of *vrn-A1* that supposedly discriminates the *vrn-A1a* short vernalization requiring phenotype against the long vernalization requiring, and *vrn-A1b* (Chen et al., 2009).

2.3.2 Population structure and the extent of linkage disequilibrium

Population structure was analyzed using admixture model-based clustering method implemented in STRUCTURE v2.3 (Pritchard et al., 2003). A total of 100,000 burn-in periods followed by 100,000 Markov Chain-Monte Carlo iterations for K values from 1 to 10, with five independent runs for each K was used. Markers with linkage disequilibrium (LD) r2 > 0.5 within a 50 SNPs window size were trimmed using '--indep-pairwise' function in PLINK 1.9 (Chang et al., 2015). The optimal number of subpopulations were determined from delta K (Δk) model as proposed by Evanno et al. (2005). The relationship between the subpopulations identified by STRUCTURE was illustrated by the neighbor-joining algorithm and a genomic population trees which uses pair-wise population Reynold's distance (Reynold et al., 1983) estimated from all markers. Estimation of Reynolds distance and drawing was performed using hapflk release 1.3 (Fariello et al., 2013). Principal component analysis (PCA) was performed in TASSELv5.2 to depict the population structure. Pairwise LD between the markers estimated in TASSEL along each chromosome and for all markers was used to visualize the LD decay according to the method described by Hill and Weir (1988), which was later modified by Remington et al. (2001). The analysis was performed in R environment for each chromosome separately and heatmaps for pairwise LD per chromosome was generated using TASSEL v5.2.

2.3.3 Whole genome scan for selection sweeps

The subpopulations defined by STRUCTURE program were used for F_{ST} analysis based on the pure drift model proposed by Nicholson et al (2002) in R v3.4.2 as illustrated by Porto-Neto et al (2013). Allele frequency for each variant was estimated within each subpopulation and in the whole population. The squared deviation of frequency for each subpopulation were divided by the variance of allele frequencies in the population to estimate the F_{ST} values. Smoothing of the F_{ST}

values was performed using R-package Lokern (Herrmann and Maechler, 2016) as suggested by Porto-Neto et al (2013). Variants and regions showing three standard deviations smoothed F_{ST} values beyond average were declared as significantly under selection and therefore SS.

HapFLK statistics were estimated using hapFLK v1.3 software available at https://forge-dga.jouy.inra.fr/projects/hapflk and using a methodology described by Fariello et al (2013). Briefly, Reynolds distances between lines were estimated using midpoint as outgroup and converted into a kinship matrix using an R script supplied with the hapFLK v1.3 package (Fariello et al., 2013). We ran hapFLK per each chromosome using the genotypes and kinship matrix assuming 15 clusters in the fastPHASE model, before the hapFLK statistic was computed as the average across 10 expectation maximization runs to fit the LD model. The P-values for hapFLK statistics were estimated using the "scaling_chi2_hapflk.py"- Python script supplied at the website (https://forge-dga.jouy.inra.fr/projects/hapflk), [04/28/2019]. The negative log p-value (-$logP$) for each variant was plotted using "ggplot2" package in R environment. The variants with P-value less than 0.01 [-$logP$ > 2] were declared as significant.

The coordinate of markers or selected regions were used to retrieve the genes from IWGSC v1.0 genome assembly (Appels et al., 2018). The literature was mined for evidence of known loci and genes for agronomic or quality traits. Nucleotide sequences of potential candidate genes were retrieved from NCBI (https://www.ncbi.nlm.nih.gov/), and their oligo primer sequences, retrieved from GrainGenes database (https://wheat.pw.usda.gov/cgi-bin/GG3/browse.cgi), were blasted against the wheat genome assembly to localize their position in relation to the regions we identified. This facilitated to form potential hypotheses why these regions were under selection.

2.3.4 Agronomic assessment

A total of 392 out of 436 entries were used in field testing in 2017-18 and 2018-19 seasons in Indiana. Due to limited availability of seeds, the experiment was planted in four unbalanced trails in randomized incomplete block designs. Each incomplete block consisted of 64 plots including 56 entries and eight replicates of a common check "PU212". Each plot was 12 ft long by 4 ft wide (~3 m x 1 m). In the 2017-18, 336 lines were evaluated in a single replicated experiment with 6 blocks (336 lines and 48 common check) at the Agronomy Center for Research

and Extension (ACRE), 40.47°N, 86.99°W, elevation 185m, West Lafayette, IN. In 2018-19 season, 392 lines were evaluated in ACRE and Romney (40.3°N, 86.9°W, elevation 225m) where each experiment consisted of 7 incomplete blocks (392 lines and 56 common check). Additionally, in ACRE, another trail was performed in 5 incomplete blocks (280 lines and 40 common check).

Days to heading (DH), plant height (PH), test weight (TW) and grain yield (YLD) were recorded from all trials and plots. Yield components i.e., number of spikes per unit area (NS) and number of kernels per unit area (NK) were collected from ACRE in 2017-18 and from Romney in 2018-19 experiments. DH was recorded when more than 50% of the plants in each plot showed completely emerged heads from boots at Zadoks stage 59. PH in centimeters (cm) was measured from the base of the plants up to the top of the spikes (excluding awns) at physiological maturity from four random spots in each plot and were then averaged. YLD (ton ha^{-1}) and TW (kg m^{-3}) data were collected from each plot, while seed weight was adjusted to 13 % moisture. Two adjacent rows of 50 cm each were cut and using for measuring NS and NK.

2.3.5   Statistical analyses of field data

For data analysis we first used lme4 (Bates et al., 2014) package in R v3.6.1 environment (R Core Team 2013) as done by (He et al., 2016; Huang et al., 2016), and fitted a mixed model by taking genotypes as fixed effect and incomplete blocks as random effects per each experiment. Then, lsmeans per trait and sets were estimated. In another follow-up analysis, both experiments and lsmeans of lines were taken as random effect to extract variance components for estimation of broad sense heritability (H^2) on entry mean basis. In this stage, the following model was fitted:

$$Y_{ij} = \mu + Line_i + Set_j + \epsilon_{ij} \qquad [1]$$

where Y_{ij} is the observed phenotype, μ is the overall mean, $Line_i$ is the random effect of the ith line, Set_j is the random effect of jth replicate, and ϵ_{ij} is the random error term. Then, using the variance components estimated form equation [1], broad sense heritability (H^2) on entry mean basis was estimated as follows:

$$H^2 = \frac{\sigma^2 g}{\sigma^2 g + \sigma^2 e / r} \qquad [2]$$

where $\sigma^2 g$ is the genotypic variance and $\sigma^2 e$ is the error variance, and r is the number of replicates from which the traits were evaluated. Best Linear Unbiased Estimates (BLUEs) of each line was estimated by setting genotype as fixed effects in equation [1]. Because the experimental design was incomplete, we also estimated heritability by the method described by Cullis et al., (2006) and suggested by Piepho and Möhring (2007) for unbalanced data using the following equation:

$$H_C^2 = 1 - \frac{\bar{v}_{BLUP}}{2\sigma^2 g} \qquad [3]$$

, where $\bar{v}_{BLUP}$ is the mean variance of all pairwise difference of BLUPs.

2.3.6 Genome-wide association studies and genomic prediction

GWAS was performed using FarmCPU package (Liu et al., 2016) in R v 3.6.1. The first four principal components were used to control the population structure and the markers were filtered to have minor allele frequency of more than or equal to 0.05. Student's t-test was used to test the significance of major loci for each trait. That is the statistical comparison between the means of two groups of homozygous individuals in the population.

In GWAS procedure for any given trait, we included those major genes that showed significant effect in t-test as fixed effect covariate. In this manuscript, we listed MTAs that were identified by the threshold of $-logP \geq 4$. The MTAs located on the same chromosome were resolved into unique regions by estimating pairwise LD between pairs of markers and picking only one representative markers from groups that had $LD \geq 0.60$. The effect of each MTA was estimated as the difference between the means of two groups of homozygous individuals for that marker in the population. The coefficient of determination (R^2) of the markers was estimated by fitting an ordinary least square regression using phenotype as the response variable and the marker as independent variable.

For grain yield only, we also performed a five-fold cross validation of genomic prediction using GBLUP model for 500 cycles in rrBLUP package (Endelman 2011) in R to evaluate the accuracy. Accuracy of genomic prediction was evaluated as correlation between the genomic-estimated breeding value and BLUE divided by the square root of heritability.

2.4 Results and Discussion

2.4.1 Marker development

Next generation sequencing of reduced libraries resulted in > 696 million reads for the 436 germplasm. The reads were merged into 1,278,152 unique tags by TASSEL5 GBS v2 pipeline. Of 1,278,152 unique tags, 1,157,312 (90.55%) tags were mapped to the IWGSC Refseq v1.0 assembly (Appels et al., 2018) while the rest of them were not mapped. After variant calling, 169,486 SNPs were identified. Only 14,907 SNPs were retained after filtering for MAF and missing data.

2.4.2 Sub-genome B revealed the greatest genetic polymorphism

The B sub-genome revealed the greatest number of SNPs 6,546 (44%), whereas sub-genomes A and D showed 1270 (37%) and 1,015 (17%), SNPs (Appendix A1). Similar relative marker numbers were observed in other studies. For example, the majority of ~ 47,000 SNPs identified using the 90K SNP chip array in wheat reported by Wang et al (2014) was located on the B sub-genome (50%), while and the D sub-genome revealed the least number of SNPs (15%). Similarly, in a wheat re-sequencing project conducted by Rimbert et al (2018) using 8 wheat lines with diverse origins, 49% of the total 3.3 million SNPs were located on the B sub-genome while only 10% located on the D sub-genome. In our study, the three chromosomes with the largest numbers of polymorphism were 2B with 1,611 (25% of the B sub-genome), 7A with 1,270 (23% of the A sub-genome), and 7D with 1,015 (41% of the D sub-genome) SNPs (Appendix A1).

2.4.3 The Linkage disequilibrium extent and decay varied among chromosomes

Across the entire genome, the LD decayed in half from the initial value of r2 = 0.23 over a physical range of 8.79 million base pairs (Mbp). However, chromosomes varied considerably in their LD decay behavior. For example, the fastest LD decay was observed within 642 Kbp for chromosome 4D. Substantially slower decay rates were observed within ranges 763 Mbp for chromosome 2B and 94.3 Mbp for chromosome 7D (Appendix A2). The highest number of SNPs and extended LD blocks, shown by LD heatmaps in Appendix A3, on chromosome 2B and 7D are potentially due to alien introgression in these chromosomes from wild relatives of wheat.

Chromosome 2B harbors an alien introgression from *Triticum timopheevi*, which is present in ~40% of the individuals in this population. This region harbors the stem rust resistance gene *Sr36* (Allards and Shands 1954; Tsilo et al., 2008). Jin and Singh (2006) evaluated the stem rust response in the US wheat cultivars and postulated that the resistance in soft winter wheat was primarily due to *Sr36* gene. The pairwise LD between all markers from chromosome 2B and KASP marker for *Sr36* (Tsilo et al., 2008) revealed that out of 1,611 SNPs on chromosome 2B, 864 located within a ~654 Mbp (from 90.4 to 744.3 Mbp) region, showed high ($r^2 > 0.7$) LD with the KASP marker representing Sr36. In contrast, the average pairwise LD of the *Sr36* KASP marker with SNPs outside this region was as low as $r^2 = 0.10$, suggesting the extended LD in 2B is due to this introgression.

Chromosome 7D harbors potentially three different high-LD islands (Appendix A3), where one of these regions is pericentromeric and the other two are located on the opposite arms. The breeding history of this population shows alien transfer of the Barley Yellow Dwarf Virus resistance gene (known as *Bdv3*) from tall wheatgrass *Thinopyrum intermedium* (Sharma et al., 1995), possibly residing the distal end of the long arm of chromosome 7D (Castra et al 2000). Genotyping for *Bdv2/3* showed that 197 and 226 individuals were homozygous for positive and negative alleles, while 13 individuals were heterozygous. Out of 1,015 SNPs on chromosome 7D, 245 covering a ~106 Mbp genomic region on the distal end of 7D showed strong LD ($r^2 > 0.73$) with *Bdv2/3* markers (Appendix A3), suggesting that this high-LD island harbors the *Bdv2/3* locus. The remaining two islands could be associated with two separate alien introgressions. A region of 7E chromosome from *Thinopyrum intermedium*, a wild relative of wheat, which consists FHB resistance locus Qfhs.pur-7EL (or so-called *Fhb7*) was transferred to hexaploid wheat (Shen and Ohm 2007, Guo et al., 2015). Similarly, another alien introgression in chromosome 7D is from wheat relative *Thinopyrum ponticum* (Bariana et al., 2007) and present in 'Wheatear', an Australian spring wheat variety, that consists of stem rust and leaf rust resistance gene block *Sr25/Lr19*.

2.4.4   Population Structure and pedigree analysis

For STRUCTURE analysis, the maximum delta K value reached at K=4, indicating that a four-subpopulation model fits best to the genotypic data. A graphical display of the ancestry

estimates for each accession in each subpopulation is shown in Fig. 2.1A. The interpretation was performed following Evanno et al., (2005). A membership coefficient > 0.5 was used to assign lines to a specific subpopulation. Individuals that did not have membership coefficient > 0.5 for any of the subpopulations were labeled as 'admixed'. The pedigree of lines with 99.9% ancestry coefficient in each subpopulation was retrieved to assign the subpopulation to the most widely known or used genetic background. The four subpopulations identified were named after the widely known germplasm in the group.
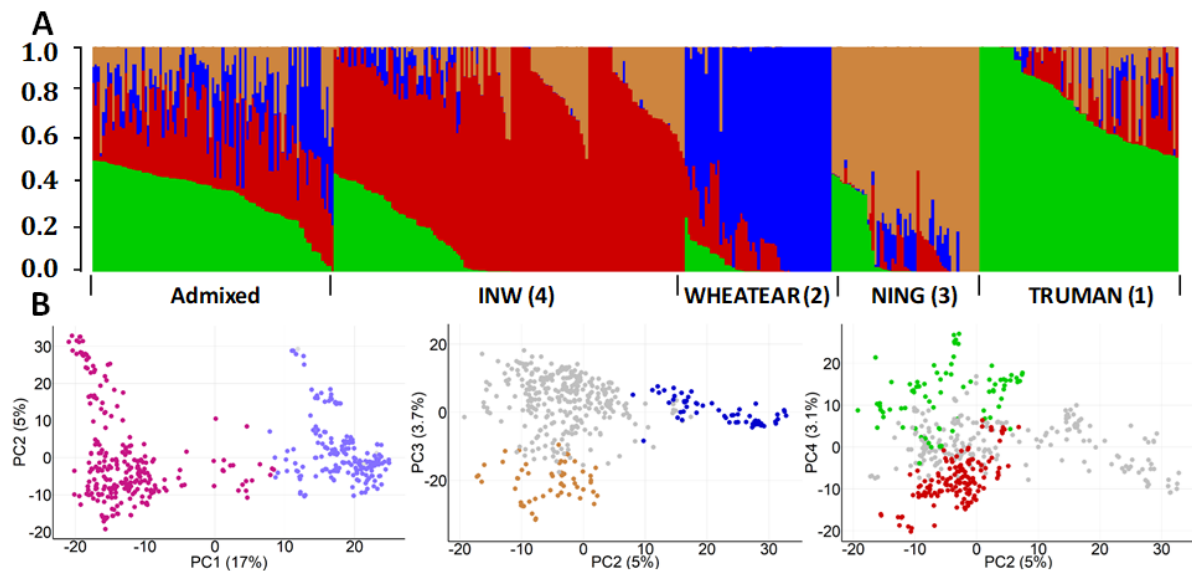
Figure 2.1. Population structure inferred by STRUCTURE software for k = 4 and principal component method. The stacked bar plot (A) shows ancestry estimates for each accession in each of the four subpopulations and in the admixture. (B) shows the scatterplot of principal components (PC). The scatterplot on the right shows the separation of germplasm based on 2B:2G introgression. The pink color represents individuals without the 2B:2G translocation and the blue color shows individuals with the 2B:2G translocation. The two scatter plots on middle and right show the relationship between four principal components and sub-populations inferred from STRUCTURE. The color of these scatterplot was matched the color of sub-populations derived from the STRUCTURE analysis for ease of comparison.

In subpopulation 1 (80 entries), eight entries showed 99.9% ancestry coefficient. Three of these eight lines were derived from a cross designated as "04606RA1". Another three lines were derived from a cross designated as "0527A1". The lines from "04606RA1" are derivatives of "TRUMAN" when it was crossed with "961341A3-1-4-6". The lines with "0527A1" pedigree are progeny of cross "04617A1" x "04688A1" lines. Similarly, the remaining two lines are "07469A1-6-1-1" and "TRUMAN". This subpopulation was named TRUMAN. Truman is a soft red winter

wheat variety developed at Missouri Agricultural Experiment Station and has been used as a native source for FHB resistance for several years (McKendry et al., 2005; Peterson et al., 2016). In Subpopulation 2 (59 entries), 16 entries are from the cross "10644C1" ("10143" x "P25R62") that showed 99.9% ancestry coefficient. The parent line "10143" itself, is a progeny of "07469A1-10-1" x "Wheatear". This subpopulation was named WHEATEAR. Wheatear is an Australian wheat variety harboring the *Sr25/Lr19* resistance gene block from *Thinopyrum ponticum* (Bariana et al., 2007; Zhang et al., 2005). KASP assay for vrn1 gene for all 5 homeologous group chromosomes showed that 44 individuals (~76%) from this group possessed at least one spring type allele suggesting that this group is segregating for spring growth habit alleles possibly inherited from Wheatear. Ten out of 59 entries in subpopulation 3 showed 99.9% ancestry coefficient. Of the 10 lines, four lines are progeny of a cross designated as "0527A1". Two lines are progeny of a cross designated as "0570A1", while other lines are progeny of four other crosses designated as "04704A1", "0537A1", "0858A1", "0925A1". According to their pedigrees, all these lines have two common ancestors "Ning-7840" and "Clark" in their recent breeding history. This subpopulation was named NING. Ning-7840 is a Chinese hard red cultivar that has been used as an exotic source of FHB resistance in breeding programs (Drake 2004). Nine lines out of 141 entries in subpopulation 4 showed 99.9% ancestry coefficient. All the lines are from a cross designated as "HORM45". They are double haploid progeny produced from a cross between "INW0411" and "INW0412". This subpopulation was named INW. INW0411 and INW0412 are soft red winter wheat cultivars developed by Purdue University. The cultivar "INW0411" has excellent milling properties and medium gluten strength. "INW0411" is the progeny of "96204A1-12" x "9723RE" cross. The parents of "96204A1-12" are "FREEDOM" and "92829A1-1-5-1-4-6". The parents of "9723RE" are "Goldfield" and "INW9824". The cultivar "INW0412", is known for its high normalized test weight and medium strong gluten and a progeny of "ACC3130_HYAPEI57-2 x "PATTERSON" cross.

In the genomic population tree, (Appendix A4), TRUMAN, INW, and admixtures clustered together in a monophyletic group, indicating these population are relatively closer to each other, probably due to the adaptation of germplasm to the US soft red winter wheat growing regions. STRUCTURE revealed that TRUMAN and INW sources dominated the ancestry coefficient in the ADMIXTURE population, and that is why the admixed subpopulation is immediately grouped to the TRUMAN and INW subpopulations (Fig. 2.1A). WHEATEAR and NING subpopulations

are distant from each other and distant from the other subpopulations, possibly because of their country of origins.

The majority of genotypic variations was explained by only the first four principal components, with only minimal marginal increase after that. The first four PCs together explained 28.8% (PC1 = 17%, PC2 = 5%, PC3 = 3.7%, and PC4 = 3.1%) of the total genotypic variation. The first principal component (PC1), explaining 17% of the total variation, distinguishes germplasm are contrasting for presence or absence of the 2B:2G translocation. Based on presence/absence of *Sr36/Pm6* regions, the individuals that carry 2B:2G translocation were almost perfectly separated by PC1 from the individuals that did not (Fig. 2.1B). Similar results were reported in soft red winter wheat grown in Virginia (Ward et al., 2019). We did not see such separation in the population using STRUCTURE. This is partly due to running of PCA on all SNPs while STRUCTURE on only reduced set of SNPs. PC2 and PC3 separated WHEATEAR and NING subpopulations from rest of the germplasm (Fig. 2.1B). Similarly, PC4 separated TRUMAN and INW subpopulations from the rest of the germplasm (Fig. 2.1B).

2.4.5   Whole genome scan for selection sweeps

Using F_{ST} approach and comparing the subpopulations identified via STRUCTURE analysis, 12 regions on eight chromosomes showed SS (Table 2.1, Fig. 2.2). Similarly, using hapFLK statistics, four genomic regions including two on chromosome 2B, one on 3B and one on 5B were identified to be significantly selected (Appendix A5). Between the smoothed F_{ST} and the hapFLK approach, only one significant genomic region on chromosome 2B was overlapping. The regions on chromosome 2B from TRUMAN, NING, and WHEATEAR subpopulations were overlapping (Table 2.1, Fig. 2.2). Based on smoothed F_{ST}, the NING group seems to have the largest region on chromosome 2B, spanning roughly over 669 Mbp, harboring the 2B:2G translocation. TRUMAN and WHEATEAR groups showed subsets of the 2B region identified in the NING group (Table 2.1). In INW group, a second region of 2B was identified that falls ~52Mbp downstream of the 2B:2G translocation. This second region is between 775.3 and 800.9 Mbp, which was also supported by the hapFLK method (Table 2.1).

Table 2.1. Genomic regions detected to be under selection by smoothed F_{ST} and hapFLK statistics. The region start and end were determined by the position of flanking markers that were above the significant threshold.

Chrom.	Region Start	Region End	Peak location	Subpopulation	Number of genes	Associated region	References
1D	411731821	414719517	412457553[a]	TRUMAN	36	*Glu-D1*	Shewry et al., 2003
	12823191	15336900	13972959[a]	INW	23	*LT*	Baga et al., 2006; Fowler et al., 2016
2B	319090	46602138	24851652[b]	Whole Panel	828	Unknown	
	101072502	570883707	146915434[a]	TRUMAN	2673	2B:2G	Tsilo et al., 2008
	53577523	723331990	483664894[a]	NING	3688	2B:2G	Tsilo et al., 2008
	53986622	655597674	259038985[a]	WHEATEAR	3684	2B:2G	Tsilo et al., 2008
	775333503	800920106	779248186[a]	INW	413	Unknown	
	763445978	800920106	776406624[b]	Whole Panel	563	Unknown	
2D	580117352	617097449	584460007[a]	INW	520	Grain protein content	Prasad et al., 2003; Mahjourimajd et al., (2016)

Table 2.1 continued …

3B	6181603	14025949	10170821[b]	Whole Panel	162	*Fhb1*	Su et al., (2019)
	711349262	753951947	711642065[a]	INW	421	Grain Number	Sukumaran et al., 2018
4B	535092535	553506651	541365659[a]	INW	107	TKW	Lozada et al., 2018
5A	503873008	552747815	550496370[a]	TRUMAN	546	*FR-A2*	Vágújfalvi et al., 2005; Miller et al 2006
	583031341	595158840	591978388[a]	INW	171	*Vrn-A1*	Yan et al., 2003
5B	550910513	585008944	556185926[b]	Whole Panel	368	*Vrn-B1*	Kato et al., 2001; Yan et al., 2003
7A	149426415	206581766	192615879[a]	INW	425	*Vrn3*	Yan et al., 2006

[a] The region was identified as significantly selected using smooth F_{ST} method.

[b] The region was identified as significantly selected using haFLK method.
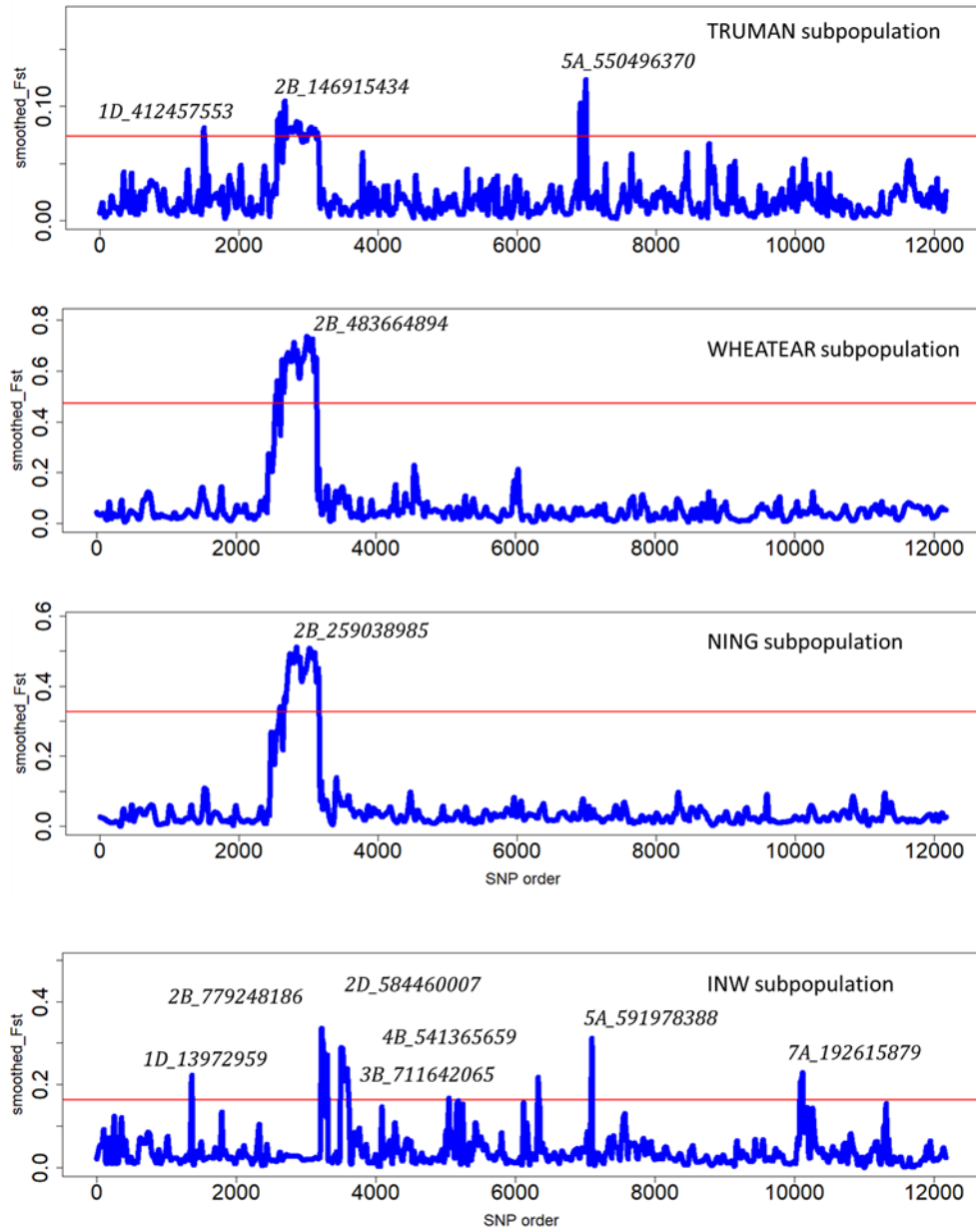
Figure 2.2. Whole genome scans for positive selection by using F_{ST} method. The x-axis shows the physical position of the variants. The horizontal red lines represent the threshold (average plus three orders of the standard deviation) for declaring significance.

Regions on chromosome 1D and 5A were identified in TRUMAN and INW groups (Table 2.1). In TRUMAN group, the region on chromosome 1D spans from 411.7 to 414.3 Mbp, while the region in INW group covers from 12.8 to 15.3 Mbp (Table 2.1), indicating that these are two distinct regions physically located ~398 Mbp apart. The significantly selected regions on chromosome 5A in TRUMAN and INW groups are 30 Mbp apart (Table 2.1, Fig. 2.2). In TRUMAN group, the 5A selected region spans from 503.9 to 552.7 Mbp whereas in INW group, it ranges from 583 to 595.2 Mbp. Other genomic regions on chromosomes 2D (580.1 to 617.1 Mbp), 3B (711.3 to 754 Mbp), and 4B (535.1 to 553.5 Mbp) were also found only in INW group (Table 2.1, Fig. 2.2).

The availability of wheat reference genome IWGSC RefSeq V1.0 (Appels et al., 2018) enabled us to retrieve the candidate genes within each region. Most of the regions identified in this study ranged from 2.5 Mb (on 1D) to 669 Mbp (on 2B), including from 23 genes to 3688 genes. Therefore, pinpointing the genes under selection is speculative. We searched the literatures to find QTLs with meaningful agronomic implications around those regions as existing support however, further evidences are required to validate these results

Two evidence suggest that the 1D region (12.8 to 15.3 Mbp) identified in INW group could be involved in low-temperature tolerance (LT). Baga et al., (2006) identified two QTL on chromosomes 5A and 1D for LT using "Norstar" (high LT) x "Manitou" (low LT) population. In the region identified in this study on 1D between microsatellite markers BARC159 and BARC169, covered from 12,203,610 bp to 328,441,452 bp is located within the region we identified. Moreover, the wheat gene Transducin/WD40 repeat protein (*TraesCS1D01G034200*) located in the region shows 79.6% identity to Arabidopsis *HOS15* (NCBI refseq: NP_178182.1), which is also a WD40 repeat protein and has been shown to be important for cold tolerance where hos15 mutant Arabidopsis plants were hypersensitive to freezing (Zhu et al., 2008). The 1D region (411.7 to 414.3 Mbp) of TRUMAN group consists of 36 genes, which includes a high molecular weight glutenin subunit *Glu-D1* gene controlling gluten quality of the wheat (Anjum et al., 2007). The 2D region (580.1 to 617.1 Mbp) of INW group consists of 520 genes. Near this region, two grain protein content QTL were reported. Prasad et al (2003) mapped QTL for grain protein content linked to wms1264 located from 577,440,661 to 577,440,823bp on chromosome 2D. Another grain

protein QTL was found between markers located from Ex_c2115_3369 (435,045,172 bp) and Kukri_c9145_1322 (571,058,395 bp) on chromosome 2D (Mahjourimajd et al 2016).

The region identified by hapFLK on chromosome 3B is most likely selected for *Fusarium* head blight resistance because it includes *Fusarium* head blight resistance locus *Fhb1*. The *Fhb1* locus from Sumai 3 cultivar is the most widely used source of FHB resistance (Bai et al., 1999; Waldron et al., 1999). Su et al., (2019) performed fine mapping of *Fhb1* locus and showed that *TaHRC*, a gene that encodes a putative histidine-rich calcium-binding protein is responsible for FHB resistance. The *TaHRC* gene (GenBan ID: CBH32655.1) is located from 8,526,628 to 8,529,572 bp, which resides inside 6.2-14 Mbp region identified in this study.

Four regions were identified on chromosomes 5A, 5B, and 7A with implication in vernalization requirements and freezing tolerance. Vernalization and freezing tolerance are controlled by several loci i.e., *Vrn1* (Yan et al., 2003), *Vrn2* (Yan et al., 2004), and *Vrn3* (Yan et al., 2006) and *Fr-A2* (Vágújfalvi et al., 2005). The *Fr-A2* locus is located 30 cM proximal to *Vrn-A1* and reported to be consisting of a cluster of eleven or more C- repeat Binding Factors (*CBF*) genes (Vágújfalvi et al., 2005; Miller et al., 2006). The *CBF* genes are member of *AP2/EREBP* transcription factor family that are essential for freezing tolerance (Thomashow et al., 2001). We retrieved gene sequences of *Vrn-A1* (GenBank ID: GQ451810.1) and *Vrn-B1* (GenBank ID: HQ130483.2) from NCBI and BLAST them against IWGSC Refseq v1.0 genome. The *Vrn-A1* and *Vrn-B1* genes were located 587,423,705 - 587,423,039 bp and 698,178,816 - 698,179,052 bp on chromosomes 5A and 5B respectively. In addition, we located 17 copies of *CBF* genes in the IWGSC V1.0 reference genome from 522,066,602 to 523,616,451 bp on chromosome 5A that is known to be associated with *Fr-A2* locus.

The 5A region (503.9-552.7 Mbp) of TRUMAN group includes the 17 copies of CBF genes. Similarly, the 5A region of INW group spans from 583 to 595.2 Mbp that includes Vrn-A1 gene. The 7A region (149.4 to 206.6 Mbp) consists of *FLOWERING LOCUS T/ TERMINAL FLOWER 1-like* protein (TraesCS7A01G229400, 199,671,796-199,673,399 bp), a gene that has homology to the *Vrn-B3* candidate gene (Yan et al., 2006). This finding suggests that the individuals in TRUMAN and INW group were potentially selected for freezing tolerance.

This study showed that of the 13 significantly selected regions, 6 loci (*LT*, *Ppd-B1*, *Fr-A2*, *Vrn-A1*, *Vrn-B1*, and *Vrn3*) are involved in environmental adaptation such as flowering time and winter hardiness. Other loci were associated with disease resistance such as *Sr36* and *Fhb1* and end-use quality traits such as *Glu-D1* and *GPC*. Regions associated with vernalization also showed SS in previous studies. Cavanagh et al (2013) used genome wide markers to identify genomic regions selected during domestication and post-domestication. They observed genetic differentiation for *Vrn-A1*, *Vrn-B1* loci but not for photoperiod regulation genes. The evidence of selection on the region containing *Vrn-A1* gene was also reported by Gao et al. (2017) in comparison of wild relatives, landraces and modern cultivars of wheat.

2.4.6 Trait correlations and heritability

The trait distributions, mean, and range are provided in Appendix A6 and Appendix A7. The germplasm headed as early as 131 days to 147 days (after January 1). The average number of spikes was 711 spikes m^{-2}, with a maximum of 1081 spikes m^{-2}. Grain yield ranged from 1 to 9, and averaged ~4.4 tons ha^{-1}. A negative correlation between DH and YLD was observed (r=-0.38), along with DH negatively correlated with the yield related traits of TW, NS and NK (r=-0.44, -0.23, -0.14 respectively). This could be due to recurrent selection for early flowering and high yielding cultivars in the past. PH showed minimal correlation with DH, NS, and NK, suggesting that these can improved in an independent manner. There was meager positive correlation between YLD and PH (r = 0.27). All the four yield related traits were positively correlated with each other as expected. TW was highly correlated with NS and YLD (0.63 and 0.64 respectively). For the rest of the trait-combinations, the coefficient of correlation ranged from 0.19 between YLD and NS to 0.35 between YLD and NK.

Broad-sense heritability based on entry mean (H^2) was moderate for NK (0.33), NS (0.42), and YLD (0.53) but relatively higher for TW (0.64), PH (0.83) and DH (0.72) (Appendix A6). Heritability based on Cullis et al., (2006) method (H^2c) was consistently lower than H^2 estimates although trend was similar (Appendix A6). Schmidt et al., (2019) showed that heritability based on entry mean method overestimates in comparison to heritability estimates using method described by Cullis et al., (2006) especially in unbalanced agriculture trails similar to this study.

2.4.7 The diversity and effects of major genes on agronomic traits

Rht-B1b allele that reduces height showed the highest allele frequency of 0.89. The *Rht-B1* gene was highly significant for PH and the *Rht-B1b* allele reduced the plant height by ~10 cm (Table 2.2). In addition, it decreased grain yield by -0.34 tons at p-value 0.048. Conversely, Hayat et al., (2019) reported a 5% yield-increasing effect for the *Rht-B1b* allele in an SRW wheat population. Only four individuals were homozygous for *Rht-D1b* allele, and only two lines were double mutants at *Rht1* loci. We did not evaluate allelic effects for these due to low frequency.

Table 2.2. T-test significance level and marker effects for major loci on agronomic traits. Allelic effects are estimated as the difference between mean of homozygous individuals with mutant and wild allele for each major locus. Numbers below the major loci names represent the number of homozygous lines with mutant and wild allele respectively.

Traits	Units	*Rht-B1*	*Ppd-A1*	*Ppd-B1*	*Ppd-D1*	*vrn-A1*	*1RS.1BL*	*Sr36/Pm6*
		(341:35)	(171:188)	(94:298)	(241:133)	(113:249)	(133:232)	(151:238)
DH	Julian days	-0.1	-0.33	-0.32	-0.83***	0.27	0.8***	-0.27
PH	Cm	-9.69***	-2.65**	-3.44***	-2.03*	-0.21	-1.28	-1.7*
TW	Kg m^{-3}	-12.84	4.04	22.25	-29.03*	-28.76*	-42.4***	21.25
YLD	Ton ha^{-1}	-0.34*	-0.1	0.07	-0.30*	-0.16	-0.19	0.16
NS	Spikes m^{-2}	-31.71	13.93	13.67	3.43	-34.17*	-38.44**	32.67**
NK	Kernels m^{-2}	-658.72	-1268.64*	795.53	-48.35	749.48	-317.62	1.95

* Allelic effect is significant at the 0.05 level; ** Allelic effect is significant at the 0.01 level; *** Allelic effect is significant at the 0.001 level
DH: Days to Heading; PH: plant height; TW= Test Weight; YLD:

The photoperiod insensitive allele *Ppd-D1a* was present in 64% of the population and had an accelerating effect on days to heading (by ~ 20 hours) and other traits (Table 2.2). The *Ppd-B1a* and *Ppd-A1a* photoperiod insensitive alleles, present in 46% and 24% of the population, respectively, did not show any significant effect of days to flowering.

The frequency of the rye translocation *1RS.1BL* was ~39%. We observed a significant negative effect from 1RS.1BL translocation on DH, TW, and NS, but not YLD (Table 2.2). McKendry et al., (1996) did not observe any significant yield advantage while heading was

delayed in lines with the translocation. Similarly, Zhao et al., (2012) reported significant reduction of TW and kernel weight in lines possessing the translocation.

The frequency of the stem rust resistance gene *Sr36* was ~37%. The presence of *Sr36* gene significantly reduced plant height and increased NS (Table 2.2). Ward et al., (2019) reported significant reduction in plant height in lines with *Sr36* genes in an SRW wheat population grown in Virginia. However, no significant effect was observed for NS by Ward et al., (2019).

2.4.8 GWAS for major agronomic traits

Association of genome-wide markers with six agronomic traits was analyzed by using FarmCPU model. For any GWAS analysis, we included major genes as fixed effect covariates to account for their effects and to avoid the association of the SNP markers that are correlated with the major genes. *Ppd-D1* was significant for most of the traits: DH, PH, TW, and TKW whereas *Ppd-B1* was only significant for PH (Table 2.2). Thirty-five marker-trait associations (MTAs) were identified at *-logP* > 4 across 13 chromosomes of wheat. After pairwise LD analysis, 35 MTAs resolved into 27 unique genomic regions. Seven of the 27 regions passed the 5% FDR threshold (Table 2.3). In addition, any MTA with $R^2 \geq 5\%$ was considered as major effect MTA.

Ppd-B1 and *1RS.1BL* translocation were used as fixed effect covariate for GWAS analysis of DH (Table 2.2). Six independent regions were identified on 4A, 5A, 6A, 6D, 7B, and 7D for DH, of which only the variant on chromosome 7B (7B_8885441) was identified at 5% FDR threshold (Table 2.3, Fig. 2.3). This variant explained the highest amount of variation (11.5%) for DH, where the early heading allele (T) reduced DH by more than two days (Table 2.3). Chromosome 7B harbors *Vrn-B3* locus that plays an important role in vernalization requirement and controlling flowering (Yan et al., 2006). The gene resides within 9,702,541-9,703,732bp and is approximately 800 kb away from the MTA and well within the average LD decay rate (dropping to half-LD within ~3 Mbp) observed in chromosome 7B suggesting the region is associated with *Vrn-3* gene.

Table 2.3 Summary of significant marker-trait associations identified by FarmCPU algorithm.

Traits	Locus*	Variants[a]	-log(p)	MAF	R^2 (%)	Effect
DH	4A_495327498	T/<u>A</u> (276:101)	4.5	0.28	2.3	0.64
(Julian days)	5A_5678539	<u>G</u>/A (307:80)	5.3	0.21	0.3	0.28
	6A_20163710	A/<u>C</u> (214:161)	4.3	0.43	0.2	0.14
	6D_1572162	<u>G</u>/A (351:39)	4.1	0.1	0.6	0.47
	7B_8885441*	C/<u>T</u> (356:34)	11.4	0.09	11.5	2.18
	7D_71569531	<u>G</u>/A (246:128)	4.6	0.35	0.4	0.25
PH (cm)	1A_494799791*	<u>A</u>/C (364:23)	5.4	0.07	3.3	5.68
	3B_6981619	C/<u>T</u> (245:139)	5.1	0.36	0.5	1.23
	4A_41973900*	<u>C</u>/G (331:48)	7.3	0.14	3.8	3.78
	5B_455980623*	A/<u>G</u> (299:81)	5.5	0.22	2.3	2.94
	6A_27020312	<u>G</u>/T (358:7)	5.1	0.05	3.8	7.55
	6A_420391165*	C/<u>G</u> (270:111)	7.9	0.3	8.1	4.96
	6B_155468149	<u>G</u>/A (360:32)	5.3	0.08	2.6	4.55
	6B_603911083	<u>C</u>/T (197:175)	4.4	0.47	1.8	2.11
	7B_714780111	G/<u>A</u> (331:57)	5.1	0.15	3.4	4.0
	7D_4501942	C/<u>T</u> (353:18)	5.3	0.07	9.2	10.34
TW (kgm^{-3})	2A_719404657	<u>C</u>/T (329:55)	4.9	0.15	0.6	26.48
	3B_487539976	A/<u>C</u> (189:183)	4.2	0.49	1.7	29.57
	4B_3295544*	<u>C</u>/T (336:20)	5.1	0.09	0.005	6.23
	4B_573294337	C/<u>T</u> (342:39)	6.2	0.11	4.4	79.11
	4B_612333025	<u>T</u>/A (307:68)	5.4	0.2	0.5	16.59
	5A_47458032	<u>C</u>/T (211:161)	4.7	0.44	0.2	10.46
NS (Spikes m^{-2})	1A_549422405*	A/<u>G</u> (368:22)	5.8	0.06	6.2	130.32
	1D_443717338	A/<u>G</u> (344:6)	4.9	0.07	6.9	93.81
	3B_419241935	<u>G</u>/A (198:187)	4.3	0.49	3.9	48.34
	5B_679675578	T/<u>C</u> (309:71)	5	0.2	2.5	48.26
NK (Kernels m^{-2})	5A_456261322	<u>C</u>/T (348:39)	4.1	0.11	12.4	5180.78

DH: days to heading, PH: plant height, TW: test weight, YLD: yield, NS: number of spikes per square meter, NK: number of kernels per square meter

*Locus that were significant at 5% FDR.

[a]Nucleotide underlined represents the favorable allele. For DH and PH, allele reducing them are considered favorable whereas for TW, NS, and NK allele increasing them are considered favorable.
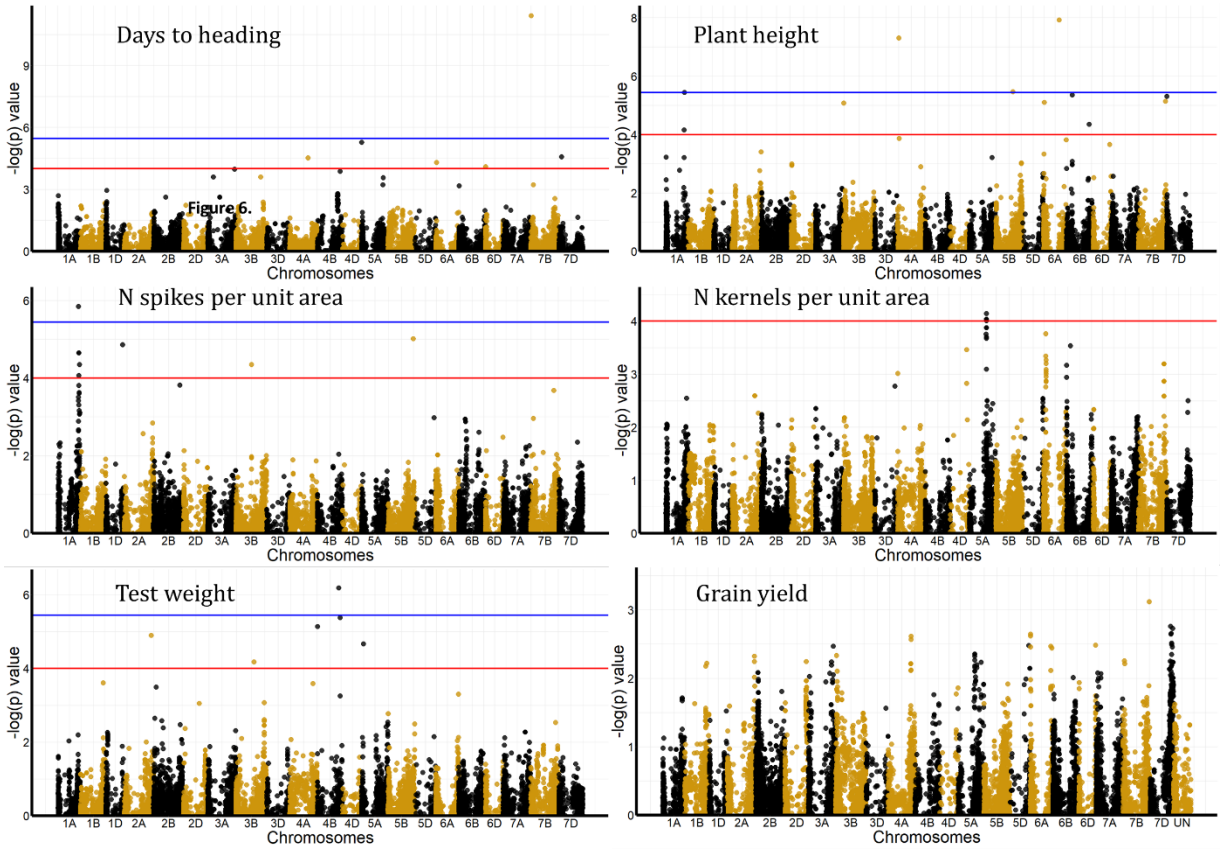
Figure 2.3. Manhattan plots of GWAS for days to heading (DH), plant height (PH), number of spikes per square meter (NS), and number of kernels per square meter (NK). The red and blue horizontal lines correspond to the *-logP* value of 4.0 and 5% FDR threshold, respectively.

The *Ppd-A1*, *Ppd-B1*, *Ppd-D1*, *Rht-B1*, and *Sr36/Pm6* loci were significant for PH and were used as fixed effect covariates in the GWAS model for PH (Table 2.2). Ten genomic regions were identified at *-logP* > 4 on chromosomes 1A, 3B, 4A, 5B, 7B, 7D, 6A, and 6B of which four MTAs i.e., 1A_494799791, 4A_41973900, 5B_455980623, and 6A_420391165 were identified at 5% FDR threshold (Table 2.3, Fig. 2.3). The variants 7D_4501942 and 6A_420391165 explained 9.2% and 8.1% of variation in PH, with height reducing effects of 10 and 5 cm, respectively (Table 2.3). The variant 6A_420391165 resides in the genomic region of *Rht24* that was previously identified by Spielmeyer et al., (2007) and further fine mapped by Li et al. (2015) between Xwmc256 and Xbarc103 markers (176,504,290 and 549,523,882 bp). In addition to 7D_4501942 and 6A_420391165, four other variants 1A_494799791, 4A_41973900, 6A_27020312, and 7B_714780111 explained ~3% variation in PH variation and marker effects were 3.8 cm for 4A_41973900 and 7.6 cm for 6A_27020312 (Table 2.3).

For TW, *Ppd-D1*, *Vrn-A1*, and 1RS.1BL were significant (Table 2.2). Using them as fixed effect covariate, six genomic regions were identified at *–logP* > 4 for TW. There were three independent regions on chromosome 4B and one each of chromosomes 2A, 3B, and 5A (Table 2.3, Fig. 2.3). The variant 4B_573294337 that explained 4.4% variation of TW, with effect size of 79 kg/m^3, was identified at 5% FDR threshold. Other studies have found QTL controlling TW (*QTW.ndsu.4B.2*) using a bi-parental population (Kumar et al., (2015) which is only 3 Mbp away from the MTA identified in this study. The other MTA explained a very small percentage of variation in TW.

For YLD no marker trait association was identified, highlighting that larger effect QTL are not segregating within the population and in addition, the study is underpowered to detect minor effect QTL. In this situation, genomic selection can be effective to achieve higher genetic gains. In this study, when did a five-fold cross validation of YLD over 500 repetitions, on average, prediction accuracy of 0.63 was achieved suggesting genomic selection could be an alternative strategy to MAS to improve yield.

For NS, major genes *Vrn-A1*, 1RS.1BL, and *Sr36/Pm6* were used as fixed effect covariates. Four regions on chromosomes 1A, 1D, 3B, and 4B were identified for NS (Table 2.3, Fig. 2.3). The favorable allele for variant 1A_549422405, identified at 5% FDR threshold, showed highest

effect, increasing NS by 130 spikes/m^2 and explaining 6.2% variation of NS (Table 2.3). A QTL responsible for number of kernels and spikes per square meter was recently identified 14 Mbp away from the region we identified. This QTL was identified in Weebill x Bacanora DH recombinant population and linked to the marker BS00009700 at physical position of 535,062,221-535,062,316bp (Griffiths et al., 2015). Similarly, the variant marker 1D_443717338 explained ~7% variation in NS with the favorable allele increasing NS by 94 spikes/m^2. The variants 3B_419241935 and 5B_679675578 explained 2.5% and 3.9% of variation in NS, respectively. In both cases, the marker effect was ~48 spikes/m^2.

After accounting for large effect of *Ppd-A1* in GWAS for NK, only one region on chromosome 5A was identified for NK at *-logP* > 4 (Table 2.3, Fig. 2.3). The variant 5A_456261322 explained 12.4% variation in NK and the favorable allele is estimated to increase NK by 5,180 kernels/m^2 (Table 2.3). Zhang et al. (2018) identified a significant MTA IWB1040 (437208847 bp on 5A) associated with grain weight in a population of spring wheats. Similarly, in a Pacific Northwest Winter Wheat population, Gizaw et al. (2018) identified marker IWA 3099 (48171543 bp) associated with thousand kernel weight (TKW).

With a focus on genotypic and phenotypic data, we characterized 436 germplasm for population structure and the extent of LD decay. Two chromosomes 2B and 7D showed extended low recombination regions possibly due to alien chromatins. STRUCTURE and PCA analyses revealed there are four sub-populations in the populations. Pedigree analysis of these sub-population suggested that two subpopulations, WHEATEAR and NING, were most likely result of introduction of exotic lines NING and WHEATEAR into the breeding program for disease resistance. The whole genome was scanned by using smoothed F_{ST} and hapFLK methods to detect SS, where 13 regions with implications in environmental adaptation, disease resistance, and end use quality were identified. Field-based phenotyping allowed for trait evaluation and GWAS analysis. We identified six major QTLs associated with agronomic (PH and DH) and yield-related traits (TW, NK, and NS). We did not find any significant regions for YLD using GWAS, however, prediction of YLD showed that reliable accuracies can be achieved for yield improvement in this population. This study helps to use large effect MTA and associated germplasm as parents in crossing nurseries with the aim to effectively pyramid the beneficial alleles. In addition, detailed characterization of population structure will also guide to optimize breeding populations for

maximized diversity and heterotic patterns. Higher genetic variances can be achieved by crossing individuals from separate sub-populations. For example, line "PU390" from INW sub-population with good YLD (7.6 ton/ha) and early flowering (135 days) and line "PU378" from TRUMAN sub-population with yield 6.4 tons/ha and flowering ~137 days seems to be ideal candidates as parents for making crosses as they come from different sub-populations and show good YLD and early flowering.

2.5 References

Adhikari, A., Steffenson, B.J., Smith, K.P., Smith, M. and Dill-Macky, R. (2020). Identification of quantitative trait loci for net form net blotch resistance in contemporary barley breeding germplasm from the USA using genome-wide association mapping. Theoretical and Applied Genetics, pp.1-19.

Allard, R.W. and Shands, R.G. (1954). Inheritance of resistance to stem rust and powdery mildew in cytologically stable spring wheats derived from *Triticum timopheevi*. Phytopathology, 44(2), pp.266-274.

Anderson, J. A., Stack, R. W., Liu, S., Waldron, B. L., Fjeld, A. D., Coyne, C. et al. (2001). DNA markers for *Fusarium* head blight resistance QTLs in two wheat populations. *Theoretical and Applied Genetics*, *102*(8), 1164-1168.

Anjum, F.M., Khan, M.R., Din, A., Saeed, M., Pasha, I. and Arshad, M.U. (2007). Wheat gluten: high molecular weight glutenin subunits—structure, genetics, and relation to dough elasticity. Journal of food science, 72(3), pp.R56-R63.

Appels, R., Eversole, K., Feuillet, C., Keller, B., Rogers, J., Stein, N., et al. (2018). Shifting the limits in wheat research and breeding using a fully annotated reference genome. Science, 361(6403), p.eaar7191.

Asseng, S., Ewert, F., Martre, P., Rötter, R.P., Lobell, D.B., Cammarano, D. et al. (2015). Rising temperatures reduce global wheat production. Nature climate change, 5(2), p.143.

Båga, M., Chodaparambil, S.V., Limin, A.E., Pecar, M., Fowler, D.B. and Chibbar, R.N. (2007). Identification of quantitative trait loci and associated candidate genes for low-temperature tolerance in cold-hardy winter wheat. Functional & integrative genomics, 7(1), pp.53-68.

Bai, G., Kolb, F.L., Shaner, G. and Domier, L.L. (1999). Amplified fragment length polymorphism markers linked to a major quantitative trait locus controlling scab resistance in wheat. Phytopathology, 89(4), pp.343-348.

Bariana, H.S., Brown, G.N., Bansal, U.K., Miah, H., Standen, G.E. and Lu, M. (2007). Breeding triple rust resistant wheat cultivars for Australia using conventional and marker-assisted selection technologies. Australian Journal of Agricultural Research, 58(6), pp.576-587.

Beissinger, T.M., Hirsch, C.N., Vaillancourt, B., Deshpande, S., Barry, K., Buell, C.R. (2014). A genome-wide scan for evidence of selection in a maize population under long-term artificial selection for ear number. Genetics, 196(3), pp.829-840.

Bonhomme, M., Chevalet, C., Servin, B., Boitard, S., Abdallah, J., Blott, S. and SanCristobal, M. (2010). Detecting selection in population trees: the Lewontin and Krakauer test extended. Genetics, 186(1), pp.241-262.

Bradbury, P.J., Zhang, Z., Kroon, D.E., Casstevens, T.M., Ramdoss, Y. and Buckler, E.S. (2007). TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics, 23(19), pp.2633-2635.

Brito, L. F., Kijas, J. W., Ventura, R. V., Sargolzaei, M., Porto-Neto, L. R., Cánovas, A., et al. (2017). Genetic diversity and signatures of selection in various goat breeds revealed by genome-wide SNP markers. BMC genomics, 18(1), 229.

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

Cavanagh, C.R., Chao, S., Wang, S., Huang, B.E., Stephen, S., Kiani, S., et al. (2013). Genome-wide comparative diversity uncovers multiple targets of selection for improvement in hexaploid wheat landraces and cultivars. Proceedings of the national academy of sciences, 110(20), pp.8057-8062.

Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M. and Lee, J.J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience, 4(1), p.7.

Chen, Y., Carver, B.F., Wang, S., Zhang, F. and Yan, L. (2009). Genetic loci associated with stem elongation and winter dormancy release in wheat. Theoretical and Applied Genetics, 118(5), pp.881-889.

Crasta, O.R., Francki, M.G., Bucholtz, D.B., Sharma, H.C., Zhang, J., Wang, R.C. (2000). Identification and characterization of wheat-wheatgrass translocation lines and localization of barley yellow dwarf virus resistance. Genome, 43(4), pp.698-706.

Cullis, B.R., Smith, A.B. and Coombes, N.E. (2006). On the design of early generation variety trials with correlated data. Journal of agricultural, biological, and environmental statistics, 11(4), p.381.

Drake, D.R. (2004). Identification of DNA markers for *Fusarium* head blight resistance QTL in wheat cultivars Ning 7840, Freedom, and Patton (Doctoral Dissertation, Purdue University). Retrieved from Purdue University Digital Repository (https://www.lib.purdue.edu/).

Evanno, G., Regnaut, S. and Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Molecular ecology, 14(8), pp.2611-2620.

Evans, L.M., Slavov, G.T., Rodgers-Melnick, E., Martin, J., Ranjan, P., Muchero, W., et al. (2014). Population genomics of *Populus trichocarpa* identifies signatures of selection and adaptive trait associations. Nature genetics, 46(10), p.1089.

FAO. 2019. http://www.fao.org/faostat/en/#data/QC (accessed 10 Sept. 2019).

Fariello, M. I., Boitard, S., Naya, H., SanCristobal, M., & Servin, B. (2013). Detecting signatures of selection through haplotype differentiation among hierarchically structured populations. Genetics, 193(3), 929-941.

Fariello, M.I., Servin, B., Tosser-Klopp, G., Rupp, R., Moreno, C., San Cristobal, M., et al. (2014). Selection signatures in worldwide sheep populations. PLoS One, 9(8), p.e103813.

Gage, J.L., White, M.R., Edwards, J.W., Kaeppler, S. and De Leon, N. (2018). Selection signatures underlying dramatic male inflorescence transformation during modern hybrid maize breeding. Genetics, 210(3), pp.1125-1138.

Gaire, R., Huang, M., Sneller, C., Griffey, C., Brown-Guedira, G. and Mohammadi, M. (2019). Association Analysis of Baking and Milling Quality Traits in an Elite Soft Red Winter Wheat Population. Crop Science, 59(3), pp.1085-1094.

Gao, L., Zhao, G., Huang, D. and Jia, J. (2017). Candidate loci involved in domestication and improvement detected by a published 90K wheat SNP array. Scientific reports, 7, p.44530.

Gholami, M., Reimer, C., Erbe, M., Preisinger, R., Weigend, A., Weigend, S., et al. (2015). Genome scan for selection in structured layer chicken populations exploiting linkage disequilibrium information. *PloS one*, *10*(7), p.e0130497.

Gill, B.S., Friebe, B.R. and White, F.F. (2011). Alien introgressions represent a rich source of genes for crop improvement. Proceedings of the National Academy of Sciences, 108(19), pp.7657-7658.

Guo, J., Zhang, X., Hou, Y., Cai, J., Shen, X., Zhou, T., et al. (2015). High-density mapping of the major FHB resistance gene *Fhb7* derived from *Thinopyrum ponticum* and its pyramiding with Fhb1 by marker-assisted selection. Theoretical and applied genetics, 128(11), pp.2301-2316.

Gupta, P.K., Kulwal, P.L. and Jaiswal, V. (2014). Association mapping in crop plants: opportunities and challenges. In Advances in genetics (Vol. 85, pp. 109-147). Academic Press.

Hayat, H., Mason, R. E., Lozada, D. N., Acuna, A., Holder, A., Larkin, D., et al. (2019). Effects of allelic variation at *Rht-B1* and *Rht-D1* on grain yield and agronomic traits of southern US soft red winter wheat. Euphytica, 215(10), 172.

He, S., Schulthess, A.W., Mirdita, V., Zhao, Y., Korzun, V., Bothe, R., et al. (2016). Genomic selection in a commercial winter wheat population. Theoretical and applied genetics, 129(3), pp.641-651.

Herrmann, E. and Mächler, M., 2003. lokern: Kernel regression smoothing with local or global plug-in bandwidth. R package version, pp.1-0.

Hill, W.G. and Weir, B.S., 1988. Variances and covariances of squared linkage disequilibria in finite populations. *Theoretical population biology*, *33*(1), pp.54-78.

Holsinger, K.E. and Weir, B.S., 2009. Genetics in geographically structured populations: defining, estimating and interpreting F_{ST}. Nature Reviews Genetics, 10(9), p.639.

Huang, M., Cabrera, A., Hoffstetter, A., Griffey, C., Van Sanford, D., Costa, J., et al. (2016). Genomic selection for wheat traits and trait stability. Theoretical and Applied Genetics, 129(9), pp.1697-1710.

Jin, F., Zhang, D., Bockus, W., Baenziger, P.S., Carver, B. and Bai, G. (2013). *Fusarium* head blight resistance in US winter wheat cultivars and elite breeding lines. Crop Science, 53(5), pp.2006-2013.

Jin, Y. and Singh, R.P. (2006). Resistance in US wheat to recent eastern African isolates of *Puccinia graminis f. sp. tritici* with virulence to resistance gene *Sr31*. Plant Disease, 90(4), pp.476-480.

Lin, T., Zhu, G., Zhang, J., Xu, X., Yu, Q., Zheng, Z., et al. (2014). Genomic analyses provide insights into the history of tomato breeding. Nature genetics, 46(11), p.1220.

Liu, X., Huang, M., Fan, B., Buckler, E.S. and Zhang, Z. (2016). Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies. PLoS genetics, 12(2).

Lozada, D.N., Mason, R.E., Sukumaran, S. and Dreisigacker, S. (2018). Validation of grain yield QTLs from soft winter wheat using a CIMMYT spring wheat panel. Crop Science, 58(5), pp.1964-1971.

Mahjourimajd, S., Taylor, J., Rengel, Z., Khabaz-Saberi, H., Kuchel, H., Okamoto, M. and Langridge, P. (2016). The genetic control of grain protein content under variable nitrogen supply in an Australian wheat mapping population. PloS one, 11(7).

McKendry, A. L., Tague, D. N., & Miskin, K. E. (1996). Effect of *1BL.1RS* on agronomic performance of soft red winter wheat. Crop science, 36(4), 844-847.

McKendry, A.L., Tague, D.N., Wright, R.L., Tremain, J.A. and Conley, S.P. (2005). Registration of 'Truman' wheat. Crop science, 45(1), p.421.

Miller, A.K., Galiba, G. and Dubcovsky, J. (2006). A cluster of 11 CBF transcription factors is located at the frost tolerance locus Fr-A m 2 in *Triticum monococcum*. Molecular Genetics and Genomics, 275(2), pp.193-203.

Money, D., Gardner, K., Migicovsky, Z., Schwaninger, H., Zhong, G.Y. and Myles, S. (2015). LinkImpute: fast and accurate genotype imputation for nonmodel organisms. G3: Genes, Genomes, Genetics, 5(11), pp.2383-2390.

Muers, M. (2011). Genome-wide association mapping in rice. Nature Reviews Genetics, 12(11), pp.741-741.

Myles, S., Peiffer, J., Brown, P.J., Ersoz, E.S., Zhang, Z., Costich, D.E. and Buckler, E.S. (2009). Association mapping: critical considerations shift from genotyping to experimental design. The Plant Cell, 21(8), pp.2194-2202.

Nagy, E.D., Chu, Y., Guo, Y., Khanal, S., Tang, S., Li, Y. (2010). Recombination is suppressed in an alien introgression in peanut harboring *Rma*, a dominant root-knot nematode resistance gene. Molecular Breeding, 26(2), pp.357-370.

Nicholson, G., Smith, A.V., Jónsson, F., Gústafsson, Ó., Stefánsson, K. and Donnelly, P. (2002). Assessing population differentiation and isolation from single-nucleotide polymorphism data. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 64(4), pp.695-715.

Ohm H., McFee W.W. (2012). Small Grains Breeding at Purdue University. Agronomy, Purdue University. Retrieved from https://ag.purdue.edu/agry/Documents/Small%20Grains%20Breeeding%20at%20Purdue KLS.pdf

Patil, G., Mian, R., Vuong, T., Pantalone, V., Song, Q., Chen, P., et al. (2017). Molecular mapping and genomics of soybean seed protein: a review and perspective for the future. Theoretical and Applied Genetics, 130(10), pp.1975-1991.

Petersen, S., Lyerly, J.H., McKendry, A.L., Islam, M.S., Brown-Guedira, G., Cowger, C. (2017). Validation of *Fusarium* head blight resistance QTL in US winter wheat. Crop Science, 57(1), pp.1-12.

Piepho, H.P. and Möhring, J. (2007). Computing heritability and selection response from unbalanced plant breeding trials. Genetics, 177(3), pp.1881-1888.

Poland, J.A. and Rife, T.W. (2012). Genotyping-by-sequencing for plant breeding and genetics. *The Plant Genome*, *5*(3), pp.92-102.

Porto-Neto, L.R., Lee, S.H., Lee, H.K. and Gondro, C. (2013). Detection of signatures of selection using F ST. In Genome-wide association studies and genomic prediction (pp. 423-436). Humana Press, Totowa, NJ.

Prasad, M., Kumar, N., Kulwal, P., Röder, M., Balyan, H., Dhaliwal, H. and Gupta, P. (2003). QTL analysis for grain protein content using SSR markers and validation studies using NILs in bread wheat. Theoretical and Applied Genetics, 106(4), pp.659-667.

Pritchard, J.K., Wen, W. and Falush, D. (2003). Documentation for STRUCTURE software: Version 2.

Ray, D.K., Mueller, N.D., West, P.C. and Foley, J.A. (2013). Yield trends are insufficient to double global crop production by 2050. PloS one, 8(6), p.e66428.

Remington, D.L., Thornsberry, J.M., Matsuoka, Y., Wilson, L.M., Whitt, S.R., Doebley, J. (2001). Structure of linkage disequilibrium and phenotypic associations in the maize genome. Proceedings of the National Academy of Sciences, 98(20), pp.11479-11484.

Reynolds, J., Weir, B.S. and Cockerham, C.C. (1983). Estimation of the coancestry coefficient: basis for a short-term genetic distance. Genetics, 105(3), pp.767-779.

Rimbert, H., Darrier, B., Navarro, J., Kitt, J., Choulet, F., Leveugle, M., et al. (2018). High throughput SNP discovery and genotyping in hexaploid wheat. PloS one, 13(1), p.e0186329.

Schmidt, P., Hartung, J., Rath, J. and Piepho, H.P. (2019). Estimating Broad-Sense Heritability with Unbalanced Data from Agricultural Cultivar Trials. Crop Science, 59(2), pp.525-536.

Sharma, H., Ohm, H., Goulart, L., Lister, R., Appels, R. and Benlhabib, O. (1995). Introgression and characterization of barley yellow dwarf virus resistance from *Thinopyrum intermedium* into wheat. Genome, 38(2), pp.406-413.

Shen, X. and Ohm, H. (2007). Molecular mapping of *Thinopyrum*-derived *Fusarium* head blight resistance in common wheat. Molecular Breeding, 20(2), pp.131-140.

Su, Z., Bernardo, A., Tian, B., Chen, H., Wang, S., Ma, H., et al. (2019). A deletion mutation in *TaHRC* confers Fhb1 resistance to *Fusarium* head blight in wheat. Nature genetics, p.1.

Sukumaran, S., Lopes, M., Dreisigacker, S. and Reynolds, M. (2018). Genetic analysis of multi-environmental spring wheat trials identifies genomic regions for locus-specific trade-offs for grain weight and grain number. Theoretical and applied genetics, 131(4), pp.985-998.

Thomashow, M.F., Gilmour, S.J., Stockinger, E.J., Jaglo-Ottosen, K.R. and Zarka, D.G. (2001). Role of the Arabidopsis *CBF* transcriptional activators in cold acclimation. Physiologia Plantarum, 112(2), pp.171-175.

Tsilo, T.J., Jin, Y. and Anderson, J.A. (2008). Diagnostic microsatellite markers for the detection of stem rust resistance gene *Sr36* in diverse genetic backgrounds of wheat. Crop Science, 48(1), pp.253-261.

USDA-NASS. (2019). https://www.nass.usda.gov/Statistics_by_Subject/result.php?4F40FE05-D755-38FE-B10C-1E93CD2EA752&sector=CROPS&group=FIELD%20CROPS&comm=WHEAT (accessed 12 Sept, 2019).

Vágújfalvi, A., Aprile, A., Miller, A., Dubcovsky, J., Delugu, G., Galiba, G. and Cattivelli, L. (2005). The expression of several *Cbf* genes at the *Fr-A2* locus is linked to frost resistance in wheat. Molecular Genetics and Genomics, 274(5), pp.506-514.

Vitti, J.J., Grossman, S.R. and Sabeti, P.C. (2013). Detecting natural selection in genomic data. Annual review of genetics, 47, pp.97-120.

Waldron, B.L., Moreno-Sevilla, B., Anderson, J.A., Stack, R.W. and Frohberg, R.C. (1999). RFLP mapping of QTL for *Fusarium* head blight resistance in wheat. Crop Science, 39(3), pp.805-811.

Wang, S., Wong, D., Forrest, K., Allen, A., Chao, S., Huang, B.E., et al. (2014). Characterization of polyploid wheat genomic diversity using a high-density 90 000 single nucleotide polymorphism array. Plant biotechnology journal, 12(6), pp.787-796.

Ward, B.P., Brown-Guedira, G., Kolb, F.L., Van Sanford, D.A., Tyagi, P., Sneller, C.H. and Griffey, C.A. (2019). Genome-wide association studies for yield-related traits in soft red winter wheat grown in Virginia. PloS one, 14(2).

Wei, D., Cui, Y., He, Y., Xiong, Q., Qian, L., Tong, C., et al. (2017). A genome-wide survey with different rapeseed ecotypes uncovers footprints of domestication and breeding. Journal of experimental botany, 68(17), pp.4791-4801.

Xiao, Y., Liu, H., Wu, L., Warburton, M. and Yan, J. (2017). Genome-wide association studies in maize: praise and stargaze. Molecular plant, 10(3), pp.359-374.

Xie, W., Wang, G., Yuan, M., Yao, W., Lyu, K., Zhao, H., et al. (2015). Breeding signatures of rice improvement revealed by a genomic variation map from a large germplasm collection. Proceedings of the National Academy of Sciences, 112(39), pp.E5411-E5419.

Xu, X., Liu, X., Ge, S., Jensen, J.D., Hu, F., Li, X. (2012). Resequencing 50 accessions of cultivated and wild rice yields markers for identifying agronomically important genes. Nature biotechnology, 30(1), p.105.

Yan, L., Fu, D., Li, C., Blechl, A., Tranquilli, G., Bonafede, M., et al. (2006). The wheat and barley vernalization gene *VRN3* is an orthologue of FT. Proceedings of the National Academy of Sciences, 103(51), pp.19581-19586.

Yan, L., Loukoianov, A., Blechl, A., Tranquilli, G., Ramakrishna, W., SanMiguel, P. (2004). The wheat *VRN2* gene is a flowering repressor down-regulated by vernalization. Science, 303(5664), pp.1640-1644.

Yan, L., Loukoianov, A., Tranquilli, G., Helguera, M., Fahima, T. and Dubcovsky, J. (2003). Positional cloning of the wheat vernalization gene *VRN1*. Proceedings of the National Academy of Sciences, 100(10), pp.6263-6268.

Yurchenko, A.A., Daetwyler, H.D., Yudin, N., Schnabel, R.D., Vander Jagt, C.J., Soloshenko, V., et al. (2018). Scans for signatures of selection in Russian cattle breed genomes reveal new candidate genes for environmental adaptation and acclimation. Scientific reports, 8.

Zhang, W., Lukaszewski, A.J., Kolmer, J., Soria, M.A., Goyal, S. and Dubcovsky, J. (2005). Molecular characterization of durum and common wheat recombinant lines carrying leaf rust resistance (*Lr19*) and yellow pigment (Y) genes from *Lophopyrum ponticum*. Theoretical and applied genetics, 111(3), pp.573-582.

Zhao, C., Cui, F., Wang, X., Shan, S., Li, X., Bao, Y., & Wang, H. (2012). Effects of *1BL/1RS* translocation in wheat on agronomic performance and quality characteristics. Field Crops Research, 127, 79-84.

Zhou, Z., Jiang, Y., Wang, Z., Gou, Z., Lyu, J., Li, W., et al. (2015). Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. Nature biotechnology, 33(4), p.408.

Zhu, J., Jeong, J.C., Zhu, Y., Sokolchik, I., Miyazaki, S., Zhu, J.K., et al. (2008). Involvement of *Arabidopsis HOS15* in histone deacetylation and cold tolerance. Proceedings of the National Academy of Sciences, 105(12), pp.4945-4950.

CHAPTER 3.   *FUSARIUM* HEAD BLIGHT RESISTANCE LOCI IDENTIFIED USING GENOME-WIDE ASSOCIATION STUDIES IN SOFT RED WINTER WHEAT

3.1   Abstract

Identification of minor *Fusarium* head blight (FHB) resistance QTL from different sources and pyramiding them into cultivars provides effective protection against FHB. We performed genome-wide association studies in a population of 392 soft red winter wheat individuals by FarmCPU model with four principal components as covariate to account for population structure. Germplasm was evaluated in misted FHB nurseries inoculated with scabby corn across 2017-18 (Y1) and 2018-19 (Y2) seasons at Purdue Agronomy Farm, West Lafayette in randomized incomplete block designs. Phenotypic data included disease incidence (INC), disease severity (SEV), *Fusarium* damaged kernels (FDK), FHB index (FHBdx), and deoxynivalenol concentration (DON). Twenty-five loci were identified at $-logP \geq 4.0$ to be associated with five FHB-related traits. Of these 25, eighteen explained more than 1% of the phenotypic variations. A major QTL on chromosome 2B i.e., $Q_{2B.1}$ that explained 36% of variation in FDK was also associated with INC, FHBdx, and DON. The marker-trait associations that explained more than 5% phenotypic variation were identified on chromosomes 1A, 2B, 3B, 5A, 7A, 7B, and 7D. To capture the effects of minor effect QTL, we lowered the threshold criterion to $-logP \geq 3.0$, which resulted in the identification of 67 unique loci. We have shown that the FHB-related traits have significant correlations with the number of favorable alleles at these loci, suggesting their utility in improving FHB resistance in this population by marker-assisted selection.

3.2 Introduction

Fusarium head blight (FHB) is the most devastating disease of wheat and barley in the United States mainly caused by the fungal pathogen *Fusarium graminearum* Schwabe [telemorph: *Gibberella zeae* Schw. (Petch)]. In United States, scab was reported in 1890s and there had been sporadic outbreaks of FHB epidemics since then (Parry et al., 1995). The re-emergence of the scab in 1990s had the long-lasting impact on the wheat industry (McMullen et al 1997). In 1991, scab hit hardest on the mid-western, southeastern, and mid-Atlantic states leading to yield reduction by 25% across 6.1 million hectares (McMullen et al., 1997). Since it's epidemic outbreak in 1993 in the United States, more than $10 billion worth of crop has been lost due to yield reduction, mycotoxin contamination, and quality deterioration (McMullen et al.,2012).

Several measures including rotation, tilling, and fungicide application are used to reduce initial inoculum, dispersal of the inoculum, or to restrict the spread of the infection in the spikes (Parry et al 1995). However, these measures either do not provide enough diseases control or are expensive, reducing the total profits. Use of resistant cultivars is the most effective way to reduce the adverse effect of scab disease on yield and quality (McMullen et al., 2012; He et al., 2013). Resistance to FHB disease is a quantitative inherited trait for which effector triggered immunity, known as "gene for gene interaction resistance", is not available. Schroeder and Christensen (1963) explained two types of FHB resistance i.e., resistance to initial infection in the spikes (Type I) and resistance to spread of the disease within the spike (type II). Type I and Type II resistance seem to be under control of different genes (Buerstmayr et al., 2003). In the developing grains, colonization of fungus leads to shriveled grains with colors ranging from pink, soft gray, to light-brown known as Fusarium damaged kernels (FDK). Severe infestation leads to lower test weights and damaged kernels resulting in loss of yield and quality. Resistance to FDK is known as Type III resistance (Mesterhazy, 1995). Furthermore, the fungus produces the mycotoxin known as Deoxynivalenol (DON) that affects animal and human health causing acute temporary nausea, vomiting, and fever (Sobrova et al., 2010). The United States Food and Drugs Administration (FDA) has established guidelines of 1 ppm maximum DON threshold for human food and 10 ppm maximum DON threshold for animal feed. The presence of higher levels of DON in wheat grains leads to discounts in price at the elevators, which brings further financial losses to growers (McMullen et al., 1997). The resistance shown by crops to accumulation of DON in the grains is known as Type IV resistance (Mesterhazy et al., 1999).

In the absence of complete resistance for FHB, identification of partial resistance from different sources and pyramiding them into cultivars is necessary to provide effective levels of protection against the devastating disease. In many cases these sources of resistance has been characterized via bi-parental quantitative trait loci (QTL) mapping. More than 250 QTL across all 21 chromosomes are associated with FHB resistance (Buerstmayr et al., 2009; Buerstmayr et al., 2019). Of all the QTL identified in genus *Triticum*, only eight have been fine mapped. They include *Fhb1*, *Fhb2*, *Fhb4*, *Fhb5*, *Qfhs.ifa-5A*, and *QFhb.nau-2B* in hexaploid wheat and *Qfhb.ndsu-3AS* and *Qfhb.mgb-2A* in durum wheat (reviewed by Buerstmayr et al., 2019, Wang et al., 2020). All the eight QTL were identified in wheat lines that are not adapted to the United States. Furthermore, alien introgression from wild relatives of wheat also provided FHB resistance traits. For example, *Fhb3*, was introgressed to chromosome 7A from *Leymus racemosus* (Qi et al., 2008); *Fhb6*, was introgressed to chromosome 1A from *Elymus tsukushiensis* (Cainong et al., 2015); and *Fhb7* was introgressed into chromosome 7D from *Thinopyrum elongatum* (Shen and Ohm 2007). Among the alien introgressed genes, *Fhb7* has been cloned and encodes a glutathione S-transferase (GST) (Wang et al., 2020). GST seems to be horizontally transferred to *Thinopyrum* from endophytic fungus *Epichloe* species and provides resistance to FHB by detoxifying the mycotoxin produced by the fungus (Wang et al., 2020). However, one major caveat of exotic sources is the linkage drag that negatively affects agronomic and end-use quality performance (Balut et al., 2013). In addition, the alien chromatin region is unable recombine when crossed with wild type chromatin due to lack of homologous pairing (Gill et al., 2011), and therefore creates an extended non-recombinant chromatin in the chromosome. Therefore, there is demand for identification and validation of QTL in native and adapted sources.

Genetic studies identified QTL in germplasm native to the US. In general, adapted soft red germplasm exhibit abundant FHB resistance compared to hard red spring wheat (HRSW) germplasm (McMullen et al., 2012). Several soft red winter wheat (SRWW) cultivars including Bess, Ernie, Freedom, Jamestown, Massey, NC-Neuse, Truman, and VA00W-38 were found as source of resistance. Mapping studies have identified QTL for FHB resistance on chromosomes 3BL from Massey (Liu et al., 2013); 3B and 5A from Ernie (Liu et al., 2007); 2B and 3B from Bess (Petersen et al., 2017); 1A, 4A, and 6A from NC-Neuse (Petersen et al., 2016); and 1B from Jamestown (Wright 2014). Similarly, hard red winter wheat cultivars that carry native resistance QTL include Everest, Overland, Lyman, Heyne, and Hondo. QTL mapping identified QTL in

chromosomes 2D from Everest (Clinesmith et al., 2019); 2A, 5A, and 6A from Overland (Eckard et al., 2019); and 3A, 4D, and 4A from Heyne (Zhang et al., 2012).

Most of mapping studies were performed by using biparental mapping strategies, which requires developing mapping populations. However, these studies are unable to capture allelic diversity in the breeding germplasm and provide low resolution of mapping. Genome-wide association studies (GWAS), on the other hand, uses population of crops for mapping and thus, bypasses the time of developing segregating populations. In addition, confidence intervals for significant loci narrows with greater probability of pinpointing shorter list of candidate genes due to greater resolution through reduced haplotype size (Cockram et al 2010, Weng et al 2011, and Ehrenreich et al 2009).

Several GWAS studies were performed for FHB resistance in SRWW population. For example, Arruda et al. (2016) performed a GWAS for FHB resistance in a population of 273 breeding lines that included 185 SRWW and identified significant markers in chromosomes 4A, 6A, 7A, 1D, 4D, and 7D. Holder (2018) performed GWAS for FHB resistance using a population of 360 southern SRWW lines that included 240 breeding lines developed by the University of Arkansas Wheat Breeding Program, where significant marker-trait association were identified on chromosomes 1A, 2D, 3B, 4A, 4B, 7A, and 7D. A GWAS was performed in population of 256 soft red winter wheat cultivars and breeding lines developed by the Triticeae Coordinated Agricultural Project (TCAP) and identified significant marker-trait associations (MTAs) on chromosomes 4A, 5B and 6B (Tessmann et al., 2019).

This study is focused on characterizing a SRWW breeding germplasm where for FHB resistance under field condition and identifying genomic regions significantly associated with FHB resistance. GWAS using breeding populations enables breeders to understand the genetic basis of traits, identify the segregating QTL and estimate frequency of desirable alleles in the population (Hoffstetter et al., 2016). We evaluated 392 lines for FHB response in field-based misted FHB nurseries across two years in three sets of experiments. We were able to identify genomic regions controlling FHB resistance traits in the populations and estimate frequency of desirable alleles in the population. This study is seminal to select parents that consists of favorable alleles for multiple QTL for crossing and developing FHB resistance enriched breeding population.

3.3   Materials and Methods

3.3.1   Plant Materials and Disease Evaluation

Three hundred and ninety-one lines of SRWW developed at Purdue University were evaluated in 2017/2018 (Y1) and 2018/2019 (Y2) seasons in Purdue Agronomy Center for Research and Education (ACRE) farm, West Lafayette, Indiana. Each line was planted in a 3ft row plot in a randomized incomplete block design. In the Y1, a single set and in Y2, two sets of experiments were planted. Each incomplete block consisted of 92 plots including 80 individual lines and four common checks each replicated three times. The common checks were Patterson, INW0411, INW0412, and Monon. Patterson is a susceptible cultivar, INW0411 and Monon are moderately susceptible, and INW0412 is a moderately resistant cultivar.

The field was inoculated using scabby corn developed by using nine isolates from IN, IL, and OH as described by Gilberts and Woods (2006). The scabby corn was applied at the rate of 40 g/m^2 approximately 2-3 weeks before heading. The field was mist irrigated to maintain high humidity for disease development. Phenotypic data were collected for days to heading (DH), disease incidence (INC), disease severity (SEV), FHB index (FHBdx), Fusarium damaged kernels (FDK), and deoxynivalenol concentration (DON). The INC and SEV data were recorded 21 days after anthesis. We measured INC as the percentage of spikes infected from a random sample of 20 spikes taken form a plot. We measured SEV as the percentage of spikelets infected within a spike averaged over 10 random spikes from a plot. The FHBdx, the index that shows overall FHB damage in spike, is calculated as INC X SEV/ 100. The FDK phenotype was assessed by comparing the samples to a known FDK standards. Known FDK standards were created by counting 300 seeds in a petri plate. Ten petri dishes with known FDK amounts ranging from 10% to 100% were prepared. For example, the 30% FDK standard, consisted of 90 FDK kernels and 210 healthy looking kernels. The experimental samples were rated against this known standard. The DON concentration was quantified using gas chromatography–mass spectrometry (GC/MS) at the Department of Plant Pathology at the University of Minnesota flowing methods described by Mirocha et al., (1998) and modified by Fuentes et al., (2005). Briefly, kernel samples from each line were grounded for 2 min followed by sample extraction and clean-up procedures described by Mirocha et al. (1998) with some modifications. Using 16 mL acetonitrile/water (84:16 v:v) extraction buffer, a 4 gram sample was extracted. After shaking the buffer for 1 hour in a rotary

shaker, 4mL of extract was passed through column with C18 and aluminium oxide. After getting the filtrate, for GC/MS analysis, 1 mL of extract was allowed to evaporate to dryness under nitrogen and derivatized by the silylating reagent (Fuentes et al., (2005).

3.3.2 Phenotypic Data Analysis

Data analysis was performed for each environment separately as well as combining all the dataset using lme4 package (Bates et al., 2007) in R v 3.6.1 (R Team, 2013). Three phenotypic data were produced including least square means *(lsmeans)* of Y1, Y2, and Best linear unbiased estimate (BLUE) derived from combined analysis of Y1 and Y2. For Y1, *lsmeans* were estimated by fitting mixed model with lines as fixed effects and incomplete block as random effects. For Y2, since there were two replications, the following mixed model was fitted to extract *lsmeans* for lines.

$$Y_{ijk} = \mu + Line_i + Rep_j + IncBlk_k(Rep_j) + \epsilon_{ik} \qquad [1]$$

Y_{ijk} is the observed phenotype for $Line_i$ located in k[th] incomplete block within a jth replication. μ is the overall mean, $Line_i$ is the fixed effect of the ith line, Rep_j is the random effect of jth replication, $IncBlk_j(Rep_k)$ is the random effect of kth incomplete block within a jth replication and ϵ_{ik} is the random error term. For combined analysis, variance components were estimated by fitting the random model:

$$Y_{ijk} = \mu + Line_i + Set_j + IncBlk(Set)_{jk} + \epsilon_{ijk} \qquad [2]$$

where Y_{ijk} is the observed phenotype, μ is the overall mean, $Line_i$ is the random effect of the ith line, Set_j is the random effect of jth experimental set (3 experimental sets), $IncBlk(Set)_{jk}$ is the random effect of kth incomplete block within the jth experimental set, and ϵ_{ijk} is the random error term. Using the variance components estimated form equation [1] and [2], broad sense heritability (H^2) on entry mean basis was estimated as follows:

$$H^2 = \frac{\sigma^2 g}{\sigma^2 g + \sigma^2 e/n}$$

, where $\sigma^2 g$ is the genotypic variance and $\sigma^2 e$ is the error variance, n is the number of experimental sets i.e., 3. Best linear unbiased estimates (BLUEs) for each trait were extracted by treating line as fixed effects in equation [1].

3.3.3 Genotyping and Population Structure

The genotyping and marker datasets of this SRWW population is well described in Chapter 2. Briefly, 436 lines were genotyped by using genotyping-by-sequencing (GBS) (Elshire et al., 2011) following the protocol described by Poland et al., (2012). The sequence data was aligned to the IWGSC v1.0 wheat genome assembly (Appels et al 2018) and the variants were called using TASSEL 5 GBS v2 Pipeline. After removing variants with less than 5% minor allele frequency and more than 20% missing data, 13,760 SNPs remained for 392 lines evaluated in this study. The imputation for missing data was performed using LDKNNimp method (Money et al., 2015) implemented in TASSELv 5.2 (Bradbury et al., 2007).

In addition, the population was genotyped for major agronomic and disease resistance genes at the USDA-ARS Eastern Regional Small Grains Genotyping Lab, with the primer sequences indicated by Ward et al., (2019). Briefly, the germplasm was assessed for the photoperiod loci located on the group 2 homoeologous chromosomes *Ppd-1*, *Fhb1* located on chromosome 3B and reduced height loci *Rht-B1* and *Rht-D1* located on 4B and 4D, respectively. The allelic effects of these major genes were estimated as the mean phenotype of lines homozygous for the mutant allele minus the mean phenotype of lines homozygous for the wild-type. Student's t-test was used to test the significance of the effect of major loci for each trait.

The hidden population structure in the population was studied using STRUCTURE (Pritchard et al., 2003) and principal component analysis. For STRUCTURE, admixture model-based clustering was performed with a total of 100,000 burn-in periods followed by 100,000 Markov Chain-Monte Carlo iterations for K values from 1 to 10. Similarly. PCA analysis was performed in TASSEL. The details of the population structure study were reported in Chapter 2.

3.3.4 Genome wide association study

GWAS was performed using FarmCPU package (Liu et al., 2016) in R v 3.6.1 using principal components to control the population structure. GWAS was performed separately for

environments Y1 and Y2, and the BLUEs extracted from equation [1]. Markers that had -logP ≥ 4 were considered as significant. For chromosomes with multiple SNPs, pairwise LD (r^2) between the SNPs were estimated. If SNPs had r^2 value more than or equal to 0.60, they were considered part of the same QTL and only one representative marker was kept. The proportion of phenotypic variance explained by individual MTAs was estimated by the coefficient of determination of a linear regression where phenotypes was dependent variable and the SNP was independent variable.

3.3.5 Additive effects of minor QTLs

It is well established that FHB resistance is controlled by many loci with minor effects. However, it is likely that using genome-wide markers for breeding is not needed. To cast a wide search process and capture the loci with impact on FHB traits, we identified all MTAs at *-logP $\geq$* 3 (lowered stringency). Correlated markers within each chromosome were removed, which resulted in a total of 67 markers.

Multiple linear regression, using ordinary least square method, was fitted for all FHB trait-year combinations, where all 67 markers were included as fixed effect independent variables. For each trait-year combination, a multiple adjusted R^2 was recorded as the amount of phenotypic variance explained by these markers. In addition, we created a variable which showed the total number of favorable alleles at these 67 markers for each individual. Then we fitted a simple linear regression with the BLUE phenotypic value as dependent variable and the number of favorable alleles as the independent variable. The Pearson's correlation coefficient between BLUE phenotype and the number of favorable alleles were estimated as the predictive ability of the number of favorable alleles in these markers for FHB traits. Finally, we performed a t-test to compare mean of 20 individuals with the highest number of favorable alleles and mean of 20 individuals with the lowest number of favorable alleles.

3.4 Results

3.4.1 Trait distributions, correlations and heritability

The response of 392 SRWW lines to FHB disease was evaluated in a misted field-based study for two seasons 2017-18 (Y1) and 2018-19 (Y2). We evaluated INC, SEV, FHBdx, FDK,

and DON. The mean, standard deviation, and ranges of the five FHB-related traits are provided in Table 3.1. Incidence ranged from 24% to 100% over the two seasons. The average INC of two-year BLUE values was 78.3%. INC followed roughly a normal distribution in Y1, Y2, and BLUE, although it was slightly left-skewed in Y1. Severity revealed a range of 3% to 100% over the two seasons. The two-year BLUE values averaged at 41.4%. The overall range and mean of BLUE observed for FHBdx were 0.5-100% and 34.4%, respectively. SEV and FHBdx followed a normal distribution in Y1, Y2, and BLUE although the phenotypes were slightly right-skewed in 2018.

Table 3.1. Descriptive statistics of the five FHB-related traits.

Trait	ENV	Mean	SD	Min.	Max.	H^2
INC (%)	Y1	77	18.5	24	100	-
	Y2	73.8	12	30.5	100	0.39
	BLUE	78.3	11.3	44.1	100	0.41
SEV (%)	Y1	44.2	21.4	3.2	100	-
	Y2	35.7	12.5	9.7	93.2	0.37
	BLUE	41.4	12.1	15.3	98.3	0.45
FHBdx (%)	Y1	36.7	23.8	0.5	100	-
	Y2	28.9	13.5	4.3	84.6	0.44
	BLUE	34.4	12.9	8.4	99.7	0.46
FDK (%)	Y1	15.2	14.4	0.5	79	-
	Y2	24.6	12.5	1.4	69.6	0.41
	BLUE	22.3	9.8	3.7	58	0.60
DON (ppm)	Y1	2.7	1.7	0.1	11.4	-
	Y2	8.2	3.8	1.5	29.8	0.39
	BLUE	6.4	2.7	1	21.8	0.63

INC: Disease Incidence; SEV: Disease Severity, FHBdx: FHB index, FDK: *Fusarium* damaged kernels, DON: Deoxynivalenol content, Y1: 2017/2018 season, Y2: 2018/2019 season, BLUE: the best linear unbiased estimate of phenotypes based on combined year analysis, SD: Standard Deviation, Min: Minimum, Max: Maximum, H^2: Broad-sense heritability estimated based on entry mean basis.

The overall range observed for FDK and DON were 0.5-79% and 0.1-29.8 ppm, with the average of BLUE values of 22.3% and 6.4, respectively. FDK and DON seems to follow a normal distribution in Y1, Y2, and BLUE, although, were slightly left-skewed in Y1. A general correlation

pattern was observed indicating two groups of traits within each environment. In this pattern, INC, SEV, and FHBdx were tightly correlated with correlation coefficient (r) ranging from 0.48 to 0.97). Similarly, FDK and DON were tightly correlated (Range of r = 0.53, -0.66). Therefore, it is expected that the lines with the best score in INC or SEV are not those that showed the best scores in FDK and DON. For example, while 10534A1-24-2 and 10442A1-19-5 showed INC of 75.2% and 60.2% and SEV of 38.3 and 23.4%, they showed the best DON levels (1, 1.5 ppm) across both years of the study.

Among the four checks, Patterson was the most susceptible with INC, SEV, FHBdx, FDK and DON values of 92.3%, 75.8%, 71%, 32.4% and 10.8ppm respectively (Appendix B1). Similarly, INW0412 was the most resistant check with INC, SEV, FHBdx, FDK and DON average values of 62.3%, 26.2%, 16.6%, 26.6%, and 7 ppm respectively. Phenotypic values of INW0411 and Monon were between Patterson and INW0412 (Appendix B1). Furthermore, we evaluated the four repeated checks only for understanding the genotype by year (G x Y) interactions. A fixed effects models that included genotypes (G), year (Y), and G x Y interaction the model. The F-test showed that GxY interactions were significant at 5% level of confidence for all FHB-related traits including INC, SEV, FHBdx, and DON. However, G x Y was not significant for FDK.

Since Y1 only consisted of single replicated experiment, we were not able to estimate H^2. In Y2, we repeated all germplasm in two replicates, so we were able to estimate the broad-sense heritability (H^2). The H^2 values for INC, SEV, FHBdx, FDK, and DON in Y2 were 0.39, 0.37, 0.44, 0.41, and 0.39, respectively. Using the phenotypic data across the two years, broad-sense heritabilities (H^2) for INC, SEV, FHBdx, FDK, and DON were 0.41, 0.45, 0.46, 0.60, and 0.63 respectively.
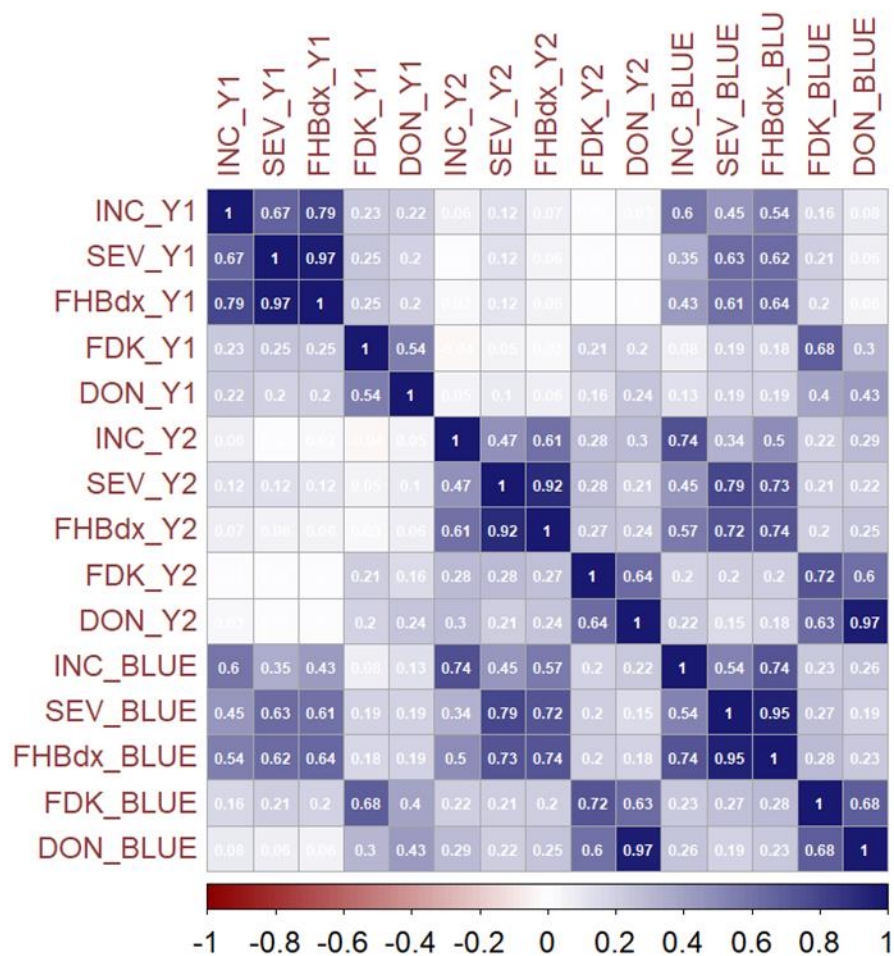
Figure 3.1. Coefficient of correlations among the traits evaluated in 2017/2018 (Y1) and 2018/2019 (Y2) and the best linear unbiased estimates (BLUE) based on combined year analysis. Traits are INC: Disease Incidence; SEV: Disease Severity, FHBdx: FHB index, FDK: *Fusarium* damaged kernels, and DON: Deoxynivalenol content.

85

3.4.2 Presence and effect of major agronomic and disease resistance genes.

One hundred and thirty-two individuals harbored the favorable allele at *Fhb1* locus, and 203 individuals did not possess the favorable allele. The presence of favorable allele of *Fhb1* was significant on SEV and FDK across Y1, Y2, and BLUE (Table 3.2). In this study, the *Fhb1* locus explained up to 1.8-3% and 5.7-10.1% of variations while reducing SEV and FDK by at least 4 and 6% across all environments, respectively (Table 3.2). The effect of *Fhb1* locus was significant on FHBdx only in Y2. *Fhb1* locus did not significantly affect INC.

The effect of *Rht-B1* locus was insignificant on INC but found to be significant on SEV, FDK, and FHBdx across all environments (Table 3.2). Three hundred and seventeen individuals harbored the height reducing *Rht-B1b* allele while 15 individuals possessed the wild-type *Rht-B1a* allele. Those with height reducing allele, on average, showed significantly higher SEV, FHBdx, and FDK scores compared to those with wild-type allele, indicating that shorted plants show more susceptibility for SEV, FHBdx, and FDK traits.

Among the *Ppd1* loci, *Ppd-D1* significantly affected days to heading in Y2 and BLUE. The dominant insensitive allele (*Ppd-D1a*) flowered earlier than the recessive sensitive allele (*Ppd-D1b*). The effect of *Ppd-D1* was not significant on any of the FHB-related traits. Although not significant for days to heading, *Ppd-A1* and *Ppd-B1* were significant for FHB resistance traits. Our data showed that Ppd-A1, present in ~47% of homozygous individuals, seems to have significant effect for INC, SEV, and FHBdx in environment Y1 (Table 3.2). Similarly, *Ppd-B1*, for which 89 individuals were homozygous, seems to have significant effect for SEV and FHBdx on environment Y2 and BLUE and for INC on BLUE. On average, the individuals with insensitive alleles at *Ppd-A1* and *Ppd-B1* loci had higher phenotypic values for FHB-related traits at environments where the genes were significantly influential.

Table 3.2. The effect of markers and their significance for major loci on FHB-related traits. Allelic effects are estimated as the difference between mean of homozygous individuals for contrasting alleles for each locus. Numbers below the major loci names represent the number of homozygous lines with the contrasting alleles.

Traits	ENV	Rht-B1 (317:15)	Ppd-A1 (157:179)	Ppd-B1 (89:278)	Ppd-D1 (222:127)	Fhb1 (132:203)
INC	Y1	5.0	4.8*	0.8	-3.0	0.1
	Y2	-0.5	0.1	2.9	1.3	-0.1
	BLUE	2.5	0.7	2.6*	-1.1	2.0
SEV	Y1	6.5*	5.1*	3.1	-4.3	-5.7*
	Y2	4.2*	-0.7	2.8*	-1.3	-4.9***
	BLUE	5.8***	0.6	3.0*	-2.2	-4.0**
FDK	Y1	5.5**	1.2	-1.1	0.2	-8.5***
	Y2	6.7**	3.4*	1.4	1.2	-6.1***
	BLUE	6.1***	1.8	0.8	0.9	-6.3***
FHBdx	Y1	7.6*	6.0*	3.6	-4.9	-4.0
	Y2	3.9*	-1.0	3.2*	-1.0	-3.9*
	BLUE	5.9***	0.7	3.6*	-2.2	-2.1
DON	Y1	1.0***	0.1	0.2	0.2	-1.4***
	Y2	1.7**	0.1	0.1	0.4	-1.8***
	BLUE	1.5***	0.2	0.1	0.1	-1.5***

* Marker effect is significant at the 0.05 level; ** Marker effect is significant at the 0.01 level; *** Marker effect is significant at the 0.001 level

INC: Disease Incidence; SEV: Disease Severity, FHBdx: FHB index, FDK: *Fusarium* damaged kernels, DON: Deoxynivalenol content, Y1: 2017/2018, Y2: 2018/2019, BLUE: the best linear unbiased estimate of phenotypes based on combined year analysis.

3.4.3 GWAS

We used the five FHB traits measured from 392 SRWW germplasm in two years and 13,784 GBS-derived markers and performed GWAS analysis. Altogether, 165 MTAs across 15 chromosomes were identified at $-logP \geq 4.0$ for the five FHB-related traits including INC, SEV, FHBdx, FDK, and DON. Pairwise LD between the markers on the same chromosome showed that four regions on chromosomes 2B, 3B, 5A, and 7D were represented by more than one marker. The marker with the highest $-logP$ value for each region is reported in Table 3.3. Twenty-five unique genomic regions, QTL, were identified (Figure 3.1). Nine loci were identified in Y1, 4 loci were identified in Y2, and 10 loci were identified using BLUE values. Two regions Q_{5A} and $Q_{2B.1}$ were identified commonly in at least two of the three phenotypic data (Y1, Y2, and BLUE). Three regions $Q_{2B.1}$, Q_{3B}, and $Q_{7A.3}$ were associated with two or more FHB-related traits (Table 3.3, Figure 3.2, and 3.3). For SEV only one locus was identified while for DON 16 loci were identified.

For INC, three loci $Q_{2B.1}$, $Q_{3A.2}$, and $Q_{5A.2}$ were identified (Table 3.3). The variant located at 2B_547543336 ($-logP = 4.3$) showed the highest effect of 21.1% for INC in Y1 (Fig. 3.2), however, only explained 0.3% variation in the phenotype. The variant located at 3A_712285960 ($-logP = 4.3$) was identified in Y1, which reduced INC in Y1 by 7.1% and explaining 4.7% of variation. The 5A.2 locus, harboring eight tightly correlated SNPs from 702.9- 708.1 Mbp, was identified ($-logP = 4.9$) for INC in Y2 and BLUE (Fig. 3.3). The peak variant in this region was located at 5A_702908675 and showed an effect size of 3% and explained 5.5% of variation (Table 3.3). Altogether, the three loci $Q_{2B.1}$, $Q_{3A.2}$, and $Q_{5A.2}$, explained 5.3, 5.3 and 5.1% variation for INC in Y1, Y2, and BLUE respectively.

The Q_{3B} region, harboring six tightly correlated markers, was identified for SEV on chromosome 3B in environment Y2 (Table 3.3). The most significant variant in this region was 3B_11165927 ($-logP = 4.6$), which reduced SEV by 3.9% and explained 3.79% of variation.

For FHBdx, five loci were identified. The $Q_{2B.1}$, $Q_{2B.2}$, $Q_{6A.1}$, and $Q_{6B.2}$ regions were identified in Y1 (Fig. 3.2) while Q_{3B} was identified in Y2. The marker for $Q_{2B.1}$ was 2B_547543336 ($-logP = 6.5$) that reduced FHBdx by 22.5% and explained 43% of phenotypic variation (Table 3.3). The marker for $Q_{2B.2}$ was 2B_787820223 that reduced FHBdx by 10.7% and explained 3.12% of phenotypic variation. The variants at Q_{3B} (3B_11165927) and Q_{6A} explained less than 1% of the

phenotypic variation while $Q_{6B.2}$ region explained 1.5% of phenotypic variation. The marker effects of the three QTL, Q_{3B}, Q_{6A}, ad $Q_{6B.2}$, ranged from 4-11.5% (Table 3.3). The amount of phenotypic variance explained altogether by the five QTL was 49.4, 2.8, and 7.4% for FHBdx across Y1, Y2, and BLUE phenotypes (Table 3.3).

Seven genomic regions were identified for FDK (Table 3.3). The regions $Q_{2B.1}$, Q_{3B}, and $Q_{7A.3}$ were identified by BLUE values (Figure 3.3) while regions $Q_{4A.1}$, $Q_{5A.1}$, and $Q_{6B.1}$ were identified in environment Y2. Similar to SEV, the peak marker at $Q_{2B.1}$ was 2B_669232996 ($-logP$ = 6.5) which reduced FDK by 7.1% and explained 36% of FDK phenotypic variation. The variant located at 7B_723092825 ($-logP$ = 4.8), reduced FDK by 4.1% and explained 6.3% of variation in environment Y2 (Table 3.3). The amount of phenotypic variations explained by other regions ranged from 1.4 -4.2% and they reduced FDK from by 2.8-4.8%. These seven regions explained 4.1% of variation in Y1, 16.1% of variation in Y2, and 46% of variation in BLUE values.

For DON, sixteen loci were identified in environment Y1 and BLUE. The peak variant at Q_{3B} was located at 3B_10170821 ($-logP$ =7.3) which reduced DON by 0.5 ppm and explained 16.8% of variation in DON in environment Y1. The next MTAs were located at 7D_592498391, 1A_499864143 and 7A_727710655 explaining 8.9, 5.9, and 5.6% of the phenotypic variation, respectively (Table 3.3). These three loci reduced DON concentration by 0.8, 0.6 and 0.5 ppm, respectively. The variant 1B_628959667 ($-logP$ = 6.1) showed reduced DON by 1.2 ppm and explained 4.5% of the variation. The peak variant at $Q_{2B.1}$ was 2B_216942046 ($-logP$ = 11.2) which explained 1.8% of variation in Y2. The amount of phenotypic variation explained by other regions ranged from 0.06% for Q_{4D} and $Q_{7B.1}$ to 2.5% for $Q_{5B.1}$. These markers reduced DON by 0.4 ppm for $Q_{7A.1}$ and by 2.9 ppm for $Q_{7A.2}$ (Table 3.3). The sixteen loci, altogether, explained 28.2, 30.8, and 33.6% variation in the DON data across Y1, Y2, and BLUE values.

Table 3.3. Summary of identified loci for FHB-related traits and seasons.

QTL	Trait_ENV[a]	Position (Mbp)	No. of SNPs	Peak marker	Alleles[b]	$-logP$	Maf[c]	Effect	R^2
Q_{1A}	DON_BLUE			1A_499864143	A/C	6.8	0.27	0.6	5.9
Q_{1B}	DON_BLUE			1B_628959667	G/A	6.1	0.08	-1.2	4.5
$Q_{2B.1}$	DON_Y1	161.2-456.2	28	2B_216942046	C/G	11.2	0.42	-1.3	1.8
	INC_Y1			2B_547543336	T/C	4.3	0.43	21.1	0.3
	FHBdx_Y1			2B_547543336	T/C	6.5	0.43	22.5	43.2
	FDK_BLUE			2B_669232996	C/G	4.2	0.43	7.1	36.0
$Q_{2B.2}$	FHBdx_Y1			2B_787820223	A/G	4.3	0.05	10.7	3.1
Q_{2D}	DON_BLUE			2D_645828400	T/C	4.1	0.1	0.8	0.8
$Q_{3A.1}$	DON_BLUE			3A_697615363	C/T	5.7	0.21	-0.7	1.3
$Q_{3A.2}$	INC_Y1			3A_712285960	C/T	4.3	0.16	-7.1	4.7
Q_{3B}	FDK_BLUE			3B_7240747	A/G	7.3	0.41	2.8	3.3
	DON_Y1			3B_10170821	G/C	7.3	0.46	-0.5	16.8
	SEV_Y2	11.2-11.9	6	3B_11165927	T/C	4.6	0.48	3.9	3.8
	FHBdx_Y2			3B_11165927	T/C	4	0.48	4.0	0.03
$Q_{4A.1}$	FDK_Y2			4A_605557514	C/T	5.8	0.11	-3.9	4.2
$Q_{4A.2}$	DON_Y1			4A_728086043	C/T	4.6	0.08	-2.1	0.6
Q_{4D}	DON_Y1			4D_450312303	G/T	4.3	0.05	-1.6	0.06
$Q_{5A.1}$	FDK_Y2			5A_489912043	T/C	5.2	0.07	4.8	2.3
$Q_{5A.2}$	INC_Y2			5A_702908675	T/C	4.2	0.34	-3.2	3.4
	INC_BLUE	702.9-708.1	8	5A_702908675	T/C	4.9	0.34	-3.0	5.5
$Q_{5B.1}$	DON_BLUE			5B_283024388	G/A	9.7	0.1	1.3	2.5
$Q_{5B.2}$	DON_BLUE			5B_712850551	A/G	4.5	0.07	0.9	0.4
Q_{6A}	FHBdx_Y1			6A_39207876	C/T	4.5	0.16	6.3	0.02
$Q_{6B.1}$	FDK_Y2			6B_157666111	G/C	4.2	0.1	4.2	1.4
$Q_{6B.2}$	FHBdx_Y1			6B_508844418	T/C	4.1	0.05	-11.5	1.5
$Q_{7A.1}$	DON_BLUE			7A_4745932	G/C	4.5	0.48	0.4	1.3
$Q_{7A.2}$	DON_Y1			7A_39064670	G/A	6.7	0.08	-2.9	0.4
$Q_{7A.3}$	FDK_BLUE			7A_727710638	C/G	4	0.07	3.2	1.9
	DON_BLUE			7A_727710655	A/G	4	0.22	0.5	5.6
$Q_{7B.1}$	DON_Y1			7B_41324506	C/G	4.5	0.05	-1.7	0.06
$Q_{7B.2}$	FDK_Y2			7B_723092825	G/A	4.8	0.08	-4.1	6.3
$Q_{7D.1}$	DON_Y1	1.53-52.8	95	7D_8587373	T/C	9.8	0.08	0.6	1.8
$Q_{7D.2}$	DON_BLUE			7D_592498391	A/G	6.9	0.14	-0.8	8.9

a, Trait_ENV: Trait and Environment combination, INC: Disease Incidence; SEV: Disease Severity, FHBdx: FHB index, FDK: *Fusarium* damaged kernels, DON: Deoxynivalenol content, Y1: 2017/2018, Y2: 2018/2019; b, The underlined nucleotide is the favorable allele that reduces disease amount, and c, Maf: Minor allele frequency
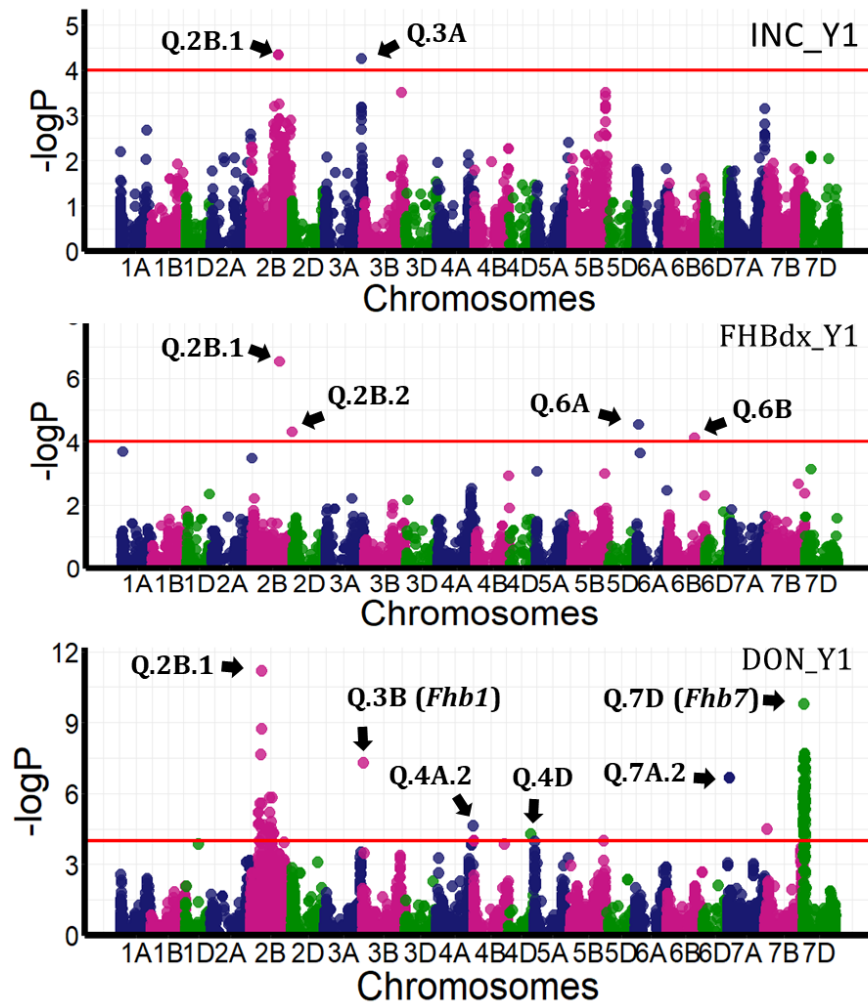
Figure 3.2 Manhattan plot showing loci associated with INC, FHBdx, and DON evaluated in 2017/2018 (Y1) season. The horizontal red line represents *-logP* = 4.0. Traits are INC: Disease incidence, FHBdx: FHB index, and DON: Deoxynivalenol content.
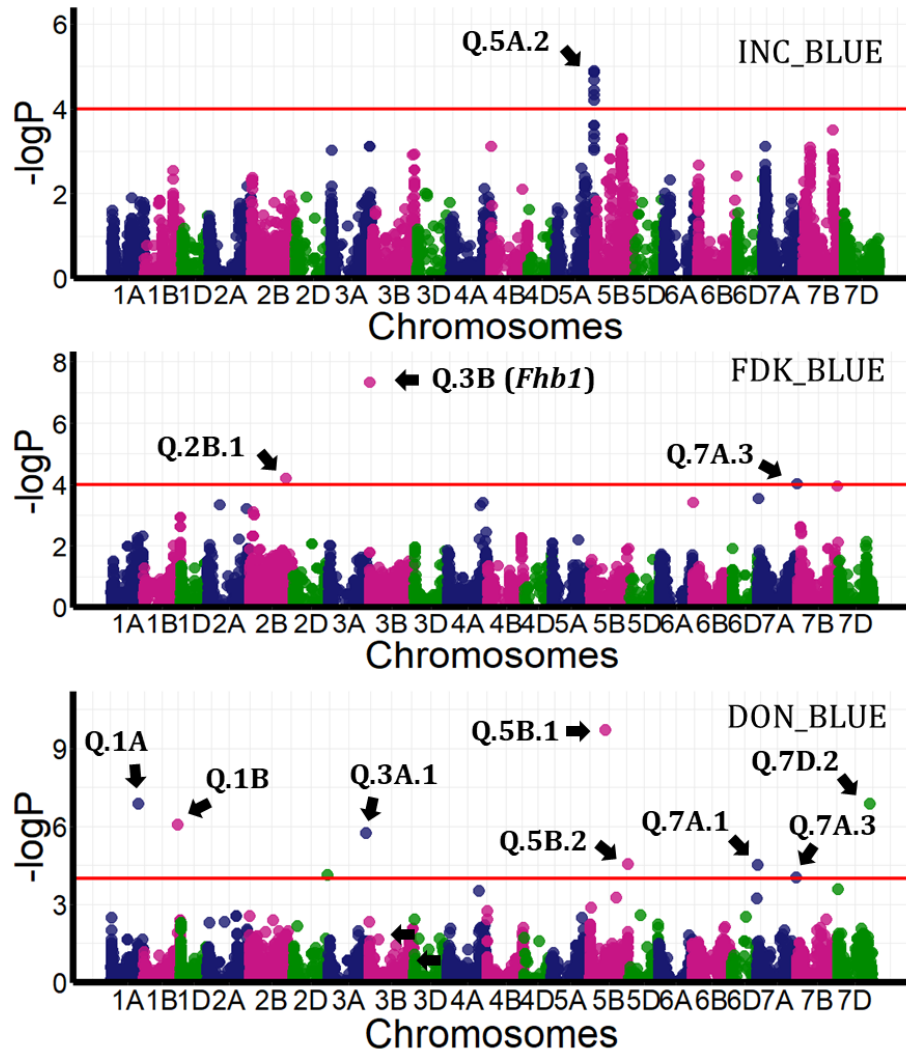
Figure 3.3 Manhattan plot showing significant loci associated with INC, FDK, and DON for BLUE estimated using combined year data analysis. The horizontal red line represents $-logP$ = 4.0. Traits are INC: Disease incidence, FDK: *Fusarium* damaged kernels, and DON: Deoxynivalenol content.

3.4.4 Additive effects of minor QTLs

It is well established that FHB resistance is a highly quantitative trait and many small-effects loci govern the traits. In this study, the amount of variation in INC and SEV explained by the significant markers was particularly low. By setting the threshold of $-logP = 4$, it is likely that many of the minor QTL are missed. To capture the minor effects loci, we reduced the cut-off criteria to $-logP = 3$ and performed ordinary regression analysis to study the amount of variation explained by these variants. Using $-logP \geq 3.0$ as criteria, 360 variants were identified across the three phenotypic datasets (Y1, Y2, and BLUE). After filtering correlated markers (r >0.6), only 67 markers remained (Appendix B2). The amount of phenotypic variation explained by the 67 markers, expressed as the adjusted R^2 of the multiple regression, listed in Table 3.4, was higher for DON (37.4, 43.2, and 46.1%) and FDK (20, 38.2, and 38.3%) across Y1, Y2, and BLUE, respectively. For INC, SEV, and FHBdx, the adjusted R^2 ranged from 14.4% for SEV in Y2 to 22.6% for INC in BLUE.

The number of favorable alleles in an individual ranged from 42 in "04620A1-1-7-4-17" to 90 in "07469A1-6-1-3" and "10531A1-4-1" (Appendix B3). It is expected that the individuals with higher number of favorable alleles exhibit less disease symptoms and phenotypes. When we fitted a linear regression model by the FHB traits as dependent variables and the number of favorable alleles as independent variable, we observed the strongest relationship with FHBdx (r = -0.44), followed by SEV (r = -0.41), and INC (r = -0.40) (Appendix B4). For FDK and DON, the correlation coefficient with the number of favorable alleles were slightly lower (-0.29 and -0.28 respectively) although the amount of variation explained by the markers were higher. Finally, when we performed t-tests to compare the means between tails of distribution for number of favorable alleles, we observed that the 20 individuals with highest number of favorable alleles had 15.1%, 13.7%, 16.1%, 13.2%, and 3.6 ppm less INC, SEV, FHBdx, FDK, and DON values than the 20 individuals with lowest number of favorable alleles (P- value < 0.0001). This analysis shows that across a variety of germplasm, the accumulating effect of favorable alleles for these loci confer improved resistance to FHB disease.

Table 3.4 The coefficient of determination expressed as multiple adjusted R^2, indicating the percentage by which the ordinary least square regression with 67 markers explains the phenotypic variation.

Trait_ENV	R^2 (%)
INC_Y1	15
INC_Y2	15.9
INC_BLUE	22.6
SEV_Y1	15.5
SEV_Y2	14.4
SEV_BLUE	18.9
FDK_Y1	20
FDK_Y2	38.2
FDK_BLUE	38.3
FHBdx_Y1	17.8
FHBdx_Y2	14.8
FHBdx_BLUE	21.7
DON_Y1	37.4
DON_Y2	43.2
DON_BLUE	46.1

3.5 Discussion

Soft red winter wheat has economic significance in the United States, producing 7.8 Tg of grains on 2.3 million ha of land in 2019 (USDA-NASS, 2020). About 54% of the total production of SRWW (4.2 Tg) was used for domestic consumption as food, seed, and feed (USDA-ERS, 2020), whereas ~41% (23.2 Tg) of the US SRWW was exported as grain, flour, and processed products (USDA-ERS, 2020). Over the last 10 years SRWW provided an estimated annual economic value of over US$2.2 billion (USDA-NASS, 2020). Despite the significant economic contribution, the total production of SRWW has significantly declined over the last three decades from 14.8 million tons in 1990 to 4.38 million tons in 2019. The rapid decline in production has been attributed to the policy changes related to the support of agricultural commodities and biofuel crops, introduction and adaptation of genetically modified (GM) corn and soybean, and epidemic outbreaks of *Fusarium* head blight disease (McMullen et al., 2012). This damage can be controlled

by using FHB resistant cultivars. In this study, we evaluated 392 SRWW lines developed at Purdue University since 1980s in misted FHB nursery for FHB resistance and performed GWAS to identify major loci governing FHB resistance in the population.

The population was genotyped for established major genes influencing several agronomic traits including plant height, days to heading and FHB resistance. Major genes that govern morphological traits such as plant height and days to heading are known to influence FHB resistance (Buerstmayer et al., 2019). *Rht-B1b* allele that confers semi-dwarfism is known to increase susceptibility to FHB (Srinivasachary et al., 2009; Liu et al., 2013). In this study, individuals homozygous to *Rht-B1b* allele showed greater INC, SEV, and FHBdx by at least 3.87%. Among the *Ppd1* homeologs, the *Ppd-D1* locus has been reported to have significant effect on FHB response such that earlier flowering individuals with insensitive allele (Ppd-D1a) show greater disease symptoms compared to the later flowering individuals with sensitive allele (*Ppd-D1b*) (McCartney et al., 2016; Liu et al., 2013). While in our study we observed an effect of Ppd-D1 on days to heading, we did not find significant effect of *Ppd-D1* on FHB traits. Individuals with insensitive allele of *Ppd-B1* and *Ppd-A1* showed greater INC, SEV, FDK, and FHBdx scores, however, were not consistent across the two environments.

Fhb1 is the most widely used major QTL for FHB resistance that is consistently associated with Type II resistance (McMullen et al., 2012). We genotyped the population for the presence of *Fhb1* allele and evaluated its effect on the resistant traits. Consistent with the previous studies, the *Fhb1* significantly reduced SEV, FHBdx, FDK, and DON in all environment, however, did not show significant effect on INC. The favorable allele for *Fhb1* was present only in 39% of individuals suggesting the potential for increasing the allele frequency in the breeding population.

GWAS identified eighteen genomic regions that explained more than 1% variation for at least one FHB-related trait in an environment. Nine out of eighteen located on chromosomes 1A, 1B, 2B, 4A, 5B, 7A, and 7B are either very close to or within a FHB resistance QTL that has been previously identified in SRWW cultivars adapted to the United States. Peterson et al., (2016) identified QTL for FHB resistance on chromosomes 1A, 1B, 1D, 2A, 4A, 5B, and 6A in SRWW RIL population derived from the cross Nc-Neuse x AGS2000. Similarly, Islam et al., (2016) mapped QTL for FHB resistance using Truman x MO-94-317 on chromosomes 1B, 2A, 2B, 2D,

3B, 3D and 6A. Later, Peterson et al., (2017) validated the QTL identified in Truman and Nc-Neuse through QTL mapping using population derived from Bess, which is a sister line of Truman and Nc-Neuse. QTLs were identified in chromosomes 1A, 1B, 2A, 2B, 4A, 4D, 5B, 5D, and 6A. The significant loci on chromosomes 1A, 1B, 2B, 4A, and 5B identified in these studies were either within or close to the QTL reported in these previous studies.

The region Q_{1A} (1A_499864143) identified for DON in BLUE explained 5.9% phenotypic variation. Peterson et al., (2016) reported a nearby QTL *Qfhb.nc-1A* flanked by markers IWA3805 (540.1 Mbp) and IWA6152 (574.9Mbp) on chromosome 1A. Similarly, Q_{1B} (1B_628959667) which explained ~4.5% of variation in DON, is nearby the previously identified *Qfhb.nc-1B.4* flanked by markers IWB31692 (658.5 Mbp) and IWB9040 (676.8 Mbp) by Peterson et al., (2017).

The region $Q_{2B.1}$, which was significantly associated with DON, INC, FHBdx, and FDK is rather wide. It spans the centromere being located between 216,942,046 to 669,232,996 bp. A second adjacent but independent region was identified as $Q_{2B.2}$ (2B_787820223) that explains 3.1% phenotypic variation of FHBdx in Y1. Liu et al., (2007) reported a QTL in chromosome 2B in SRWW cultivar 'Ernie' linked with SSR marker Xgwm276b. Similarly, Islam et al., (2016) reported a QTL (*Qfhb.nc-2B.1*) for SEV in chromosome 2B derived from SRWW cultivar TRUMAN. Based on the sequences of the flanking DArT markers (wPt8548–wPt8916), *Qfhb.nc-2B.1* is located between 159,487,181 to 785,251,367bp. Petersen et al., (2017) validated the 2B QTL from TRUMAN by mapping FHB resistance in a population derived from Bess and NC-Neuse. The marker associated with SEV was IWA5830, located at 165,115,668 bp whereas the peak marker for DON, IWB1769, was located at 767,021,628 bp. The regions identified in these studies seems to cover a very large region potentially due to presence of 2B:2G translocation on NC-Neuse that reduces recombination. This made the comparative analysis of our QTL regions difficult. The pedigrees of the lines in our study showed Truman as parent suggesting that MTAs identified in this study could potentially be related to *Qfhb.nc.2B.1*. Interestingly, the two loci $Q_{4A.1}$ (4A_605557514) and $Q_{4A.2}$ (S4A_728086043) we identified for FDK and DON are very close to the two QTL *Qfhb.nc-4A.2a* and *Qfhb.nc-4A.2b* (609.8- 665.5 Mbp and 707-742.6 Mbp) that Peterson et al., (2017) reported on chromosome 4A.

The DON QTL $Q_{7A.1}$ (7A_4745932) explained 1.3% of variation in BLUE. Our finding is consistent with the MTA Holder (2018) reported ~4.7 Mbp down the chromosome 7A that was associated with INC, SEV, and FDK in a SRWW population developed by the University of Arkansas Wheat Breeding Program (Holder, 2018). Interestingly, Liu et al., (2009), through meta-analysis, identified a QTL for SEV linked with SSR markers located between 5.4 Mbp (Xfbb121) and 11.3 Mbp (Xgwm433). The FDK QTL $Q_{7B.2}$ (S7B_723092825) in the distal end of long arm of chromosome 7B, explained 6.5% of variation in Y2. A QTL for disease incidence was identified on chromosome 7B in a bi-parental population derived from Goldfield X Patterson (Gilsinger et al., 2005). Like Patterson, Goldfield is a SRWW line developed and released from Purdue University breeding program. Other supports for the presence of a QTL in this region are a SEV QTL found in a double haploid population derived from DH181 x AC Foremost cross flanked by wmc276 (716.5 Mbp) and wmc526 (734.8) (Zhuping et al., 2005) and a QTL identified by using a progeny from Goldfield x DH181 cross (Liu et al., 2009).

Nine loci identified on chromosomes 3A, 3B, 5A, 6B, and 7D are either within or very close to FHB resistance QTLs identified in unadapted primary gene pool. For example, Q_{3B} and $Q_{6B.1}$ seems to be located nearby the already identified *Fhb1* and *Fhb*2 QTL, respectively. Via fine mapping, Su et al., (2019) determined that the TaHRC is the candidate gene for the Fhb1 locus. TaHRC gene (GenBan ID: CBH32655.1) is located from 8,526,628 to 8,529,572 bp, which resides within the region identified in this study, and the peak marker of Q_{3B} i.e., S3B_11165927 is highly correlated (~0.80) with the *Fhb1* diagnostic marker. Fine mapping showed that FHB2 on 6B originated from Sumai-3 derivative line BW278 (Cuthbert et al., 2007) which later was assigned to a narrower region flanked by markers Xwmc737 (154.3 Mbp) and Xmag 3017 (159.3 Mbp) by Jia et al., (2018). The variant representing $Q_{6B.1}$, 6B_157666111, resides within the fine-mapped *Fhb2* QTL region.

We identified two QTL $Q_{3A.1}$ and $Q_{3A.2}$,~15Mbp apart, on chromosome 3A that control DON and INC. Earlier reports shows QTL on chromosome 3A derived from plant introductions Huapei 57-2 and Fundulea 201R introduced to Purdue breeding program as novel source of FHB resistance (Bourduncle and Ohm 2003; Shen et al., 2003). However due to low resolution of mapping in the previous study, it is impossible to explain whether the QTL identified in this study is same or not.

We identified two QTL on chromosome 5A. The $Q_{5A.1}$ (5A_489912043) controlled FDK in Y2 and explained 2.3% of variation. The $Q_{5A.2}$ (5A_ 5A_702908675) controlled INC in Y2 and BLUE and explained 3.4% and 5.5% of variation, respectively. A well-known QTL *Qfhs.ifa-5A* has been fine mapped and is reported to have two QTLs - from 70.7-119.9 Mbp and 245.9-290 Mbp (Buerstmayr et al. 2002; Steiner et al., 2019). Based on physical locations, we concluded that these are different from those two QTL we identified on chromosome 5A. Our study revealed $Q_{5A.1}$ (5A_489912043) and $Q_{5A.2}$ (5A_ 5A_702908675) which are far apart from *Qfhs.ifa-5A*. The QTL $Q_{5A.2}$ seems to be closer to another European QTL derived from Arina, Pirat, and Apache (Draeger et al., 2007; Gervais et al., 2003; Holzapfel et al., 2008, Paillard et al., 2004). Based on the primer sequences of flanking markers Xgwm6 and Xgwm410 (Holzapfel et al., 2008), the QTL region covered 678.3 to 709.7 Mbp which included the MTA identified in this study. Furthermore, a gene, *Tipped1* (*B1*), that suppresses the awn in wheat is mapped in chromosome 5A. Würschum et al., (2020) reported that the candidate gene for *Tipped1* is a *C2H2* zinc finger (*TraesCS5A01G542800*) that is located at 698.5 Mbp. Given the proximity of the MTA 5A_702860236 and *TraesCS5A01G542800*, it is possible that the MTA is associated with awn characteristics. Awn length is known to confer passive FHB resistance such that awnless genotypes show lesser disease severity than awned genotypes (Mesterhazy 1995).

Two regions $Q_{7D.1}$ and $Q_{7D.2}$ on chromosome showed elevated signals for DON in Y1 and BLUE. $Q_{7D.1}$ region consists of 95 markers with elevated LD (>0.90) with each other. Such extended LD blocks are observed when an alien chromatin is introgressed in wheat chromosome and unable to recombine when crossed with wild type chromatin due to lack of homologous pairing (Gill et al., 2011). Our data shows this QTL is *Qfhs.pur-7EL*, also known as, *Fhb7*. Shen and Ohm (2007) mapped *Qfhs.pur-7EL* for disease severity using population derived from exotic Thinopyrum-7E translocation lines K2620 (intermediate FHB resistant) x K11463 (FHB susceptible line). The QTL mapped on the distal end of translocated 7E chromosome translocated to chromosome 7D of wheat (Wang et al., 2020). Furthermore, it appears that $Q_{7D.2}$ (7D_592498391) is nearby a previously identified QTL derived from Wangshuibai and Riband (Draeger et al., 2007; Jia et al., 2018).

FHB resistance is a quantitatively inherited trait and a single major QTL such as *Fhb1* do not provide sufficient protection against the disease (McMullen et al., 2012). Pyramiding multiple

major QTL while increasing the frequency of minor effects QTLs prevalent in the population is an effective strategy to improve FHB resistance in breeding. In the absence of a single major QTL, increasing the frequency of favorable alleles for many minor QTL will provide effective resistance. By reducing the threshold criteria $-logP = 3.0$, we were able to capture minor loci such that number of significant loci increased from 25 to 67. A significant correlation between the number of favorable alleles at over these 67 loci and the FHB traits showed that genetic gains can be made by accumulating favorable alleles of many minor-effects QTL, and showing that a lower density marker set could potentially be as efficient as large scale and genome-wide genotyping. The favorable alleles of sixty-seven markers identified in this study can be collectively targeted for marker-assisted selection to enhance FHB resistance in the breeding population.

Using SRWW population developed at Purdue University, we performed association mapping study to identify and characterize the genomic regions governing FHB resistance in the population. Altogether twenty-five loci were identified to be associated with FHB resistance. Of the twenty-five loci, 18 loci explained more than 1% variation in the phenotype in at least one environment. The SRWW breeding population at Purdue University consists of many QTLs derived from both native and exotic wheat lines. The major loci governing FHB resistance can be pyramided through MAS. Furthermore, this study shows that a number of minor QTLs govern FHB resistance and pyramiding QTLs from these diverse sources have potential to develop lines with efficient and long-term resistance to the devastating disease.

3.6   References

Appels, R., Eversole, K., Feuillet, C., Keller, B., Rogers, J., Stein, N., et al (2018). Shifting the limits in wheat research and breeding using a fully annotated reference genome. Science, 361(6403), eaar7191.

Arruda, M. P., Brown, P., Brown-Guedira, G., Krill, A. M., Thurber, C., Merrill, K. R., et al (2016). Genome-wide association mapping of *Fusarium* head blight resistance in wheat using genotyping-by-sequencing. The Plant Genome, 9(1).

Balut, A. L., Clark, A. J., Brown-Guedira, G., Souza, E., & Van Sanford, D. A. (2013). Validation of Fhb1 and QFhs. nau-2DL in several soft red winter wheat populations. Crop science, 53(3), 934-945.

Bates, D., Sarkar, D., Bates, M. D., & Matrix, L. (2007). The lme4 package. R package version, 2(1), 74.

Bourdoncle, W., & Ohm, H. W. (2003). Quantitative trait loci for resistance to *Fusarium* head blight in recombinant inbred wheat lines from the cross Huapei 57-2/Patterson. Euphytica, 131(1), 131-136.

Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., & Buckler, E. S. (2007). TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics, 23(19), 2633-2635.

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.

Buerstmayr, H., Lemmens, M., Hartl, L., Doldi, L., Steiner, B., Stierschneider, M., & Ruckenbauer, P. (2002). Molecular mapping of QTLs for *Fusarium* head blight resistance in spring wheat. I. Resistance to fungal spread (Type II resistance). Theoretical and Applied Genetics, 104(1), 84-91.

Buerstmayr, H., Steiner, B., Hartl, L., Griesser, M., Angerer, N., Lengauer, D., et al (2003). Molecular mapping of QTLs for *Fusarium* head blight resistance in spring wheat. II. Resistance to fungal penetration and spread. Theoretical and Applied Genetics, 107(3), 503-508.

Buerstmayr, M., Steiner, B., & Buerstmayr, H. (2019). Breeding for *Fusarium* head blight resistance in wheat—Progress and challenges. Plant Breeding.

Cainong, J. C., Bockus, W. W., Feng, Y., Chen, P., Qi, L., Sehgal, S. K., et al (2015). Chromosome engineering, mapping, and transferring of resistance to *Fusarium* head blight disease from *Elymus tsukushiensis* into wheat. Theoretical and Applied Genetics, 128(6), 1019-1027.

Clinesmith, M. A., Fritz, A. K., Lemes da Silva, C., Bockus, W. W., Poland, J. A., Dowell, F. E., & Peiris, K. H. (2019). QTL Mapping of *Fusarium* Head Blight Resistance in Winter Wheat Cultivars 'Art'and 'Everest'. Crop Science, 59(3), 911-924.

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.). Theoretical and Applied Genetics, 114(3), 429-437.

Draeger, R., Gosman, N., Steed, A., Chandler, E., Thomsett, M., Schondelmaier, J., et al (2007). Identification of QTLs for resistance to *Fusarium* head blight, DON accumulation and associated traits in the winter wheat variety Arina. Theoretical and Applied Genetics, 115(5), 617-625.

Eckard, J. T., Gonzalez-Hernandez, J. L., Caffe, M., Berzonsky, W., Bockus, W. W., Marais, G. F., & Baenziger, P. S. (2015). Native *Fusarium* head blight resistance from winter wheat cultivars 'Lyman,''Overland,''Ernie,'and 'Freedom'mapped and pyramided onto 'Wesley'-Fhb1 backgrounds. Molecular breeding, 35(1), 6.

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PloS one, 6(5).

Fuentes, R. G., Mickelson, H. R., Busch, R. H., Dill-Macky, R., Evans, C. K., Thompson, W. G., et al (2005). Resource allocation and cultivar stability in breeding for *Fusarium* head blight resistance in spring wheat. Crop science, 45(5), 1965-1972.

Gervais, L., Dedryver, F., Morlais, J. Y., Bodusseau, V., Negre, S., Bilous, M., et al (2003). Mapping of quantitative trait loci for field resistance to *Fusarium* head blight in an European winter wheat. Theoretical and Applied Genetics, 106(6), 961-970.

Gill, B.S., Friebe, B.R. and White, F.F. (2011). Alien introgressions represent a rich source of genes for crop improvement. Proceedings of the National Academy of Sciences, 108(19), pp.7657-7658.

Gilsinger, J., Kong, L., Shen, X., & Ohm, H. (2005). DNA markers associated with low *Fusarium* head blight incidence and narrow flower opening in wheat. Theoretical and Applied Genetics, 110(7), 1218-1225.

He, X., Singh, P. K., Duveiller, E., Dreisigacker, S., & Singh, R. P. (2013). Development and characterization of International Maize and Wheat Improvement Center (CIMMYT) germplasm for *Fusarium* head blight resistance. In *Fusarium* Head Blight in Latin America (pp. 241-262). Springer, Dordrecht.

Hoffstetter, A., Cabrera, A., & Sneller, C. (2016). Identifying quantitative trait loci for economic traits in an elite soft red winter wheat population. Crop Science, 56(2), 547-558.

Holder, A. (2018). A Genome Wide Association Study for *Fusarium* Head Blight Resistance in Southern Soft Red Winter Wheat.

Holzapfel, J., Voss, H. H., Miedaner, T., Korzun, V., Häberle, J., Schweizer, G., et al (2008). Inheritance of resistance to *Fusarium* head blight in three European winter wheat populations. Theoretical and Applied Genetics, 117(7), 1119-1128.

Islam, M. 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.

Jia, H., Zhou, J., Xue, S., Li, G., Yan, H., Ran, C., et al (2018). A journey to understand wheat *Fusarium* head blight resistance in the Chinese wheat landrace Wangshuibai. The Crop Journal, 6(1), 48-59.

Liu, S., Abate, Z. A., Lu, H., Musket, T., Davis, G. L., & McKendry, A. L. (2007). QTL associated with *Fusarium* head blight resistance in the soft red winter wheat Ernie. Theoretical and Applied Genetics, 115(3), 417-427.

Liu, S., Griffey, C. A., Hall, M. D., McKendry, A. L., Chen, J., Brooks, W. S., et al (2013). Molecular characterization of field resistance to *Fusarium* head blight in two US soft red winter wheat cultivars. Theoretical and applied genetics, 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 resistance in wheat. Crop Science, 49(6), 1955-1968.

Liu, X., Huang, M., Fan, B., Buckler, E. S., & Zhang, Z. (2016). Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies. PLoS genetics, 12(2).

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.

McMullen, M., Jones, R., & Gallenberg, D. (1997). Scab of wheat and barley: a re-emerging disease of devastating impact. Plant disease, 81(12), 1340-1348.

Mesterhazy, A. (1995). Types and components of resistance to *Fusarium* head blight of wheat. Plant breeding, 114(5), 377-386.

Mesterházy, Á., Bartók, T., Mirocha, C. G., & Komoroczy, R. (1999). Nature of wheat resistance to *Fusarium* head blight and the role of deoxynivalenol for breeding. Plant breeding, 118(2), 97-110.

Mirocha, C. J., Kolaczkowski, E., Xie, W., Yu, H., & Jelen, H. (1998). Analysis of deoxynivalenol and its derivatives (batch and single kernel) using gas chromatography/mass spectrometry. Journal of Agricultural and Food Chemistry, 46(4), 1414-1418.

Money, D., Gardner, K., Migicovsky, Z., Schwaninger, H., Zhong, G. Y., & Myles, S. (2015). LinkImpute: fast and accurate genotype imputation for nonmodel organisms. G3: Genes, Genomes, Genetics, 5(11), 2383-2390.

Paillard, S., Schnurbusch, T., Tiwari, R., Messmer, M., Winzeler, M., Keller, B., & Schachermayr, G. (2004). QTL analysis of resistance to *Fusarium* head blight in Swiss winter wheat (Triticum aestivum L.). Theoretical and Applied Genetics, 109(2), 323-332.

Parry, D. W., Jenkinson, P., & McLeod, L. (1995). *Fusarium* ear blight (scab) in small grain cereals—a review. Plant pathology, 44(2), 207-238.

Petersen, S., Lyerly, J. H., Maloney, P. V., Brown-Guedira, G., Cowger, C., Costa, J. M., et al (2016). Mapping of *Fusarium* head blight resistance quantitative trait loci in winter wheat cultivar NC-Neuse. Crop Science, 56(4), 1473-1483.

Petersen, S., Lyerly, J. H., McKendry, A. L., Islam, M. S., Brown-Guedira, G., Cowger, C., et al (2017). Validation of *Fusarium* head blight resistance QTL in US winter wheat. Crop Science, 57(1), 1-12.

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).

Pritchard, J. K., Wen, W., & Falush, D. (2003). Documentation for STRUCTURE software: Version 2.

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. Theoretical and Applied Genetics, 117(7), 1155-1166.

Schroeder, H. W., & Christensen, J. J. (1963). Factors affecting resistance of wheat to scab caused by *Gibberella zeae*. Phytopathology, 53(7, 1), 831-838.

Shen, X., & Ohm, H. (2007). Molecular mapping of Thinopyrum-derived *Fusarium* head blight resistance in common wheat. Molecular Breeding, 20(2), 131-140.

Shen, X., Ittu, M., & Ohm, H. W. (2003). Quantitative trait loci conditioning resistance to *Fusarium* head blight in wheat line F201R. Crop Science, 43(3), 850-857.

Sobrova, P., Adam, V., Vasatkova, A., Beklova, M., Zeman, L., & Kizek, R. (2010). Deoxynivalenol and its toxicity. Interdisciplinary toxicology, 3(3), 94-99.

Srinivasachary, N.G., Steed, A., Hollins, T. W., Bayles, R., Jennings, P., & Nicholson, P. (2009). Semi-dwarfing *Rht-B1* and *Rht-D1* loci of wheat differ significantly in their influence on resistance to *Fusarium* head blight. Theoretical and applied genetics, 118(4), 695.

Steiner, B., Buerstmayr, M., Wagner, C., Danler, A., Eshonkulov, B., Ehn, M., & Buerstmayr, H. (2019). Fine-mapping of the *Fusarium* head blight resistance QTL *Qfhs. ifa-5A* identifies two resistance QTL associated with anther extrusion. Theoretical and applied genetics, 132(7), 2039-2053.

Su, Z., Bernardo, A., Tian, B., Chen, H., Wang, S., Ma, H., et al (2019). A deletion mutation in *TaHRC* confers Fhb1 resistance to *Fusarium* head blight in wheat. Nature genetics, 51(7), 1099-1105.

Team, R. C. (2013). R: A language and environment for statistical computing.

Tessmann, E. W., Dong, Y., & Van Sanford, D. A. (2019). GWAS for *Fusarium* head blight traits in a soft red winter wheat mapping panel. Crop Science, 59(5), 1823-1837.

USDA-ERS. 2020. Wheat data. USDA Econ. Res. Serv. https://www.ers.usda.gov/topics/crops/wheat/wheat-sector-at-a-glance/#trade. Accessed on 15 Jan. 2020.

USDA-NASS. 2017. Quick stats. USDA Natl. Agric. Stat. Serv. https://www.nass.usda.gov/Statistics_by_Subject/result.php?CE7983DA-D56B-3F1C-9D4D-3D2E4DC8B381&sector=CROPS&group=FIELD%20CROPS&comm=WHEAT. Accessed on 6 Jan. 2020.

Wang, H., Sun, S., Ge, W., Zhao, L., Hou, B., Wang et al., (2020). Horizontal gene transfer of Fhb7 from fungus underlies Fusarium head blight resistance in wheat. Science.

Ward, B. P., Brown-Guedira, G., Kolb, F. L., Van Sanford, D. A., Tyagi, P., Sneller, C. H., & Griffey, C. A. (2019). Genome-wide association studies for yield-related traits in soft red winter wheat grown in Virginia. PloS one, 14(2).

Wright, E. E. (2014). Identification of Native FHB Resistance QTL in the SRW Wheat Cultivar Jamestown (Doctoral dissertation, Virginia Tech).

Würschum, T., Jähne, F., Phillips, A. L., Langer, S. M., Longin, C. F. H., Tucker, M. R., & Leiser, W. L. (2020). Mis-expression of a transcriptional repressor candidate provides a molecular mechanism for the suppression of awns by *Tipped 1* in wheat. Journal of Experimental Botany.

Yang, Z., Gilbert, J., Fedak, G., & Somers, D. J. (2005). Genetic characterization of QTL associated with resistance to *Fusarium* head blight in a doubled-haploid spring wheat population. Genome, 48(2), 187-196.

Zhang, X., Bai, G., Bockus, W., Ji, X., & Pan, H. (2012). Quantitative trait loci for *Fusarium* head blight resistance in US hard winter wheat cultivar Heyne. Crop science, 52(3), 1187-1194.

CHAPTER 4. CONCLUSIONS AND FUTURE DIRECTIONS

Wheat occupies almost one-fifth of the world's cultivated land and is a major staple food for almost two-thirds of world population (~ 2.5 billion people from 89 different countries). To meet the growing demands of food and uncertainty brought by climate change, more responsibility is shouldered upon plant breeders to develop high-yielding, climate-resilient wheat varieties. In the US, wheat production has declined over the past two decades despite its significance in the United States. Nevertheless, plant breeders have made tremendous progress during the last century using conventional breeding methodology in combination with MAS. The advancement in next generation sequence methodologies and statistical tools to perform GWAS and genomic selection open avenues for plant breeders to make higher genetic gains by reducing the time and resources required to develop high-performing cultivars. However, a prerequisite of a targeted breeding program is to characterize the germplasm for important traits and for genomic structure, which was the goal of this PhD dissertation. In this PhD dissertation, I characterized the soft red winter wheat lines developed at Purdue University by means of genome wide marker development, population genomics methods, and field-based phenotyping for grain yield and FHB traits.

Genome-wide markers provide valuable information to understand genomic architecture and hidden population structures of crop populations. Using genotyping-by-sequencing (GBS) methodology (Elshire et al., 2011), I genotyped 436 Purdue-bred SRWW lines and generated 14,907 single nucleotide polymorphism (SNP) markers. LD decay analysis showed that two chromosomes 2B and 7D showed extended low recombining regions possibly due to alien chromatins integration. STRUCTURE and PCA analyses revealed there are four sub-populations in the germplasm. These subpopulations reveal historical patterns of breeding and introduction of adapted and exotic lines into the breeding program. Pedigree evaluation suggested that two subpopulations, INW and TRUMAN represents clusters of lines that are originated from lines adapted to the SRWW growing regions of the United States. Similarly, the remaining two sub-populations, WHEATEAR and NING, were most likely result of introduction of exotic lines NING and WHEATEAR into the breeding program for disease resistance. Furthermore, by scanning the whole genome using smoothed F_{ST} and hapFLK methods, genomic regions associated with selection history were identified. Altogether, 13 regions were identified with 10 regions have

implications in environmental adaptation, disease resistance, and end use quality were identified. This study provides a foundation to formulate biological hypothesis about selection process in winter wheat and associated candidate genes for further investigation. For example, in INW sub-population, a region in chromosome 1D (12.8 to 15.3 Mbp) depicted selection signatures and seems to be coinciding with a QTL for low-temperature tolerance (LT)T in the similar region (12.2-328.4 Mbp). I identified a gene Transducin/WD40 repeat protein (*TraesCS1D01G034200*) located in the region that showed 79.6% identity to Arabidopsis *HOS15* (NCBI refseq: NP_178182.1), known to be important for cold tolerance.

I evaluated three hundred and ninety-two lines for yield traits in field across two years in two location. I collected phenotypic data for days to heading (DH), plant height (PH), test weight (TW), grain yield (YLD), number of spikes per unit area (NS), and number of kernels per unit area (NK). Altogether, twenty-seven genomic regions were associated with these traits except YLD with seven of them significant at 5% FDR criteria. The seven loci included MTAs 7B_8885441 for DH, 1A_494799791, 4A_41973900, 5B_455980623, and 6A_27020312 for PH, 4B_3295544 for TW, 1A_549422405 for NS, and 5A_456261322 for NK.

I also evaluated the germplasm for FHB resistance in misted FHB nursery across two years in three sets of experiments. Phenotypic data were collected for disease incidence (INC), disease severity (SEV), FHB index (FHBdx), *Fusarium* damaged kernels (FDK), and deoxynivalenol concentration (DON). GWAS identified twenty-five genomic regions associated with FHB resistance of which seven regions explained significant proportion of phenotypic variation (>5%). The 67 SNP information obtained in this study can be used to make breeder-friendly markers to be used in MAS for improvement of FHB resistance. In addition, with the availability of the annotated reference genome of wheat (Appels et al., 2018), it is possible to narrow-down the list of candidate genes associated with the traits.

Genomic selection increases selection accuracy, and genetic gains per unit of time. In this study, I did not identify any significant association with YLD. Preliminary assessment of genomic prediction model showed that genomic selection could be a viable strategy for improving yield. The genotype and phenotype data produced in this study will be valuable to train genomic prediction models and study the optimal design of genomic selection training sets.

This study laid foundation for the design and breeding decisions to increase the efficiency of pyramiding strategies and achieving transgressive segregation. Crossing parents with favorable alleles at multiple loci should increase the probability of fixing these alleles in the progenies. In addition, the parental combination should produce progeny with greater genetic variances.

Learnings from population genomics and field-based phenotypes are specifically important in the design and expectations from the breeding populations. An illustration of using this genomic and field-based study for selecting parents for breeding is described here. Using the allele information and phenotypic data, I selected one hundred lines from the population that show good FHB resistance traits and grain yield > 6 tons ha^{-1}. To avoid crossing similar lines, I clustered these 100 lines based on SNP marker data using K-means clustering method. The sum of squares of distances within clusters was assessed from 2 to 15. The within cluster sum of squares dropped sharply until six and roughly plateaued (Figure 4.1). Therefore, the germplasm was classified into 6 groups (Figure 4.2). An ideal line is early flowering, semi-dwarf, have higher test weight and FHB resistance, and consists favorable alleles across major loci. As an example, I selected two lines PU209(05262A1-5-9-35-1) and PU108 (10534A1-17-17) as parents for crossing. In general, both of the lines showed desirable agronomic performance and FHB resistance. Line PU209 show good agronomic traits such as early flowering (~137 days), higher yield (6.4 ton ha^{-1}) and test weight (1049.81 kg m^{-3}) and good height (75 cm). However, it exhibited low FHB resistance (INC = 80% and SEV = 40%, FDK =9.5%). In contrast, line PU108 performed slightly lower than PU209 agronomically with days to heading of ~140 days, yield of 6 ton ha^{-1} and test weight of (~981 kg m^{-3}) but did good in terms of FHB resistance (INC = 64%, SEV = 34%, FDK = 7.4%). Line PU209 was homozygous for *Fhb1* allele however did not possessed *Q2B.1* QTL for FHB resistance whereas line PU108 does not have the *Fhb1* allele but is homozygous for favorable allele at *Q2B.1*. Both of the lines are homozygous for *Rht-B1b* (dwarf) and *Ppd-B1a* (photoperiod insensitive) alleles. In addition, K-means clustering assigned these lines to cluster 3 and 6 suggesting, based on genome-wide markers, they are distantly related and thus is expected to generate progeny with large genetic variance.
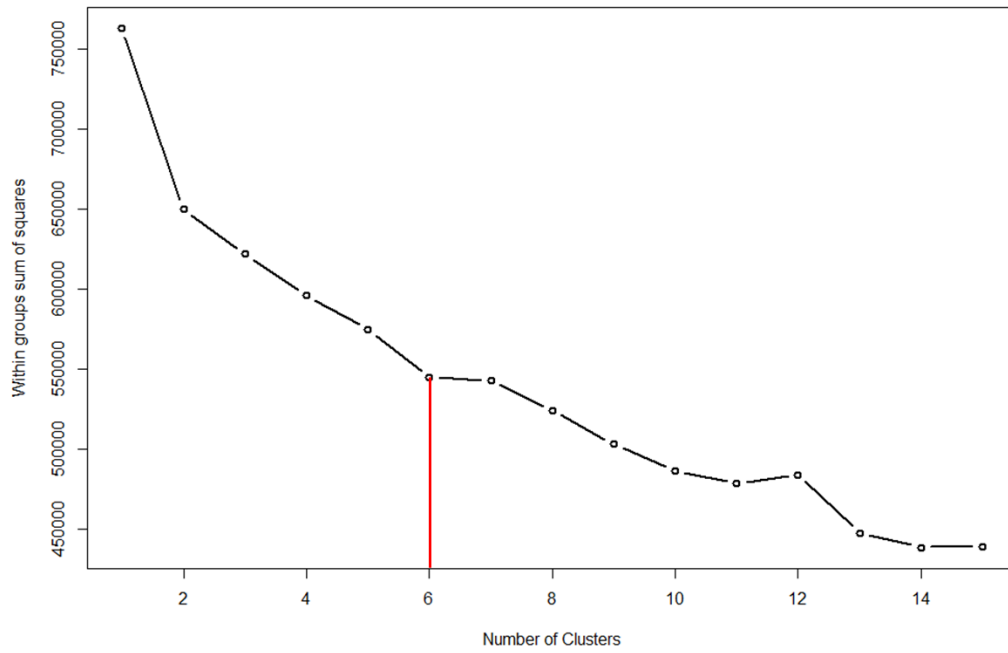
Figure 4.1 Scree plot showing decrease in sum of squares decreasing as the number of cluster increases. The within cluster sum of squares dropped sharply until cluster 6 and then roughly plateaued.
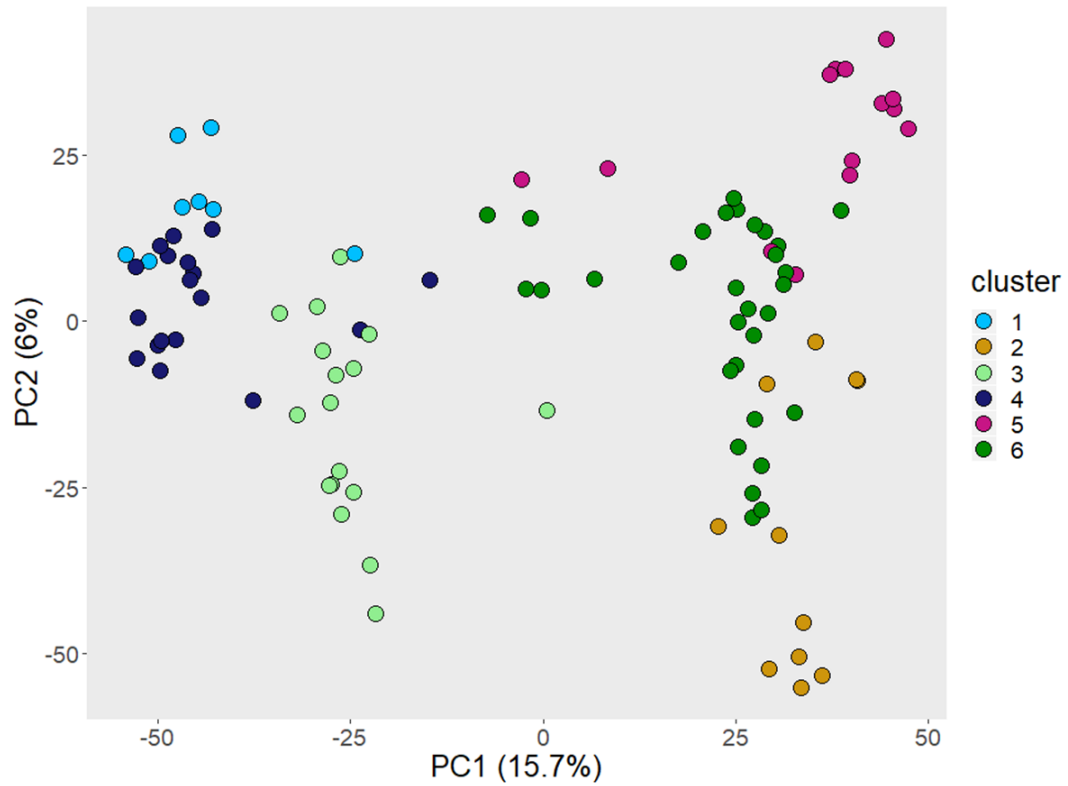
Figure 4.2. The six clusters identified by K-means clustering method in relation to their first two principal components.

APPENDIX A. SUPPLEMENTAL INFORMATION OF CHAPTER 2.

A1 Distribution of SNPs across the sub-genomes and the 21 chromosomes of hexaploid wheat.

A2. The table shows the LD decay distances for individual chromosomes. The chromosomes 2B and 7D with unusually high LD decay rates are underlined.

A3. The heatmap of linkage disequilibrium for chromosomes 2B (left) and 7D (right), showing extended islands of high-LD.

A4. Genomic population tree based on Reynold's distance matrix of lines, where all markers were used to derive distances.

A5. Manhattan plots showing the *-logP* values based on the hapFLK method. The horizontal line shows the cut-off at *-logP* value of 2 (P-value = 0.01).

A6. Summary of descriptive statistics and heritability of the six agronomic traits.

A7. Histograms (diagonals), pairwise scatterplots (lower triangle from the diagonal of histograms), and the pairwise Pearson's correlations (upper triangle from the diagonal of histograms) are presented. The line through the scatterplot is the least squares fit. The asterisks show significance levels at 0.05 (*), 0.01 (**), and 0.001 (***).
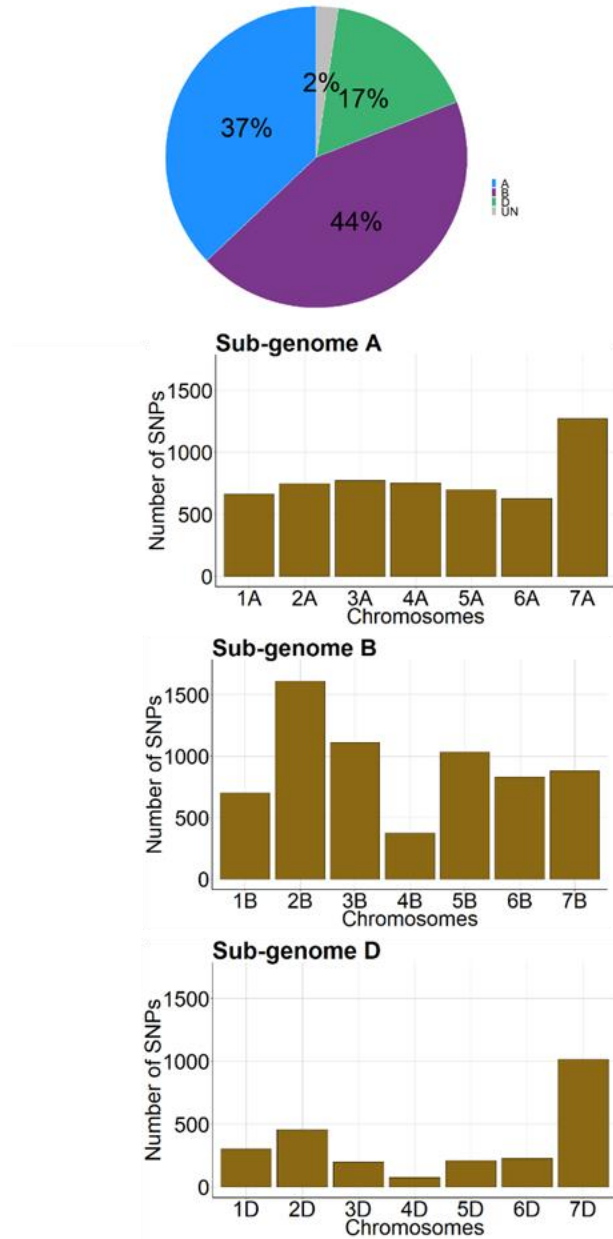
A1. Distribution of SNPs across the sub-genomes and the 21 chromosomes of hexaploid wheat.

A2. The table shows the LD decay distances for individual chromosomes. The chromosomes 2B and 7D with unusually high LD decay rates are underlined.

Chromosome	Half decay distance (Mbp)
1A	2.64
1B	6.10
1D	4.97
2A	3.01
2B	<u>763</u>
2D	12.93
3A	1.93
3B	3.15
3D	1.88
4A	2.77
4B	5.42
4D	0.64
5A	4.69
5B	5.37
5D	1.68
6A	2.73
6B	3.82
6D	0.77
7A	1.69
7B	3.05
7D	<u>94.32</u>
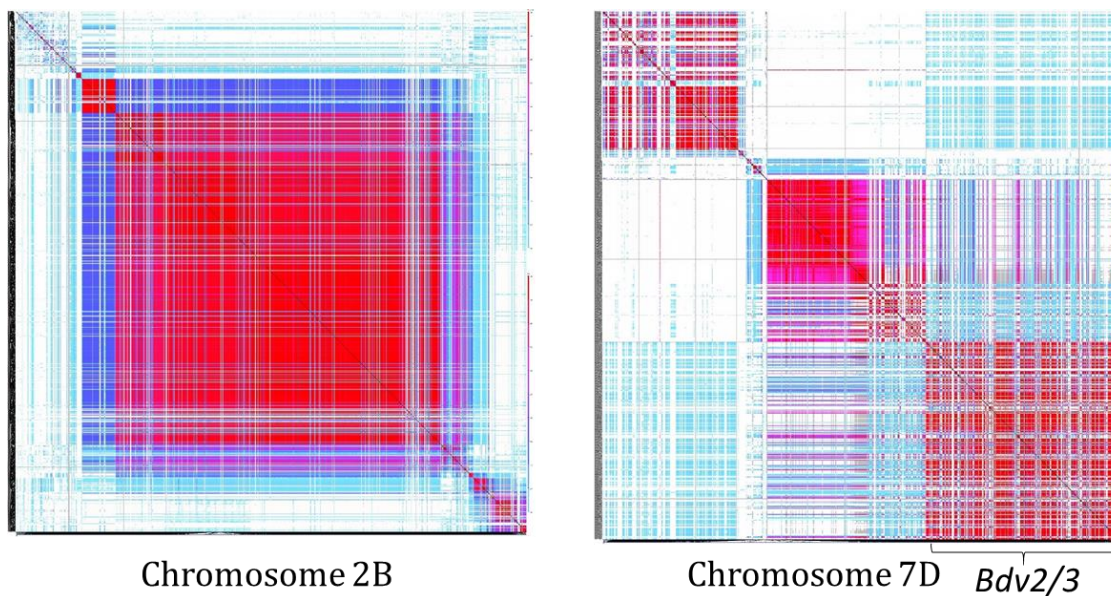
Chromosome 2B   Chromosome 7D   *Bdv2/3*

A3. The heatmap of linkage disequilibrium for chromosomes 2B (left) and 7D (right), showing extended islands of high-LD.
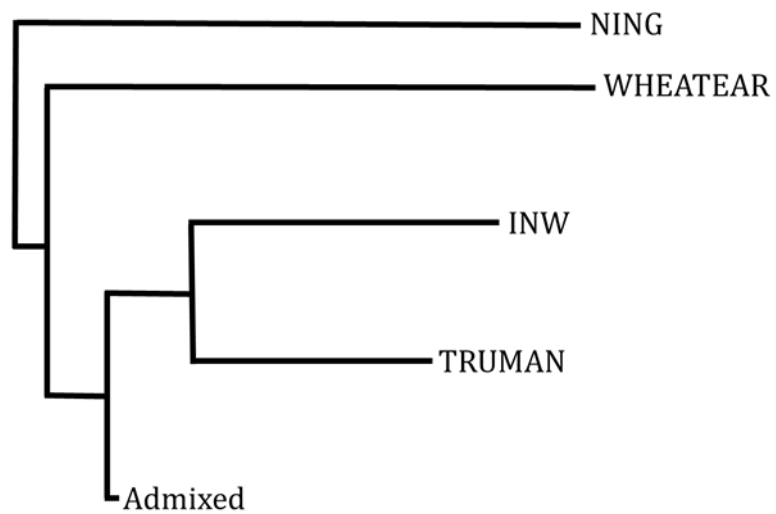
A4. Genomic population tree based on Reynold's distance matrix of lines, where all markers were used to derive distances.
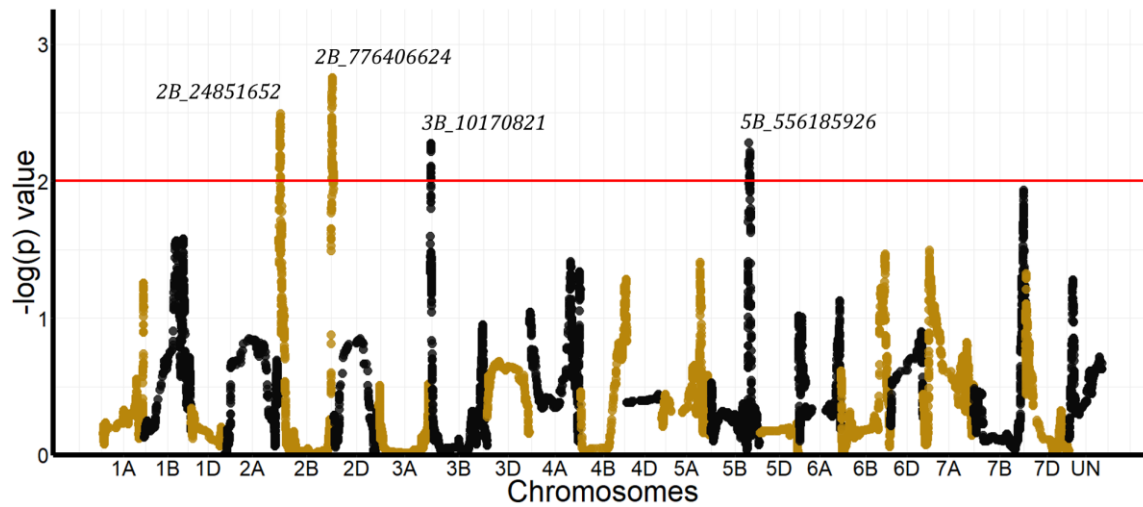
A5. Manhattan plots showing the *-logP* values based on the hapFLK method. The horizontal line shows the cut-off at *-logP* value of 2 (P-value = 0.01).

A6. Summary of descriptive statistics and heritability of the six agronomic traits.

Traits	Units	Mean	SD	Min	Max	H^2	H^2_C
Days to heading	Julian days	140.1	1.8	131.7	147.5	0.74	0.69
Plant height	Cm	78.5	7.8	55.4	124.7	0.84	0.78
Test weight	Kg m^{-3}	931.9	112.7	463.5	1282	0.49	0.41
Grain yield	Ton ha^{-1}	4.4	1.1	1	9	0.54	0.49
#Spikes per unit area	Spikes m^{-2}	710.7	121.4	337.8	1081.3	0.52	0.39
#Kernels per unit area	Kernels m^{-2}	15019.4	4530.1	2497.6	32378.6	0.4	0.27

H^2: broad sense heritability estimated on entry mean, H^2c is the heritability measure following Cullis et al., (2006).
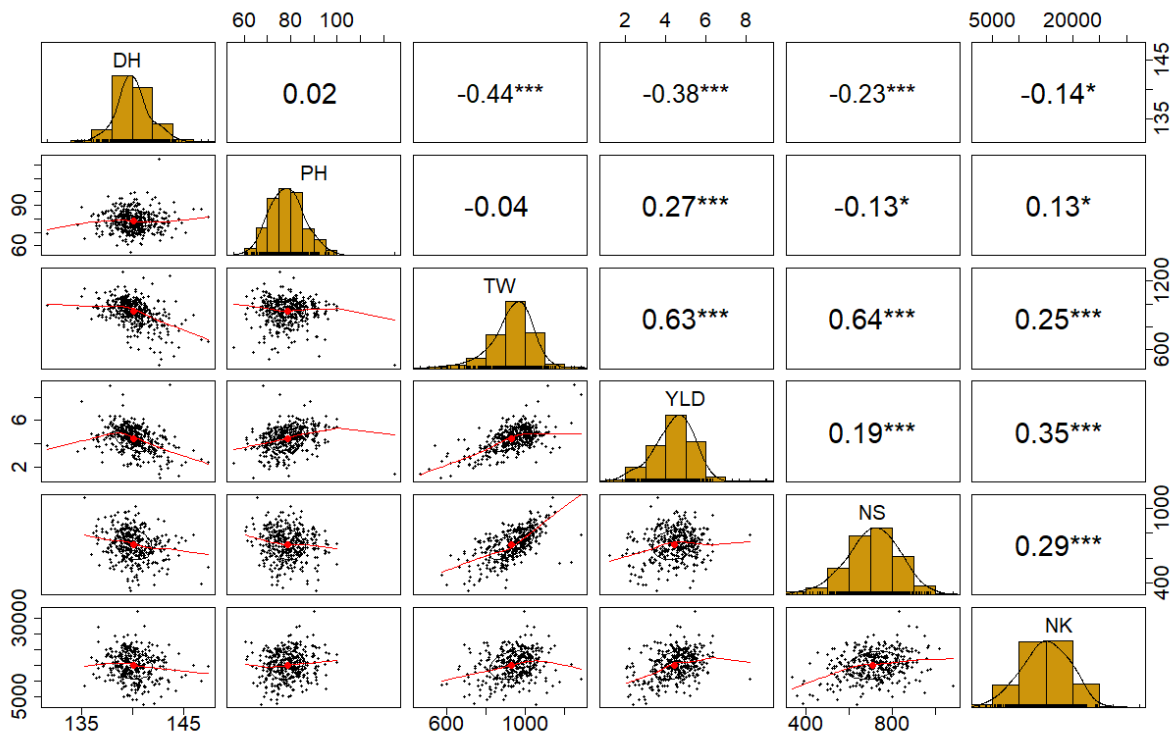
A7. Histograms (diagonals), pairwise scatterplots (lower triangle from the diagonal of histograms), and the pairwise Pearson's correlations (upper triangle from the diagonal of histograms) are presented. The line through the scatterplot is the least squares fit. The asterisks show significance levels at 0.05 (*), 0.01 (**), and 0.001 (***).

APPENDIX B. SUPPLEMENTAL INFORMATION OF CHAPTER 3.

B1. Mean and standard deviations of the four check cultivars across the experimental sets.

B2. List of names, allelic polymorphism, chromosomal assignment, and physical position of the 67 markers chosen on the basis of –logP > 3 criteria for ordinary least square model fitting and marker-assisted selection.

B3. List and BLUE estimates of FHB phenotypes for twenty lines with more and 20 lines with less favorable alleles. T-test revealed that the mean difference in the phenotypes between the groups is highly significant (p <0.001).

B4. Summary of simple linear regression with number of favorable alleles as independent variable and phenotypes as response variable.

B1. Mean and standard deviations of the four check cultivars across the experimental sets.

Check name	Site/rep	INC	SEV	FHBdx	FDK	DON
Patterson	ACRE18	97.6±5.1	83.8±13.6	82.1±15.3	27.7±13.2	4.1±1.4
	ACRE19_I	90±13.7	72.2±16	66.6±21.1	34.7±10.7	14±4.9
	ACRE19_II	89.8±6.8	72.3±12.2	65.3±13.8	34±11.3	13.1±2.5
	Mean	92.3±9.9	75.8±14.8	71±18.4	32.4±11.9	10.8±5.4
Monon	ACRE18	88.3±9.7	61.5±17.4	55.5±19.9	25±16.8	1.6±1.1
	ACRE19_I	70.6±13.8	37.2±12.7	27±12.3	28.8±12.2	6.8±2.9
	ACRE19_II	78±12.9	45.5±14.9	35.8±13.1	34.3±10	8.3±2.4
	Mean	78.9±14.1	48±17.9	39.3±19.2	29.5±13.5	5.7±3.7
INW0411	ACRE18	94.2±9.4	66.2±16	62.9±18.2	20.4±14.7	1.7±0.7
	ACRE19_I	82.3±12	41.2±12	34.5±13.7	21.4±9.2	6.8±2.1
	ACRE19_II	85.5±10.1	36.8±11.7	31.9±11.4	25±9.3	6.1±2.1
	Mean	87±11.5	47.3±18.3	42.2±19.9	22.5±11	5±2.9
INW0412	ACRE18	60.8±15.3	29.1±7	18.1±7.8	21.8±13.5	1.5±0.8
	ACRE19_I	59.8±16.4	28.6±12.8	17.4±9.6	28.7±8.2	9.1±4.2
	ACRE19_II	66.3±11	20.9±6.9	14.2±6.6	28.9±11.8	10.1±2.8
	Mean	62.3±14.5	26.2±10	16.6±8.1	26.6±11.6	7±4.8

B2. List of names, allelic polymorphism, chromosomal assignment, and physical position of the 67 markers chosen on the basis of –logP > 3 criteria for ordinary least square model fitting and marker-assisted selection.

Marker names	Alleles	Chromosome	Physical position
1A_23000344	C/A	1A	23000344
1A_499864143	A/C	1A	5E+08
1B_628959667	G/A	1B	6.29E+08
1D_249560681	G/A	1D	2.5E+08
2A_224906132	G/A	2A	2.25E+08
2A_695316061	G/A	2A	6.95E+08
2A_758950193	C/G	2A	7.59E+08
2B_65370227	C/T	2B	65370227
2B_226687379	C/T	2B	2.27E+08
2B_787820223	A/G	2B	7.88E+08
2D_521578938	A/G	2D	5.22E+08
2D_645828400	T/C	2D	6.46E+08
3A_15462580	C/T	3A	15462580
3A_697615363	C/T	3A	6.98E+08
3A_703914260	A/G	3A	7.04E+08
3A_712302895	T/C	3A	7.12E+08
3A_721301853	A/G	3A	7.21E+08
3B_11289023	A/G	3B	11289023
3B_26974711	C/G	3B	26974711
3B_756667955	G/A	3B	7.57E+08
4A_47393653	T/C	4A	47393653
4A_598226912	A/G	4A	5.98E+08
4A_605557514	C/T	4A	6.06E+08
4A_614026054	G/A	4A	6.14E+08
4A_649885899	A/T	4A	6.5E+08
4A_654871135	C/A	4A	6.55E+08
4A_691467965	C/G	4A	6.91E+08
4A_728086043	C/T	4A	7.28E+08
4B_622455	A/G	4B	622455
4B_7263114	T/G	4B	7263114
4B_594693911	A/G	4B	5.95E+08
4B_653970542	C/T	4B	6.54E+08
4D_450312303	G/T	4D	4.5E+08
5A_19682318	C/A	5A	19682318
5A_55385530	C/G	5A	55385530

5A_489912043	T/C	5A	4.9E+08
5A_559517035	A/G	5A	5.6E+08
5A_561121102	G/T	5A	5.61E+08
5A_695581531	A/G	5A	6.96E+08
5A_702988371	G/A	5A	7.03E+08
5B_283024388	G/A	5B	2.83E+08
5B_501957243	A/G	5B	5.02E+08
5B_553933522	G/A	5B	5.54E+08
5B_576896245	T/C	5B	5.77E+08
5B_668461278	G/A	5B	6.68E+08
6A_39207876	C/T	6A	39207876
6A_65844428	C/G	6A	65844428
6A_559342209	C/G	6A	5.59E+08
6B_8798063	G/T	6B	8798063
6B_157666111	G/C	6B	1.58E+08
6B_508844418	T/C	6B	5.09E+08
7A_4745932	G/C	7A	4745932
7A_39064682	C/T	7A	39064682
7A_40101095	G/C	7A	40101095
7A_516508921	C/T	7A	5.17E+08
7A_705490416	C/T	7A	7.05E+08
7A_727710638	C/G	7A	7.28E+08
7B_41324506	C/G	7B	41324506
7B_60855111	C/G	7B	60855111
7B_126787326	C/G	7B	1.27E+08
7B_539196314	C/T	7B	5.39E+08
7B_708860224	A/C	7B	7.09E+08
7B_720700818	T/A	7B	7.21E+08
7B_723092825	G/A	7B	7.23E+08
7D_26720068	T/G	7D	26720068
7D_113792653	T/C	7D	1.14E+08
7D_592498391	A/G	7D	5.92E+08

B3. List and BLUE estimates of FHB phenotypes for twenty lines with more and 20 lines with less favorable alleles. T-test revealed that the mean difference in the phenotypes between the groups is highly significant (p <0.001).

Genotype	INC	SEV	FHBdx	FDK	DON	#favorable alleles	Type
07469A1-6-1-3	82.19	40.33	37.95	17.62	2.72	90	more
10531A1-4-1	46.50	32.81	17.00	25.62	5.46	90	more
10531A1-14-42	56.01	23.58	16.13	9.20	4.08	88	more
10523RA1-22-4	76.62	43.70	35.78	26.62	7.09	88	more
10534A1-17-10	70.26	41.08	32.56	4.94	5.27	87	more
03528A1-10-6-7-2-1	73.06	51.04	43.37	9.50	3.24	87	more
10641B1-13-15-3-05	71.70	16.24	10.56	16.55	4.89	85	more
10531A1-8-3	64.21	29.53	22.51	28.81	5.95	85	more
05247A1-7-7-3-1	66.65	30.41	20.72	NA	9.25	84	more
10531A1-14-24	68.19	26.66	19.86	13.86	3.85	84	more
10535A1-15-25	53.85	29.47	17.02	6.42	3.00	84	more
10534A1-17-5	60.44	23.91	14.32	11.00	5.97	84	more
05251A1-1-77-68	77.47	40.16	33.19	18.60	5.76	84	more
05269A1-4-9-13	83.21	43.24	37.66	5.11	3.68	84	more
10535A1-15-4	73.44	25.56	19.31	9.92	5.07	84	more
10641B1-9-8-4-02	79.32	37.07	30.67	18.50	6.80	84	more
05247A1-7-7-2-3-3-4	69.53	30.87	22.46	15.96	4.78	83	more
10523RA1-21-10	48.21	45.08	20.87	14.84	4.79	83	more
INW0412	62.16	27.81	18.28	24.45	6.74	83	more
10534A1-17-17	67.30	34.46	27.06	7.40	5.27	83	more
04620A1-1-7-4-17	82.75	61.21	51.37	15.29	5.73	42	less
0527A1-9-9-2-4	91.68	61.45	58.20	39.75	9.47	47	less
0527A1-9-10-1-15	79.30	47.59	39.24	24.62	8.89	47	less
10649RB1-4-17-3-01	96.28	65.94	63.67	32.34	8.85	49	less
05251A1-1-77-12	76.36	46.30	40.70	30.41	6.43	51	less
0570A1-2-39-2-1-1	84.85	53.15	43.55	21.44	7.02	52	less
0570A1-2-39-2-4	75.02	57.63	52.39	31.78	7.55	52	less
05262A1-5-9-31-1	82.71	31.17	26.57	16.68	6.33	54	less
04620A1-1-17-4-17	87.81	41.94	36.29	17.18	6.23	54	less
0570A1-2-58	89.54	44.39	39.81	24.21	9.87	56	less
04620A1-1-7-4-10	85.89	31.73	27.50	32.01	9.40	56	less
961341A3-1-2	84.60	50.96	43.17	28.35	8.54	56	less
961341A3-1-4-1	83.57	63.89	53.44	38.44	10.22	57	less
011007A1-14-16-192	71.00	31.42	24.36	25.98	7.58	58	less
0527A1-9-9-2-31	83.78	44.28	38.30	22.84	7.94	58	less

10523RA1-23-2	76.37	35.97	28.64	46.18	21.85	58	less
0537A1-7-8-22	87.79	44.63	38.76	34.00	9.28	58	less
961341A3-1-8-2	76.13	47.84	39.09	30.27	11.02	58	less
011007A1-14-16-50	89.98	48.41	44.24	26.76	6.78	58	less
05251A1-1-77-32	67.76	37.80	29.62	25.72	6.34	58	less

B4. Summary of simple linear regression with number of favorable alleles as independent variable and phenotypes as response variable.

Trait_ENV	R^2	Slope	Pearson's r
INC_Y1	6.6	-0.65	-0.26
INC_Y2	11.8	-0.56	-0.35
INC_BLUE	15.7	-0.6	-0.4
SEV_Y1	6.2	-0.73	-0.25
SEV_Y2	14.2	-0.63	-0.38
SEV_BLUE	16.4	-0.67	-0.41
FHBdx_Y1	7	-0.86	-0.27
FHBdx_Y2	15.6	-0.71	-0.4
FHBdx_BLUE	19.1	-0.76	-0.44
FDK_Y1	3	-0.35	-0.18
FDK_Y2	3.6	-0.33	-0.2
FDK_BLUE	8.4	-0.39	-0.3
DON_Y1	6.2	-0.06	-0.25
DON_Y2	5.6	-0.12	-0.24
DON_BLUE	7.6	-0.1	-0.28