University of Arkansas, Fayetteville
ScholarWorks@UARK

Theses and Dissertations

8-2018

Association Mapping and Genomic Selection for Yield and Agronomic Traits in Soft Winter Wheat

Dennis Bulusan Lozada
University of Arkansas, Fayetteville

Follow this and additional works at: https://scholarworks.uark.edu/etd

Part of the Agronomy and Crop Sciences Commons, Plant Biology Commons, and the Plant Breeding and Genetics Commons

Recommended Citation
Lozada, Dennis Bulusan, "Association Mapping and Genomic Selection for Yield and Agronomic Traits in Soft Winter Wheat" (2018). *Theses and Dissertations*. 2941.
https://scholarworks.uark.edu/etd/2941

This Dissertation is brought to you for free and open access by ScholarWorks@UARK. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of ScholarWorks@UARK. For more information, please contact scholar@uark.edu, ccmiddle@uark.edu.

Association Mapping and Genomic Selection for Yield and Agronomic Traits in Soft Winter Wheat

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy in Cell and Molecular Biology

by

Dennis Nicuh Lozada
University of the Philippines - Los Baños
Bachelor of Science in Biology, 2010

August 2018
University of Arkansas

This dissertation is approved for recommendation to the Graduate Council.

____
Richard Esten Mason, Ph.D.
Dissertation Director

____
Pengyin Chen, Ph. D.
Committee Member

____
Ainong Shi, Ph. D.
Committee Member

____
Andy Pereira, Ph. D.
Committee Member

Abstract

Tools such as genome-wide association study (GWAS) and genomic selection (GS) have expedited the development of crops with improved genetic potential. While GWAS aims to identify significant markers associated with a trait of interest, the goal of GS is to utilize all marker effects to predict the performance of new breeding lines prior to testing. A GWAS for grain yield (GY), yield components, and agronomic traits was conducted using a diverse panel of 239 soft winter wheat (SWW) lines evaluated in eight site-years in Arkansas and Oklahoma. Broad sense heritability of GY (H^2=0.48) was moderate compared to other traits including plant height (H^2=0.81) and kernel weight (H^2=0.77). Markers associated with multiple traits on chromosomes 1A, 2D, 3B, and 4B serve as potential targets for marker assisted breeding to select for GY improvement. Validation of GY-related loci using spring wheat from the International Maize and Wheat Improvement Center (CIMMYT) in Mexico confirmed the effects of three loci in chromosomes 3A, 4B, and 6B. Lines possessing the favorable allele at all three loci (A-C-G allele combination) had the highest mean GY of possible haplotypes. The same population of 239 lines was used in a GS study as a training population (TP) to determine factors that affect the predictability of GY. The TP size had the greatest effect on predictive ability across the measured traits. Adding covariates in the GS model was more advantageous in increasing prediction accuracies under single population cross validations than in forward predictions. Forward validation of the prediction models on two new populations resulted in a maximum accuracy of 0.43 for GY. Genomic selection was “superior” to marker-assisted selection in terms of response to selection and combining phenotypic selection with GS resulted in the highest response. Results from this study can be used to accelerate the process of GY improvement and increase genetic gains in wheat breeding programs.

©2018 by Dennis Nicuh Lozada
All Rights Reserved

Acknowledgments

I would like to thank my adviser, Dr. Richard Esten Mason for giving me the opportunity to study under his supervision. I will always be grateful for his support from the very beginning of my graduate school journey up to the end.

To my Committee Members, Dr. Pengyin Chen, Dr. Andy Pereira, and Dr. Ainong Shi- for their insights during the writing process.

To our collaborators from the International Maize and Wheat Improvement Center (CIMMYT) in Mexico, Dr. Susanne Dreisigacker, Dr. Sivakumar Sukumaran, and the Wheat Molecular Breeding Laboratory- CIMMYT. *Gracias por todo*!

To Dr. Eduard Runge and the Monsanto Beachell-Borlaug International Scholars Program for funding my research in Mexico and my study at the University of Arkansas.

To the Wheat Breeding and Genetics staff and students at the University of Arkansas, Randy, Mr. Dave, Mr. Peter, Andrea, Amanda, Habib, Zach, Jamison, and Dylan- for their help in conducting my experiments and for the encouragements.

To the Chi Alpha Student Ministries for their prayers and friendship. Special thanks to Tim, Franckie, Juanchi, Tony, and Jossie. I also would like to thanks my good friend Lanre.

To my family- *Papang*, *Mamang*, *Manang*, *Manong* and *Ate* Princess. Thank you for the unconditional love, prayers, and support. Thank you for reminding me that I am and will never be on this adventure alone.

And *lastly* to Marvelyn, for believing that we can make it to this point.

"For from Him and to Him and through Him are all things…" Rom. 11:36

Dedication

To *Pa*, *Ma*, *Manang*, *Manong,*
and to Marvelyn…

Table of Contents

Page

CHAPTER I Introduction and Literature Review……………………………........................1

Wheat……………………………………………………………………………............2

Wheat genetics and genome sequencing………………………………………………...3

The need to increase wheat production…………………………………………………4

Marker-assisted breeding………………………………………………………………..4

Molecular markers for marker-assisted selection………………………………………7

Genetic factors controlling growth habit in wheat……………………………………....8

Vernalization genes……………………………………………………………...8

Photoperiod genes………………………………………………………………11

Reduced height genes…………………………………………………………...13

Effects of major growth habit genes on yield and adaptation of wheat………………...14

Genome-wide association study………………………………………………………...17

Genotyping-by-sequencing ……………………………………………………………19

Genomic selection………………………………………………………………………21

Approach of the current study…………………………………………………………..22

References………………………………………………………………………………24

CHAPTER II Genome-wide association study for grain yield and agronomic traits in soft winter wheat………………………………………………………….............39

Abstract…………………………………………………………………………………40

Introduction……………………………………………………………………………..41

Materials and Methods…………………………………………………………………42

Plant material and experimental design………………………………………...42

Trait measurements…………………………………………………………..43

Statistical analysis……………………………………………………………44

SNP marker genotyping……………………………………………………...44

Linkage disequilibrium, population structure, and genetic
diversity……………………………………………………………………45

GWAS for grain yield and agronomic traits…………………………………47

Results……………………………………………………………………………48

Genotype-by-environment interactions and heritability………………………..48

Principal components analyses (PCA) and phenotypic correlations…………...49

Analysis of LD…………………………………………………………………49

Population structure……………………………………………………………50

Genetic diversity for developmental genes……………………………………50

Summary for marker-trait associations identified……………………………...51

MTA for GY and yield components……………………………………………51

MTA for agronomic traits………………………………………………………52

Discussion…………………………………………………………………………52

Rapid LD decay………………………………………………………………...52

Moderate genetic stratification…………………………………………………53

Genome location of identified MTA compared to previous studies……………54

Current and future genetic improvement of southern US winter wheat………..57

Conclusions………………………………………………………………………..58

References…………………………………………………………………………59

Chapter III Validation of yield QTL from soft winter wheat using a CIMMYT spring wheat panel ……71

Abstract……72

Introduction……73

Materials and Methods……74

Wheat association mapping initiative (WAMI)……74

Selection of QTL for validation, primer design, and KASP assay genotyping……75

WAMI phenotypic data analyses……76

Genome-wide association study and candidate gene analysis……77

Results……78

Heritability across environments and trait correlations……78

Principal components analyses (PCA) for measured traits across environments……78

KASP analysis and association mapping……79

Haplotype analysis of validated winter wheat SNPs……80

Candidate genes associated with validated SNP loci……81

Discussion……82

Winter wheat QTL on 3A, 4B, and 6B were validated on WAMI spring wheat panel……82

Positive alleles at all three validated loci resulted in the highest mean grain yield……83

Candidate genes associated with the validated SNPs
reflect genetic complexity of grain yield……………………………….84
Conclusions…………………………………………………………………......85
References……………………………………………………………………….87
Chapter IV Genomic selection for grain yield and agronomic traits in soft winter wheat………………………………………………………………97
Abstract…………………………………………………………………………98
Introduction……………………………………………………………………..99
Materials and Methods…………………………………………………………...100
Plant Material……………………………………………………………100
Genotype data……………………………………………………………101
Phenotype data…………………………………………………………...101
Genomic selection model………………………………………………...102
Genomic selection scenarios……………………………………………..103
Different factors affecting prediction accuracy for the TP………………103
Size of the training population……………………………………103
Number of markers…...…………………………………………..103
Relatedness and population structure………………………….....104
Covariates in the model…………………………………………...104
Response to selection for grain yield using the TP………………………104
Forward validation of genomic selection model
using biparental populations……………………………………105
Results……………………………………………………………………………106

Phenotypic data……………………………………………………………………..106

Effect of TP size……………………………………………………………………106

Effect of marker number…………………………………………………………106

Effect of relatedness……………………………………………………………...107

Effect of covariates…………………………………………………………………107

Response to selection for grain yield……………………………………………108

Forward predictions in bi-parental populations…………………………………108

Discussion………………………………………………………………………………108

Marker number, training population size, and relatedness affect the accuracy of genomic prediction………………………………………109

Using markers for major genes as covariates improve predictability for grain yield in the TP………………..……...…………………………….111

Combining genome-wide prediction with phenotypic selection resulted to the highest response to selection for grain yield…………………………......112

Forward validations showed potential for predictive breeding of complex traits in winter wheat…………………………………………………………....112

Conclusions……………………………………………………………………………113

References……………………………………………………………………………...114

Conclusions……………………………………………………………………………...126

Over-all conclusions…………………………………………………………………127

Appendices…………………………………………………………………………….....128

List of Tables

CHAPTER II **Page**

Table 1. Adjusted means and analysis of variance (ANOVA) of the measured traits for the soft winter wheat association mapping panel……………………66

Table 2. Phenotypic correlations (*r*) of the measured traits for the soft winter wheat association mapping panel……………………………………………67

Table 3. Markers associated with multiple traits identified for yield and agronomic traits for the soft winter wheat association mapping panel……………………………………68

CHAPTER III

Table 1. List of QTL from winter wheat for validation in WAMI spring wheat panel from CIMMYT, Mexico……………………………………………………..……………...91

Table 2. SNPs from winter wheat that were validated and showed significant associations on the WAMI spring wheat panel………………………………92

CHAPTER IV

Table 1. Descriptive statistics and heritability values for each of the grain yield datasets used for genome-wide predictions……………………………………………119

Table 2. Response to selection, *R* for grain yield in the training population across different selection strategies……………………………………………………120

List of Figures

CHAPTER II **Page**

Figure 1. PCA biplots for the (a) measured traits and (b) adjusted GY across different site-years for the soft winter wheat AMP……………………………………………69

Figure 2. Manhattan plot showing genome-wide SNP loci associated with grain yield, heading date, and plant height PH……………………………………...70

CHAPTER III

Figure 1. Mean and standard error comparisons of seven segregating SNP markers on the WAMI spring wheat panel for grain yield, mean across 29 environments.…………93

Figure 2. Manhattan plots showing association of grain yield (GY) and grain number (GNO) QTL from winter wheat on the WAMI spring wheat panel across all environments (ABLUP dataset) for (a) GY and (b) GNO…………………………………………..94

Figure 3. Marker segregation for the designed assays: a) *IWA3560* (3A; *PA*- 'Pioneer 26R61'/'AGS 2000' biparental mapping population); b) *IWA1818* (4B; *AMP*- winter wheat association mapping panel); and c) *IWA755* (6B; PA)………………………..95

Figure 4. Average grain yield across all environments (t ha^{-1}) for the different combinations of alleles for the winter wheat QTL validated at 3A (*IWA3560,* **A**/G), 4B (*IWA1818,* **C**/T), and 6B (*IWA755,* A/**G**) (allele combination in that order) loci.………………..96

CHAPTER IV

Figure 1. Effects of training population (TP) size on accuracy of genomic selection for GY, yield components and agronomic traits; rrBLUP model, 10-fold CV for BLUP across all environments (ABLUP) at a constant validation population (VP) size (*N*=60) ………...121

Figure 2. Effect of marker number on the accuracy of genome-wide prediction for GY, yield components and agronomic traits under an rrBLUP model, 10-fold CV for BLUP across all environments (ABLUP)……………………………………………………………122

Figure 3. Effect of using different subgroups, *Q* as training population to predict grain yield for other subgroups; *Q1*-59 lines; *Q2*-54 lines; *Q3*-126 lines……………………………123

List of Figures (cont.)

Figure 4. Effect of adding covariates to the predictability of GY across different datasets. Predictions were done using a standard single population CV; TP size= 144…………124

Figure 5. Accuracy for genomic selection using TP (*N*=239; ABLUP, NBLUP, and SBLUP datasets) to predict GY in the NB (*N*=100) and PA-RIL (*N*=156) across different site years and clusters…………………………………………………………………....125

List of Published Papers

Lozada, D.N., Mason, E. Babar, M. A., Carver, B., Brown-Guedira, G., Merrill K., Arguello, N., Acuna, A., Vieira, L., Holder, A., Addison, C., Moon, D., Miller, R., Dreisigacker, S. (2017) Association mapping reveals loci associated with multiple traits that affect grain yield and adaptation in soft winter wheat. Euphytica 213:213-222. DOI: 10.1007/s10681-017-2005-2 (Chapter II)

Lozada, D.N., Mason, E. Sukumaran, S. Dreisigacker, S. Validation of grain yield QTL from soft winter wheat using a CIMMYT spring wheat panel. Crop Science, doi: 10.2135/cropsci2018.04.0232 (Chapter III)

CHAPTER I

INTRODUCTION AND LITERATURE REVIEW

Wheat

Bread wheat (*Triticum aestivum)* is an important food crop and provides 20% of calories to the world's population and a similar proportion of daily protein for about 2.5 billion people in less-developed countries (Braun et al., 2010). Wheat is the principal food grain produced in the United States and ranks third among field crops in both planted acreage and gross farm receipts, after maize and soybeans (US Department of Agriculture, 2013). The USDA Economic Research Service (USDA-ERS) reported a total acreage of 56.82 million acres (22.99 million hectares) and a national average yield of 43.7 bushels per acre for 2014 in the United States.

There are six recognized classes of *T. aestivum* in the US which are classified based on their hardiness, consistency, and season of planting (Beuerlein, 2001). These classes include durum wheat, hard red spring, hard red winter, hard white, soft white, and soft red winter. Durum wheat, the hardest class, is grown primarily in North Dakota and is used for pasta products. Hard red spring contains the highest protein among the classes making it excellent for baking and is grown mostly in Montana, North and South Dakota, and Minnesota. Hard red winter is used mostly for bread and all-purpose flour, and grown in the Great Plains, between the Mississippi River and the Rocky Mountains, and from Texas to the Dakotas and Montana. Hard white is the newest class in the United States and is closely related to red wheat except that it has a milder, sweeter flavor. It is used in hard rolls, bulgur, tortillas, yeast breads, and oriental noodles. Soft white, with low protein and high yields is used for bakery products other than bread and is grown mostly in the Pacific Northwest, California, Wisconsin, Michigan, and New York.

Soft red winter wheat is seeded in the fall and has low to medium protein content with a soft endosperm. It is primarily used to make cakes, pastries, flat breads, and crackers. It is grown throughout the southeastern U.S., with Ohio, Arkansas, Illinois, and Missouri being the major

producers. According to a report from the USDA-ERS, soft red winter is next to hard red winter wheat in terms of planted acreage in the US with a total of 8.50 million acres (3.44 million hectares) for 2014 and had the highest mean yield with 63.6 bushels/acre.

Wheat genetics and genome sequencing

The allohexaploid nature of the wheat makes it the species with the largest genome among cereals (William et al., 2007). The polyploidy was a consequence of the hybridization of the diploid DD genome of *Aegilops tauschii* with the AABB tetraploid genome of *T. turgidum* (Dubcovsky and Dvorak, 2007) resulting in the extant hexaploid bread wheat (AABBDD 2n = 6x = 42) (Kamran et al., 2014a). *T. urartu* is a known progenitor of the A genome while the B genome is thought to be derived from an unknown species of the *Sitopsis* genus (Feldman and Levy, 2005). The large size of the *T. aestivum* genome is a major constraint to sequencing as physical map construction remains a challenge (Kamran et al., 2014a) and there is no fully-ordered published sequence to date. Regardless, draft sequences of bread wheat and its progenitors have already been previously published consequently providing a richer panel of genomic resources that can be used to study important genes.

Using a whole-genome shotgun approach, Brenchley et al. (2012) initially reported the sequencing of the hexaploid genome of bread wheat. Publications of draft sequences of the ancestral species eventually followed. Whole genome shotgun draft sequence of the A-genome progenitor *T. urartu* (Ling et al., 2013) and draft sequence of the D-genome ancestor *A. tauschii* (Jia et al., 2013) were subsequently published. Shortly after, Saintenac et al. (2013) published a sequence-based map of the polyploid wheat genome through the application of NGS technique to a double-haploid population of wheat. Mayer et al. (2014) successively issued a chromosome-

based draft sequence of the 17 GB genome of bread wheat by sequencing isolated chromosome arms.

The need to increase wheat production

According to the US Census Bureau (2013), world population is projected to hit 9.5 billion by the year 2050. While linear growth in wheat production and productivity in the past has been observed, gains at current levels are insufficient to meet the demands of population growth (Gupta et al., 2010). Ensuring food security amidst the rapidly increasing population, together with the threats of the constantly changing climate, and the presence of biotic and abiotic stresses (heat, drought, waterlogging, etc.) have catalyzed efforts to improve wheat varieties through various breeding programs and initiatives. Reynolds et al. (2012) and Edgerton (2009) emphasized that the most direct solution to these problems will be to increase productivity on currently cultivated land through adoption of cultivars with improved genetic potential. Improved varieties are anticipated to be higher yielding, more tolerant to stresses, and more adaptable to a wide range of environmental conditions than the traditional ones.

Molecular techniques have been tapped in the recent years to facilitate the development of improved cultivars of important crops. New advances in molecular marker technologies have allowed researchers to explore the potential of improving varieties by examining the genetic makeup of a particular cultivar. Molecular marker approaches have been integrated with plant breeding through the process of marker-assisted selection (MAS).

Marker-assisted breeding

Molecular or DNA-based markers which represent genetic differences between individual organisms or species have helped in accelerating the development of improved varieties of crops through an approach called marker assisted selection (MAS) (Vogel, 2009). DNA markers

increase the efficiency of conventional plant breeding by making selections not directly on the trait of interest but on the molecular markers linked to the trait (Mohan et al., 1997). The development and use of markers for the detection of DNA polymorphisms is one of the most significant developments in the field of molecular genetics (Kesawat and Das, 2009).

Collard et al. (2005) listed some advantages of MAS: 1) time saving; 2) unreliable phenotypic evaluation associated with field trials due to environmental effects are eliminated, thus significantly enhancing genetic gain for these traits (Moose and Mumm, 2008); 3) selection of genotypes at seedling stage is possible; 4) gene 'pyramiding' or combining multiple genes simultaneously; 5) transfer of undesirable or deleterious genes is avoided; 6) selecting for traits with low heritability; and 7) testing for specific traits where phenotypic evaluation is not feasible. MAS was also noted to accelerate the deployment of transgenes in commercial cultivars (Moose and Mumm, 2008).

MAS, on the one hand also has its limitations. These include lack of strong trait-marker relationships, polymorphisms and/or diagnostic markers, cost, and genome structure (Gupta et al., 2010). Availability of markers for major traits of interest is regarded to be a limitation as there is a lack of reliable markers for abiotic stress tolerance (e.g. drought) and quantitative disease resistance. Another issue raised is the low level of polymorphisms in markers used in most wheat breeding programs since the germplasm used is often based on a narrow gene pool. Markers used in polyploid species such as wheat must be capable of distinguishing between the large polymorphisms seen in homeologous and paralogous genes compared with the relatively infrequent varietal polymorphisms (Barker and Edwards, 2009). The cost of marker assays is also considered to be a limiting factor, together with the complex genome structure of wheat which affects the regulation of important traits (William et al., 2007). On another note, Bonnett

et al. (2005) observed that F_2 enrichment, increasing homozygosity through inbreeding or double haploids, and backcrossing to increase recurrent parent allele frequencies are all efficient strategies to improve MAS.

Recognizing the enormous potential of DNA markers in plant breeding, many agricultural, research, and commercial institutions have adopted the capacity for marker development and MAS (Collard et al., 2005). In the late 1990's, Monsanto® decided to implement MAS for quantitative traits in their global breeding programs which consequently led to an increased mean performance of their elite breeding populations (Eathington et al., 2007).

Examples of target traits where MAS strategies have been successful in wheat include resistance against *Fusarium* head blight (del Blanco et al., 2003), scab (Zhou et al., 2005), powdery mildew (Tucker et al., 2006), and leaf rust (Nocente et al., 2007). Through a MAS approach, University of California, Davis was able to develop their first commercially available variety 'Patwin,' a hard white spring type which contains the introgressed stripe rust resistant gene *Yr17* and leaf rust resistant gene *Lr37* (Gupta et al., 2010; Helguera et al., 2003; Hospital, 2009).

Other reported success stories involving the use of MAS approach in wheat include the development of disease resistant varieties 'BIOINTA 2004', a hard red winter wheat from Argentina possessing the *Lr47* gene responsible for leaf rust resistance (Bainotti et al., 2009); the hard red spring Canadian varieties 'Goodeve' which is resistant to the insect orange blossom wheat midge and contains the gene *Sm1* (DePauw et al., 2009) and 'Lillian' which contains the gene *Gpc-B1* which gives a high grain protein content (DePauw et al., 2005).

Molecular markers for marker-assisted selection

Simple sequence repeats (SSRs, also called microsatellites) which are tandemly repeated DNA sequences of short repetitive motives (Ganal and Roder, 2007) have been particularly popular in MAS as markers because they are observed to be co-dominant, robust, reproducible, and reliable as a PCR based system (William et al., 2007). Additionally, these markers are genome specific and highly polymorphic even among related germplasm and thus are the marker type of choice for diversity analyses (Ganal and Roder, 2007; Wurschum et al., 2013). Large scale microsatellite maps for wheat (Roder et al., 1998; Somers et al., 2003) greatly expedited their utilization for molecular breeding. However, it was observed that the recovery rates for these markers are generally low and that they require *de novo* development which is costly and time consuming (Whankaew et al., 2012; Zane et al., 2002). Allen et al. (2011) on the one hand, underscored that a crucial step in the successful application of MAS in breeding programs is the development of cheap and easy to use molecular markers such as single nucleotide polymorphisms or SNPs.

SNP markers have acquired popularity as an alternative marker system for breeding applications. As suggested by the acronym, SNP or single nucleotide polymorphism is a single-base change in the DNA sequence at which different nucleotides occur in different individuals of populations (Kesawat and Das, 2009). SNPs have recently gained considerable interest as they occur in virtually unlimited numbers in the genome (Ganal and Roder, 2007). The subsequent shift to SNP markers from SSRs has made excellent progress to characterize genetic diversity of major crop species, to map QTL for key traits, and to clone genes important for crop improvement (Thomson, 2014).

Advantages of SNPs as DNA-based markers include their high abundance in the genome, ease of data management- scoring and interpretation of results; flexibility, speed, high throughput platforms, relative simplicity in assay design, and low cost (Bajgain et al., 2016; Kanazin et al., 2002; Thomson, 2014). Using information from expressed sequence tags (ESTs), Somers et al. (2003) estimated that there is an occurrence of a SNP for every 540 base pairs (bp) in the wheat genome. Semagn et al. (2013) noted that SNPs have largely replaced SSRs in crop species that have been extensively sequenced, and they are expected to replace other molecular marker types in most species given the increased use of next-generation sequencing (NGS) technologies for genotyping. There are many genetic applications of SNPs including germplasm characterization (genetic diversity, relationships, population structure), allele mining, linkage mapping, marker-assisted backcrossing (MABC), marker-assisted recurrent selection (MARS), dissection of complex genetic traits such as genome-wide association studies, and genomic selection (He et al., 2014a; Semagn et al., 2013).

Genetic factors controlling growth habit in wheat

Genes controlling response to vernalization and photoperiod, and those involved in plant stature are three genetic systems that control growth habit in wheat (Blake et al., 2009). A broader understanding of the effect of these major genes controlling growth habit is crucial to continuing to unlock the potential for breeding of wheat cultivars that are higher yielding and more adapted to target environments.

Vernalization genes

Vernalization is the physiological effect of chilling corresponding to the awakening of nature in spring (Chouard, 1960). It is the requirement of a long exposure to low temperature to induce and accelerate flowering in wheat and other cereals (Dubcovsky et al., 2006; Yan et al., 2004).

Vernalization is an important adaptation in response to cold environments for the plant's transition from the vegetative to the reproductive phase (Kamran et al., 2013). It helps prevent flowering during the winter which can consequently damage the plant's cold-sensitive meristem and permits flowering under favorable conditions in the spring (Chen et al., 2013). Flower development will only start once the risk of damage as a consequence of cold is minimal, i.e. flowering is delayed until winter and the danger of frost damage has passed (Cockram et al., 2007).

Differences in the vernalization genes present divide wheat cultivars into the "non-vernalization requiring" spring, the facultative, and the "cold-requiring" winter habits (Blake et al., 2009; Distelfeld et al., 2009; Kamran et al., 2014b; Yan et al., 2003). Winter wheat varieties require up to 45 days (1,080 hours) accumulated exposure to temperature between 32 to 45^0F (0 to 7^0C) to vernalize (Morgan et al., 2008). Vernalization, together with photoperiod are two of the main environmental cues that plants monitor to determine the appropriate time to flower (Dennis and Peacock, 2009; Sung and Amasino, 2004). The ability of wheat to synchronize its flowering during favorable conditions is central to its global adaptability and hence to its success (Allard et al., 2011; Kamran et al., 2013; Shewry, 2009).

Response to vernalization in hexaploid wheat is mainly controlled by a series of orthologous genes, namely, the *Vrn-A1* on chromosome 5A, *Vrn-B1* on 5B, and *Vrn-D1* on 5D, collectively known as the *Vrn-1* series (Rousset et al., 2011; Santra et al., 2009). Dominant alleles at *Vrn-A1* confers complete insensitivity to vernalization (Iqbal et al., 2011a) while the spring *Vrn-B1* and *Vrn-D1* alleles provide a reduced vernalization requirement relative to the winter alleles (Blake et al., 2009). In addition, other vernalization genes on wheat have been

mapped on chromosomes 3B, 4B, 4D, *Vrn-D5* on 5D, and *Vrn-B3* on 7B, all together known as the *Vrn-2* series (Iwaki et al., 2000).

A variation on the potency conferred by these genes exists from complete insensitivity to partial or weak sensitivity, depending on the type of *Vrn* alleles present (Diaz et al., 2012; Kamran et al., 2013). *Vrn-A1* has the strongest effect on inhibiting vernalization requirement, followed by *Vrn-D1*, *Vrn-D5*, and *Vrn-B1*, respectively (Goncharov, 2004); thus, plants with dominant *Vrn-A1* will head first while those having *Vrn-B1* will head last, provided that other genetic factors remain constant (Kamran et al., 2014a). Variation on the copy number for *Vrn-A1* was also found out to cause an increased vernalization requirement for cultivated bread wheat, rendering the potential role of copy number variation (CNV) in wheat adaptation (Diaz et al., 2012). It has also been shown that wheat responds linearly to vernalization duration, suggesting the quantitative nature of vernalization response (Streck et al., 2003).

Wheat and barley (*Hordeum vulgare*) are the only plant species aside from *Arabidopsis* in which vernalization genes have been well-characterized molecularly (Kim et al., 2009). Using a positional cloning approach, Yan et al. (2003) were able to map and clone the *VRN1* (now *Vrn-*A1) gene in the wild relative *T. monococcum* and found out at that it is completely linked to the MADS-box genes *AP1* (APETALA1) and *AGLG1* (agamous-like gene from grasses). Analyses of gene expression profiles eventually led to identifying the earlier as a better candidate for the *VRN1* gene and that a deletion in its promoter was associated with spring growth habit. A follow-up examination of the allelic variation at the promoter region of *VRN1* revealed duplication at the promoter region of the *Vrn-A1a* allele (Yan et al., 2004). Moreover, it was found out that *Vrn-A1b* allele has two mutations in the host direct duplication (HDD) region and a 20-bp deletion in

the 5'-UTR (untranslated region). Ultimately, it was thought that *VRN-1* genes should have extra sites of regulation localized outside the region of the promoter.

Previous studies have characterized the *Vrn* and *Ppd* response genes for various sets of germplasm from different geographical regions of the world. Such information is crucial to understanding the adaptability of wheat cultivars to different environments (Zhang et al., 2008). Eagles et al. (2010) examined the effects of *Vrn* and *Ppd* genes in southern Australian wheat cultivars, Chen et al. (2013) characterized vernalization and photoperiod response genes of wheat from the Yellow and Huai Valley of China, Iqbal and colleagues (2007) surveyed the effects of *Vrn* genes in Canadian spring wheat, and Singh et al. (2013) examined these genes from varieties coming from different agro-climatic zones of India. Other reported analyses of vernalization response genes from varieties across different regions include those genotypes from Russia (Shcherban et al., 2012), Turkey (Andeden et al., 2011), Bulgaria (Kolev et al., 2011), Pakistan (Iqbal et al., 2011b), China, Korea, and Japan (Iwaki et al., 2000), and the Pacific Northwest region of the US (Santra et al., 2009).

Photoperiod genes

Photoperiodism is the phenomenon where plants respond to variable day and/or night length by receiving signals in the form of cryptochrome or phytochrome to initiate flowering (Fosket, 1994). In wheat, photoperiod sensitive cultivars require long days for induction of flowering while photoperiod insensitive genotypes flower independently of day length (Blake et al., 2009). Photoperiod insensitive cultivars of wheat immediately shift to reproductive growth with a rise in temperature in the spring, while photoperiod sensitive continue in the vegetative phase until the day length sufficiently increases to satisfy photoperiod requirement (Snape et al., 2001). Next to vernalization requirement, photoperiod response is regarded as the second most important

genetic system determining flowering time, and hence adaptation of wheat to different agro-climatic conditions (Kamran et al., 2014a).

Photoperiod response is mainly controlled by the *Ppd-1* loci, namely the *Ppd-A1, Ppd-B1, and Ppd-D1* located on the short arms of chromosomes 2A, 2B, and 2D, respectively (Law et al., 1978b; Scarth and Law, 1983; Snape et al., 2001). The *Ppd-D1* allele is considered to be the most potent in conferring insensitivity to photoperiod, followed by *Ppd-B1* and *Ppd-A1* (Chen et al., 2013; Worland, 1996). Insensitive alleles for photoperiod are designated by the suffix 'a' while sensitive alleles are designated as 'b'. *Ppd-A1a, Ppd-B1a, and Ppd-D1a* hence indicate insensitive whereas *Ppd-A1b, Ppd-B1b, and Ppd-D1b* indicate sensitive at the three loci (McIntosh et al., 2008).

Recent studies have focused on the molecular characterization and mapping of the major genes involved in photoperiod response. Sun et al. (2014) reported two different methylation patterns or haplotypes in the regulatory region of *Ppd-B1* alleles that are associated with copy number variation and photoperiod insensitivity. Earlier, Beales et al. (2007) identified a "misexpressed" pseudo-response regulator (PRR) in the photoperiod insensitive *Ppd-D1a* mutant of wheat and demonstrated the gene to be collinear with the *Ppd-H1* of barley. Hanocq et al. (2004) detected four different photoperiod sensitivity QTL from chromosomes 2B, 2D, 5A, and 7D using an F_7 RIL population derived from the cross between cultivars 'Renan' and 'Recital'. Prior to this, Shindo et al. (2003) identified markers linked to photoperiod sensitivity on chromosomes 2B, 4B, 5A, 5B, and 7A when they examined an F_8 RIL population derived from a cross between *T. aestivum* (cv. 'Chinese spring') and *T. spelta* (var. *'dumalemiamum'*).

Reduced height genes

Impressive increases in yield during the 'Green Revolution' have been primarily attributed to the introduction of dwarfing genes (*Rht*) which rendered resistance to lodging and higher harvest index (HI). (Gale and Youssefian, 1985; Hedden, 2010; Pearce et al., 2011). A higher HI signifies that a greater proportion of the products of photosynthesis accumulates in the grains rather than in the leaves (Flintham et al., 1997; Hedden, 2010). This increased in HI is a consequence of reduced internal competition for assimilate supply between the developing ear and the stem during elongation before flowering (Chapman et al., 2007).

Slafer and Araus (2007) observed that reducing height to a certain level has no effect on the crop's ability to capture resources while markedly improving the efficiency with which these resources are used to produce yield. Through the utilization of F_3, F_4, and F_5 lines of wheat, the genetic relationship between height and yield has long been established and was demonstrated to be positively correlated (Law et al., 1978a).

Dwarfing genes are classified to be either gibberellin (GA)-sensitive or GA-insensitive, based on whether applied GA did or did not result in increased stem elongation (Gale and Youssefian, 1985). *Rht1*, *Rht-B1, Rht-B1b*, and *Rht-D1* among others were classified to be GA-insensitive while *Rht4*, *Rht5*, *Rht12*, and *Rht13* were regarded to be the GA-sensitive alleles (McIntosh et al., 2008). *Rht-B1* and *Rht-D1* encode DELLA proteins which act to repress GA-responsive growth; a limited response to GA for GA-insensitive alleles results in improved resistance to stem lodging and yield benefits through an increase in grain number (Pearce et al., 2011). Pearce et al. (2011) also demonstrated that severe dwarfism caused by *Rht-B1c* is caused by intragenic insertion while extreme dwarfism due to *Rht-D1c* is attributed to the overexpression of the *Rht-D1b* allele. Peng et al. (1999) earlier demonstrated that *Rht-B1* and

Rht-D1 encode mutant gibberellin response modulators that are orthologues of the *Arabidopsis* Gibberellin Insensitive (GAI) gene.

PCR-based markers for *Rht-B1b* and *Rht-D1b* were developed to detect point mutations responsible for these genes in wheat and eventually dubbed as "perfect markers" since they are specific for the base pair change responsible for the semi-dwarf phenotype (Ellis et al., 2002). Ellis et al. (2005) were able to identify the chromosomal locations of several height-reducing genes by screening populations of recombinant inbred and double haploid lines of bread wheat. Linked markers were found for *Rht5* on chromosome 3BS, *Rht12* on 5AL, and *Rht13* on 7BS, which accounted for most of the phenotypic variance. The height-reducing effect of these genes across target environments was also observed. Semi-dwarfing genes *Rht-B1b* (*Rht1*) and *Rht-D1b* (*Rht2*) were introduced into commercial wheat cultivars from the Japanese variety 'Norin 10' in the 1960s as part of wheat improvement programs in USA and Mexico (Ellis et al., 2002).

Effects of major growth habit genes on yield and adaptation of wheat

Understanding the effects of genes involved in adaptation is crucial for the breeding and development of varieties that are more adjusted to local environments. Improvement in grain yield is a primary objective of wheat breeding programs (Green et al., 2012), including that of the University of Arkansas (Esten Mason, pers. communication). Slafer (2003) defined "yield" as "the final outcome of the crop growth and development process occurring throughout the growing season" while Evans and Fischer (1999) referred to it as the "mass of product at final harvest, for which dry matter content should be specified." In order to maximize yield, it is essential to tailor a plant's life cycle to the agro-environments in which they are grown ensuring that the appropriate flowering time and life cycle duration are met (Cockram et al., 2007; Snape et al., 2001). Wheat yield has been increased globally through modification of its developmental

pattern that best suit specific growing condition (Kamran et al., 2014a). Stelmakh (1998) observed that *Vrn1*, *Vrn2*, and *Vrn3* genes have different effect values in relation to heading date, plant height, and yield components.

Vrn response genes are known to contribute indirectly to the yield of wheat by influencing flowering time (Kumar et al., 2012), and tiller and spikelet number in sensitive genotypes (Iqbal et al., 2007). Genotypes having two dominant alleles in combination at two *Vrn* loci tended to be early maturing and higher yielding, suggesting the possibility of combining specific dominant genes in spring wheat to improve yield potential (Kamran et al., 2014a; Stelmakh, 1998). Additionally, Iqbal et al. (2011b) emphasized that early maturing spring cultivars with desirable grain yield potential may be developed if specific dominant *Vrn* alleles are combined in a genotype. After examining a collection of Canadian spring wheat germplasm, Kamran et al. (2013) reported that 74% of soft white lines possessing a less potent vernalization gene, *Vrn-B1* alone or in combination with other *Vrn* genes are higher yielding. The findings of a subsequent study by Kamran et al. (2014b) suggested the possible role of *Vrn-D1* in producing higher grain yield on a set of Canadian spring wheat lines. On a recent study, Zhang et al. (2014b) identified combinations of vernalization response genes that resulted to high yield in drought and well-watered conditions for a double haploid population of wheat segregating for *Vrn-A1a*, *Vrn-B1a*, and *Vrn-D1a*. The genotype *vrn-A1/vrn-B1/vrn-D1* showed high kernel number (KN) and grain weight (GW) in well-watered environments. On the one hand, the genotype *Vrn-A1a/vrn-B1/Vrn-D1a* gave high GW and KN in drought conditions.

A relationship between photoperiod response and yield was earlier established by Worland and colleagues (1998) when they examined the gene's influence on the adaptability of winter wheat varieties from Europe. The authors showed that early flowering *Ppd-1* genotypes

produce larger grains and greater yields in the Southern European region. Kamran et al. (2013) noted that yield advantages with photoperiod insensitive cultivars were possibly due to escapes from hot summers by maturing earlier as hot, dry conditions are associated with decreased tiller number and decreased grain weight.

Photoperiod alleles, in combination with vernalization response genes have also been observed to have effects on yield of wheat cultivars. Under early spring sowing conditions, Kolev et al. (2010) showed that allele combinations *Ppd-D1a/Vrn-A1a* and *Ppd-D1b/Vrn-A1a* were higher yielding in a set of Bulgarian varieties. Field studies have also shown that photoperiod genes play an important role in accelerating or delaying flowering time in spring after vernalization requirement has been satisfied (Snape et al., 2001). An examination of the effect of the insensitive allele *Ppd-A1a* on the heading date of Japanese wheat revealed that cultivars from the Kanto region possessing the allele headed ~7-10 days earlier while varieties from Hokkaido headed 2.5 days earlier than the sensitive genotypes (Seki et al., 2013).

Foulkes et al. (2004) observed an average advanced flowering by 9-12 days of wheat NILs coming from the UK and Kamran et al. (2013) noted reduction for time of flowering from 1.52-1.57 days for wheat genotypes from Canada. Using introgression lines developed from the spring wheat variety 'Paragon' population, Shaw et al. (2013) found out that wheat lines lacking *Ppd-B1* flowered 10-15 days later than controls under long day conditions, while candidate loss of function *Ppd-A1* delayed flowering by 1-5 days confirming the effects of loss of function mutations to flowering under long days. Similarly, Kiss and coworkers (2014) observed that entries possessing photoperiod-insensitive alleles in *Ppd-D1* and *Ppd-B1* headed the earliest among a worldwide collection of 683 wheat genotypes. A recent study by Guedira et al. (2014) identified QTL related with photoperiod response and vernalization sensitivity on chromosomes

2B and 5B, respectively. These QTL associated with the environmentally-sensitive photoperiod and vernalization genes were shown to be the major determinants of heading dates in eastern soft wheat winter germplasm.

The increase on mean yield of wheat varieties during the 'Green Revolution' was attributed primarily to the presence reduced height (*Rht*) genes in wheat (Hedden, 2010). Yield advantages of shorter wheat plants over tall controls were earlier observed by Flintham et al. (1997) when they conducted yield trials in eastern England and Central Germany. Addisu et al. (2010) observed that *Rht-D1b* was associated with reduced height, increased harvest index (HI), and grain yield when they examined near isogenic lines (NILs) of wheat under two contrasting production systems. The semi-dwarfing *Rht-B1b* and *Rht-D1b* are usually associated with increased wheat yields (Rebetzke et al., 2011) but their effects vary with environment (Chapman et al., 2007). Reduction in height was observed to be correlated with reduced lodging score and increased grain number on a set of four inbred wheat populations segregating for one or more gibberellin-responsive dwarfing genes (Rebetzke et al., 2012) and on a set of near-isogenic (NILs) and recombinant inbred lines (RILs) derived from the cross between 'Magnif M1' and 'Chuan-mai 18' (Rebetzke et al., 2011).

Genome-wide association study

Identification of marker-trait associations is the first step towards marker-assisted selection (Wang et al., 2014). Genome-wide association study (GWAS) is a method that relies on linkage disequilibrium (LD), which is the nonrandom combination of alleles at two genetic loci to study the relationship between phenotypic variation and single nucleotide polymorphisms (Breseghello and Sorrells, 2006; Flint-Garcia et al., 2003). GWAS can be an effective approach for bridging the gap between QTL analysis and MAS (Myles et al., 2009). Moreover, it is useful for

dissecting complex traits controlled by multiple QTL when LD decays rapidly and is anticipated to be an efficient method for the study of complex traits in wheat (Wang et al., 2014).

There are some advantages of AM over bi-parental mapping. Its main advantage is that it exploits all the recombination events that have occurred in the individuals' evolutionary history (Myles et al., 2009) in contrast to linkage analysis where there are only a few opportunities for recombination to occur within families and pedigrees with known ancestry (Zhu et al., 2008). Moreover, a much larger and more representative gene pool can be surveyed and screened for genetic variation in complex traits (Neumann et al., 2011; Zhao et al., 2011). On another note, it was emphasized that the statistical tools required to perform the analysis are more complex due to the probable presence of false positive associations in the population (Neumann et al., 2011). The power of detecting significant marker-trait associations also depends on the quality of the phenotypic data, sample size, and the genetic architecture and heritability of the trait under study (Barabaschi et al., 2016).

Association mapping studies have been previously conducted in wheat. Wang et al. (2014) recently reported marker-trait associations using GWA analysis for five important agronomic traits, namely, kernel hardness, thousand-kernel weight, grain protein content, test weight, and plant height in a diverse set of 94 wheat lines. Similarly, Neumann et al. (2011) reported a genome-wide association study for 20 agronomic traits in a winter wheat core collection using diversity array technology (DArT) markers where significant marker-trait associations were detected for plant height, grain yield, and disease resistance. On the one hand, a total of six known stem rust resistance genes were detected by Zhang et al. (2014a) when they conducted association mapping for resistance genes in US winter wheat germplasm using SSR and sequence-tagged sites (STS) markers.

Genomic regions associated with resistance to aluminum toxicity were earlier identified by Raman et al. (2010) using a set of 178 polymorphic DArT markers. Prior to this, Roy et al. (2006) identified associations for 14 agronomic traits using SSR, selective amplification of microsatellite polymorphic loci (SAMPL), and amplified fragment length polymorphism (AFLP) markers in elite genotypes of wheat. Sukumaran et al. (2014) identified thirty-one significant loci associated with grain yield and yield related traits in a population consisting of 287 elite lines of spring wheat. Using a candidate gene association mapping approach, the association of *Vrn-1A* functional gene with heading date and days to anthesis was also demonstrated. Recently, Hoffstetter et al. (2016b) identified important loci governing yield and economic traits in an elite collection of soft red winter wheat (SRWW) lines grown in the northeastern US through a genotyping-by-sequencing (GBS)- GWAS approach.

Marker-trait association analyses have also been conducted on other crops. Much emphasis has been given to find markers associated with flowering time particularly in diverse inbred lines of maize (Salvi et al., 2007; Thornsberry et al., 2001; Wilson et al., 2004) and *Arabidopsis* (Olsen et al., 2004). Zhao and co-workers (2011) later examined a global collection of diverse rice (*Oryza sativa*) germplasm for 34 different traits and observed significant genetic heterogeneity among the four subgroups of rice. Additionally, association analyses have been conducted for functional gene markers for pro-vitamin A levels in maize inbred lines (Azmach et al., 2013).

Genotyping-by-sequencing

The advent of high-throughput and cost-effective genotyping technologies has further driven the use of GWAS as an alternate strategy for finding marker-trait associations. Recent improvements in sequencing throughput combined with an overall decrease in costs per gigabase

(Gb) of sequence is allowing next-generation sequencing (NGS) technologies to be used not only for the evaluation of small subsets of parental inbred lines, but also for the mapping and characterization of traits of interest in much larger populations (Deschamps et al., 2012). GWA mapping is further believed to be powerful tool to increase our understanding of complex traits, including tillering and branching of panicles, through which we can validate their molecular mechanisms and pyramid multiple genes to breed desired elite rice varieties (Wang and Li, 2011).

Genotyping by sequencing (GBS) is an application of NGS for discovering and scoring segregating markers in the population under study (Spindel et al., 2013). The key objective of GBS is to simultaneously discover polymorphisms and obtain genotypic information across the whole population of interest (Poland and Rife, 2012). GBS involves genomic DNA digestion with restriction enzymes coupled with DNA barcoded adapters to reduce genome complexity and sequencing of the ends of the resulting restriction fragments (Elshire et al., 2011; Poland and Rife, 2012). GBS technology offers a wider range of polymorphisms than PCR-based assays and eliminates the need to pre-discover and validate polymorphisms and thus can be used in any polymorphic species and segregating population (Schnable et al., 2013). Other advantages of GBS include a simplified library preparation, less starting DNA requirement, random shearing and size selection of DNA samples are avoided and contains few PCR and purification steps (Poland et al., 2012). This strategy is becoming increasingly important as a cost-effective and unique tool for genomics-assisted breeding in a range of plant species (He et al., 2014b). GBS approach for association mapping has been conducted in different crop species such as soybeans (Iquira et al., 2015; Sonah et al., 2015), oat (Huang et al., 2014), rice (Spindel et al., 2013), cotton (Islam et al., 2015) and potato (Uitdewilligen et al., 2013), among others.

Genomic selection

Genomic selection (GS) is a marker-assisted selection (MAS) tool that aims to predict and perform selection based on genomic estimated breeding values (GEBV) of individuals that are generated using genome-wide marker data through training and validation of a prediction model (Meuwissen et al., 2001). GS is a complement to traditional breeding strategies, potentially reducing the need for large-scale phenotyping and accelerating genetic gain through shorter breeding cycles (Heffner et al., 2010; Muranty et al., 2015; Nakaya and Isobe, 2012).

Pioneering studies on GS in animal breeding, particularly of cattle (Hayes et al., 2009; Meuwissen et al., 2001) have now been extended to crops, including rice (Onogi et al., 2016; Spindel et al., 2015), tomato (Duangjit et al., 2016; Hernández-Bautista et al., 2016), maize (Zhao et al., 2012), soybean (Bao et al., 2014), and barley (Lorenzana and Bernardo, 2009). In soft winter wheat, GS studies have been conducted for *Fusarium* head blight resistance (Arruda et al., 2016), grain yield and stability traits (Huang et al., 2016), yield, softness equivalence, and flour yield (Hoffstetter et al., 2016a), grain yield, plant height, heading date, and flour quality traits (Heffner et al., 2011b), and normalized difference vegetative index (NDVI) (Mason et al., 2017).

The performance of GS depends primarily on prediction accuracy, defined as the Pearson's correlation between the selection criterion and the true breeding value to select individuals with unknown phenotypes (Desta and Ortiz, 2014). Factors affecting GS accuracy include gene effects, genetic composition of the training population (TP), level of LD, marker density, model performance, QTL number, relationship between TP and the validation population (VP) or selection candidates, TP size, and trait heritability (Desta and Ortiz, 2014; Rutkoski et al., 2015; Zhong et al., 2009).

Approach of the current study

While previous studies give insights on genomic regions that render significant variation for GY and agronomic traits in wheat, reports on the use of winter wheat lines adapted to the Southeastern region of the US for association mapping and genomic selection remain limited. The objective of the current study was to identify yield-related loci in soft winter wheat and perform genomic predictions for these traits. The working hypothesis is that SNP loci that control grain yield and agronomic traits are distributed in multiple chromosomes and that genome-wide selection accuracy is affected by several factors, among which the size of the training population having the greatest effect on predictive ability. There are three specific objectives for this paper and each objective corresponds to a chapter. The specific objectives are:

Objective 1: Determine genomic regions associated with GY and agronomic traits in a soft red winter wheat panel adapted to the Southeastern region of the US. This was accomplished through a genome-wide association mapping approach employing several mixed models to identify these genetic loci. The working hypothesis is that loci that control variation for the measured traits are distributed in multiple chromosomes.

Objective 2: Validate yield-related loci identified from winter wheat using a panel of spring wheat from CIMMYT, Mexico. Based on previous meta-analyses showing co-localization of QTL detected across wheat classes, it was hypothesized that stable GY related QTL could be identified across winter and spring wheat. Designed allele specific primers were tested on the Wheat Association Mapping Initiative (WAMI) from CIMMYT, Mexico and QTL were validated through a GWAS approach.

Objective 3: Predict grain yield and agronomic traits in soft winter wheat through cross validations and forward predictions. A ridge regression (RR) model was used to

evaluate the effects of different factors, namely training population size, number of markers, relatedness, and covariates on the accuracy of genomic selection in soft winter wheat. It was hypothesized that these parameters have varying effects on the prediction accuracy and that the size of the training population has the greatest impact on the predictive ability of genomic selection.

References

Addisu M., Snape J.W., Simmonds J.R., Gooding M.J. (2010) Effects of reduced height (Rht) and photoperiod insensitivity (Ppd) alleles on yield of wheat in contrasting production systems. Euphytica 172:169-181. DOI: 10.1007/s10681-009-0025-2.

Allard V., Veisz O., Kõszegi B., Rousset M., Le Gouis J., Martre P. (2011) The quantitative response of wheat vernalization to environmental variables indicates that vernalization is not a response to cold temperature. Journal of Experimental Botany. DOI: 10.1093/jxb/err316.

Allen A.M., Barker G.L.A., Berry S.T., Coghill J.A., Gwilliam R., Kirby S., Robinson P., Brenchley R.C., D'Amore R., McKenzie N., Waite D., Hall A., Bevan M., Hall N., Edwards K.J. (2011) Transcript-specific, single-nucleotide polymorphism discovery and linkage analysis in hexaploid bread wheat (Triticum aestivum L.). Plant Biotechnology Journal 9:1086-1099. DOI: 10.1111/j.1467-7652.2011.00628.x.

Andeden E.E., Yediay F.E., Baloch F.S., Shaaf S., Kilian B., Nachit M., Ozkan H. (2011) Distribution of Vernalization and Photoperiod Genes (Vrn-A1, Vrn-B1, Vrn-D1, Vrn-B3, Ppd-D1) in Turkish Bread Wheat Cultivars and Landraces. Cereal Research Communications 39:352-364. DOI: 10.1556/crc.39.2011.3.5.

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

Azmach G., Gedil M., Menkir A., Spillane C. (2013) Marker-trait association analysis of functional gene markers for provitamin A levels across diverse tropical yellow maize inbred lines. Bmc Plant Biology 13. DOI: 10.1186/1471-2229-13-227.

Bainotti C., Fraschina J., Salines J.H., Nisi J.E., Dubcovsky J., Lewis S.M., Bullrich L., Vanzetti L., Cuniberti M., Campos P., Formica M.B., Masiero B., Alberione E., Helguera M. (2009) Registration of 'BIOINTA 2004' Wheat. Journal of Plant Registrations 3:165-169. DOI: 10.3198/jpr2008.12.0713crc.

Bajgain P., Rouse M.N., Anderson J.A. (2016) Comparing Genotyping-by-Sequencing and Single Nucleotide Polymorphism Chip Genotyping for Quantitative Trait Loci Mapping in Wheat. Crop Science 56:232-248.

Bao Y., Vuong T., Meinhardt C., Tiffin P., Denny R., Chen S., Nguyen H.T., Orf J.H., Young N.D. (2014) Potential of association mapping and genomic selection to explore PI 88788 derived soybean cyst nematode resistance. Plant Genome 7. Doi:10.3835/plantgenome2013.11.0039.

Barabaschi D., Tondelli A., Desiderio F., Volante A., Vaccino P., Vale G., Cattivelli L. (2016) Next generation breeding. Plant Science 242:3-13. DOI: 10.1016/j.plantsci.2015.07.010.

Barker G.L.A., Edwards K.J. (2009) A genome-wide analysis of single nucleotide polymorphism diversity in the world's major cereal crops. Plant Biotechnology Journal 7:318-325. DOI: 10.1111/j.1467-7652.2009.00412.x.

Beales J., Turner A., GriYths S., Snape J.W., Laurie D.A. (2007) A Pseudo-Response Regulator is misexpressed in the photoperiod insensitive Ppd-D1a mutant of wheat (Triticum aestivum L.). Theoretical and Applied Genetics 115:721-733. DOI: 10.1007/s00122-007-0603-4.

Beuerlein J. (2001) Classes and Uses of Wheat, Horticulture and Crop Science Extension Fact Sheet, The Ohio State University, Colombus, Ohio.

Blake N.K., Lanning S.P., Martin J.M., Doyle M., Sherman J.D., Naruoka Y., Talbert L.E. (2009) Effect of Variation for Major Growth Habit Genes on Maturity and Yield in Five Spring Wheat Populations. Crop Science 49:1211-1220. DOI: 10.2135/cropsci2008.08.0505.

Bonnett D.G., Rebetzke G.J., Spielmeyer W. (2005) Strategies for efficient implementation of molecular markers in wheat breeding. Molecular Breeding 15:75-85. DOI: 10.1007/s11032-004-2734-5.

Braun H.J., G. A., T. P. (2010) Multi-location testing as a tool to identify plant response to global climate change, in: M. P. Reynolds (Ed.), Climate Change and Crop Production CABI Climate Change Series, Surrey, UK. pp. 115–138.

Brenchley R., Spannagl M., Pfeifer M., Barker G.L.A., D'Amore R., Allen A.M., McKenzie N., Kramer M., Kerhornou A., Bolser D., Kay S., Waite D., Trick M., Bancroft I., Gu Y., Huo N., Luo M.-C., Sehgal S., Gill B., Kianian S., Anderson O., Kersey P., Dvorak J., McCombie W.R., Hall A., Mayer K.F.X., Edwards K.J., Bevan M.W., Hall N. (2012) Analysis of the breadwheat genome using whole-genome shotgun sequencing. Nature 491:705-710. DOI: 10.1038/nature11650.

Breseghello F., Sorrells M.E. (2006) Association analysis as a strategy for improvement of quantitative traits in plants. Crop Science 46:1323-1330. DOI: 10.2135/cropsci2005.09-0305.

Chapman S.C., Mathews K.L., Trethowan R.M., Singh R.P. (2007) Relationships between height and yield in near-isogenic spring wheats that contrast for major reduced height genes. Euphytica 157:391-397. DOI: 10.1007/s10681-006-9304-3.

Chen F., Gao M., Zhang J., Zuo A., Shang X., Cui D. (2013) Molecular characterization of vernalization and response genes in bread wheat from the Yellow and Huai Valley of China. Bmc Plant Biology 13. DOI: 10.1186/1471-2229-13-199.

Chouard P. (1960) Vernalization and its relations to dormancy. Annual Review of Plant Physiology 11:191-238.

Cockram J., Jones H., Leigh F.J., O'Sullivan D., Powell W., Laurie D.A., Greenland A.J. (2007) Control of flowering time in temperate cereals: genes, domestication, and sustainable productivity. Journal of Experimental Botany 58:1231-1244. DOI: 10.1093/jxb/erm042.

Collard B., Jahufer M., Brouwer J., Pang E. (2005) An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. Euphytica 142:169-196. DOI: 10.1007/s10681-005-1681-5.

del Blanco I.A., Frohberg R.C., Stack R.W., Berzonsky W.A., Kianian S.F. (2003) Detection of QTL linked to Fusarium head blight resistance in Sumai 3-derived North Dakota bread wheat lines. Theoretical and Applied Genetics 106:1027-1031. DOI: 10.1007/s00122-002-1137-4.

Dennis E.S., Peacock W.J. (2009) Vernalization in cereals. Journal of Biology (London) 8:57-Article No.: 57. DOI: 10.1186/jbiol156.

DePauw R.M., Knox R.E., Thomas J.B., Smith M., Clarke J.M., Clarke F.R., McCaig T.N., Fernandez M.R. (2009) Goodeve hard red spring wheat. Canadian Journal of Plant Science 89:937-944.

DePauw R.M., Townley-Smith T.F., Humphreys G., Knox R.E., Clarke F.R., Clarke J.M. (2005) Lillian hard red spring wheat. Canadian Journal of Plant Science 85:397-401.

Deschamps S., Llaca V., May G.D. (2012) Genotyping-by-sequencing in plants. Biology 1:460-483.

Desta Z.A., Ortiz R. (2014) Genomic selection: genome-wide prediction in plant improvement. Trends in Plant Science 19:592-601.

Diaz A., Zikhali M., Turner A.S., Isaac P., Laurie D.A. (2012) Copy Number Variation Affecting the Photoperiod-B1 and Vernalization-A1 Genes Is Associated with Altered Flowering Time in Wheat (Triticum aestivum). Plos One 7. DOI: 10.1371/journal.pone.0033234.

Distelfeld A., Li C., Dubcovsky J. (2009) Regulation of flowering in temperate cereals. Current Opinion in Plant Biology 12:178-184. DOI: 10.1016/j.pbi.2008.12.010.

Duangjit J., Causse M., Sauvage C. (2016) Efficiency of genomic selection for tomato fruit quality. Mol Breed 36:1-16. DOI: 10.1007/s11032-016-0453-3.

Dubcovsky J., Dvorak J. (2007) Genome plasticity a key factor in the success of polyploid wheat under domestication. Science 316:1862-1866. DOI: 10.1126/science.1143986.

Dubcovsky J., Loukoianov A., Fu D., Valarik M., Sanchez A., Yan L. (2006) Effect of Photoperiod on the Regulation of Wheat Vernalization Genes VRN1 and VRN2. Plant Molecular Biology 60:469-480. DOI: 10.1007/s11103-005-4814-2.

Eagles H.A., Cane K., Kuchel H., Hollamby G.J., Vallance N., Eastwood R.F., Gororo N.N., Martin P.J. (2010) Photoperiod and vernalization gene effects in southern Australian wheat. Crop & Pasture Science 61:721-730. DOI: 10.1071/cp10121.

Eathington S., Crosbie T., Edwards M., Reiter R., Bull J. (2007) Molecular Markers in a Commercial Breeding Program. Crop Sci 47(S3):S154-S163. DOI: 10.2135/cropsci2007.04.0015IPBS.

Edgerton M.D. (2009) Increasing Crop Productivity to Meet Global Needs for Feed, Food, and Fuel. Plant Physiology 149:7-13. DOI: 10.1104/pp.108.130195.

Ellis M.H., Rebetzke G.J., Azanza F., Richards R.A., Spielmeyer W. (2005) Molecular mapping of gibberellin-responsive dwarfing genes in bread wheat. Theoretical and Applied Genetics 111:423-430. DOI: 10.1007/s00122-005-2008-6.

Ellis M.H., Spielmeyer W., Gale K.R., Rebetzke G.J., Richards R.A. (2002) "Perfect" markers for the Rht-B1b and Rht-D1b dwarfing genes in wheat. Theoretical and Applied Genetics 105:1038-1042. DOI: 10.1007/s00122-002-1048-4.

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:e19379.

Evans L.T., Fischer R.A. (1999) Yield potential: Its definition, measurement, and significance. Crop Science 39:1544-1551.

Feldman M., Levy A.A. (2005) Allopolyploidy - a shaping force in the evolution of wheat genomes. Cytogenetic and Genome Research 109:250-258. DOI: 10.1159/000082407.

Flint-Garcia S.A., Thornsberry J.M., Buckler E.S. (2003) Structure of linkage disequilibrium in plants. Annual Review of Plant Biology 54:357-374. DOI: 10.1146/annurev.arplant.54.031902.134907.

Flintham J.E., Borner A., Worland A.J., Gale M.D. (1997) Optimizing wheat grain yield: Effects of Rht (gibberellin-insensitive) dwarfing genes. Journal of Agricultural Science 128:11-25. DOI: 10.1017/s0021859696003942.

Fosket D. (1994) Plant growth and development, a molecular approach Academic Press, San Diego.

Foulkes M.J., Sylvester-Bradley R., Worland A.J., Snape J.W. (2004) Effects of a photoperiod-response gene Ppd-D1 on yield potential and drought resistance in UK winter wheat. Euphytica 135:63-73. DOI: 10.1023/B:EUPH.0000009542.06773.13.

Gale M.D., Youssefian S. (1985) The genetics of dwarism in wheat, in: G. Russell (Ed.), Progres in Plant Breding, Butterworths, London, UK.

Ganal M., Roder M. (2007) Microsatellite and SNP Markers in Wheat breeding, in: R. Varshney and R. Tuberosa (Eds.), Genomics Assisted Crop Improvement: Genomics Applications in Crops, Springer. pp. 1-24.

Goncharov N.P. (2004) Response to vernalization in wheat: its quantitative or qualitative nature. Cereal Research Communications 32:323-330.

Green A.J., Berger G., Griffey C.A., Pitman R., Thomason W., Balota M., Ahmed A. (2012) Genetic Yield Improvement in Soft Red Winter Wheat in the Eastern United States from 1919 to 2009. Crop Science 52:2097-2108. DOI: 10.2135/cropsci2012.01.0026.

Guedira M., Maloney P., Xiong M., Petersen S., Murphy J.P., Marshall D., Johnson J., Harrison S., Brown-Guedira G. (2014) Vernalization Duration Requirement in Soft Winter Wheat is Associated with Variation at the Locus. Crop Science.

Gupta P.K., Langridge P., Mir R.R. (2010) Marker-assisted wheat breeding: present status and future possibilities. Molecular Breeding 26:145-161. DOI: 10.1007/s11032-009-9359-7.

Hanocq E., Niarquin M., Heumez E., Rousset M., Le Gouis J. (2004) Detection and mapping of QTL for earliness components in a bread wheat recombinant inbred lines population. Theoretical and Applied Genetics 110:106-115. DOI: 10.1007/s00122-004-1799-1.

Hayes B.J., Bowman P.J., Chamberlain A.C., Verbyla K., Goddard M.E. (2009) Accuracy of genomic breeding values in multi-breed dairy cattle populations. Genet Select Evol 41:1.

He C., Holme J., Anthony J. (2014a) SNP Genotyping: The KASP Assay. Methods in molecular biology (Clifton, N.J.) 1145:75-86. DOI: 10.1007/978-1-4939-0446-4_7.

He J., Zhao X., Laroche A., Lu Z.-X., Liu H., Li Z. (2014b) Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. Frontiers in Plant Science 5:484. DOI: 10.3389/fpls.2014.00484.

Hedden P. (2010) Green Revolution Genes, Plant Physiology Online. pp. Essay 20.1.

Heffner EL, Lorenz AJ, Jannink J-L, Sorrells ME (2010) Plant breeding with genomic selection: gain per unit time and cost. Crop science 50 (5):1681-1690

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

Hernández-Bautista A., Lobato-Ortiz R., García-Zavala J.J., Parra-Gómez M.A., Cadeza-Espinosa M., Canela-Doñan D., Cruz-Izquierdo S., Chávez-Servia J.L. (2016) Implications of genomic selection for obtaining F 2: 3 families of tomato. Sci Hort 207:7-13.

Hoffstetter A., Cabrera A., Huang M., Sneller C. (2016a) Optimizing Training Population Data and Validation of Genomic Selection for Economic Traits in Soft Winter Wheat. G3 (Bethesda). 116.032532.

Hoffstetter A., Cabrera A., Sneller C. (2016b) Identifying Quantitative Trait Loci for Economic Traits in an Elite Soft Red Winter Wheat Population. Crop Sci 56:1-12.

Hospital F. (2009) Challenges for effective marker-assisted selection in plants. Genetica 136:303-310. DOI: 10.1007/s10709-008-9307-1.

Huang M., Cabrera A., Hoffstetter A., Griffey C., Van Sanford D., Costa J., McKendry A., Chao S., Sneller C. (2016) Genomic selection for wheat traits and trait stability. Theor Appl Genet. 1-14.

Huang Y.-F., Poland J.A., Wight C.P., Jackson E.W., Tinker N.A. (2014) Using genotyping-by-sequencing (GBS) for genomic discovery in cultivated oat.

International Wheat Genome Sequencing C. (2014) A chromosome-based draft sequence of the hexaploid bread wheat (Triticum aestivum) genome. Science (New York, N.Y.) 345:1251788-1251788. DOI: 10.1126/science.1251788.

Iqbal M., Fayyaz M., Shahzad A., Ahmed I., Ali G.M., Masood S., Spaner D. (2011a) Effect of vernalization on grain fill duration and grain weight in SPRING WHEAT. Pakistan Journal of Botany 43:689-694.

Iqbal M., Navabi A., Yang R.-C., Salmon D.F., Spaner D. (2007) Molecular characterization of vernalization response genes in Canadian spring wheat. Genome 50:511-516. DOI: 10.1139/g07-028.

Iqbal M., Shahzad A., Ahmed I. (2011b) Allelic variation at the Vrn-A1, Vrn-B1, Vrn-D1, Vrn-B3 and Ppd-D1a loci of Pakistani spring wheat cultivars. Electronic Journal of Biotechnology 14. DOI: 10.2225/vol14-issue1-fulltext-6.

Iquira E., Humira S., Francois B. (2015) Association mapping of QTLs for sclerotinia stem rot resistance in a collection of soybean plant introductions using a genotyping by sequencing (GBS) approach. Bmc Plant Biology 15. DOI: 10.1186/s12870-014-0408-y.

Islam M.S., Thyssen G.N., Jenkins J.N., Fang D.D. (2015) Detection, validation, and application of genotyping-by-sequencing based single nucleotide polymorphisms in upland cotton. The Plant Genome 8.

Iwaki K., Nakagawa K., Kuno H., Kato K. (2000) Ecogeographical differentiation in east Asian wheat, revealed from the geographical variation of growth habit and Vrn genotype. Euphytica 111:137-143. DOI: 10.1023/a:1003862401570.

Jia J., Zhao S., Kong X., Li Y., Zhao G., He W., Appels R., Pfeifer M., Tao Y., Zhang X., Jing R., Zhang C., Ma Y., Gao L., Gao C., Spannagl M., Mayer K.F.X., Li D., Pan S., Zheng F., Hu Q., Xia X., Li J., Liang Q., Chen J., Wicker T., Gou C., Kuang H., He G., Luo Y., Keller B., Xia Q., Lu P., Wang J., Zou H., Zhang R., Xu J., Gao J., Middleton C., Quan Z., Liu G., Wang J., Yang H., Liu X., He Z., Mao L., Wang J., Int Wheat Genome Sequencing C. (2013) Aegilops tauschii draft genome sequence reveals a gene repertoire for wheat adaptation. Nature 496:91-95. DOI: 10.1038/nature12028.

Kamran A., Iqbal M., Spaner D. (2014a) Flowering time in wheat (Triticum aestivum L.): a key factor for global adaptability. Euphytica 197:1-26. DOI: 10.1007/s10681-014-1075-7.

Kamran A., Randhawa H.S., Pozniak C., Spaner D. (2013) Phenotypic Effects of the Flowering Gene Complex in Canadian Spring Wheat Germplasm. Crop Science 53:84-94. DOI: 10.2135/cropsci2012.05.0313.

Kamran A., Randhawa H.S., Yang R.-C., Spaner D. (2014b) The effect of VRN1 genes on important agronomic traits in high-yielding Canadian soft white spring wheat. Plant Breeding 133:321-326. DOI: 10.1111/pbr.12149.

Kanazin V., Talbert H., See D., DeCamp P., Nevo E., Blake T. (2002) Discovery and assay of single-nucleotide polymorphisms in barley (Hordeum vulgare). Plant Molecular Biology 48:529-537. DOI: 10.1023/a:1014859031781.

Kesawat M., Das B. (2009) Molecular Markers: It's Application in Crop Improvement. J. Crop Sci. Biotech. 12:169-181. DOI: 10.1007/s12892-009-0124-6.

Kim D.-H., Doyle M.R., Sung S., Amasino R.M. (2009) Vernalization: Winter and the Timing of Flowering in Plants. Annual Review of Cell and Developmental Biology 25:277-299. DOI: 10.1146/annurev.cellbio.042308.113411.

Kiss T., Balla K., Veisz O., Láng L., Bedő Z., Griffiths S., Isaac P., Karsai I. (2014) Allele frequencies in the VRN-A1, VRN-B1 and VRN-D1 vernalization response and PPD-B1 and PPD-D1 photoperiod sensitivity genes, and their effects on heading in a diverse set of wheat cultivars (Triticum aestivum L.). Molecular Breeding 34:297-310. DOI: 10.1007/s11032-014-0034-2.

Kolev S., Ganeva G., Christov N., Belchev I., Kostov K., Tsenov N., Rachovska G., Landgeva S., Ivanov M., Abu-Mhadi N., Todorovska E. (2010) Allele variation in loci for adaptive response and plant height and its effect on grain yield in wheat. Biotechnology & Biotechnological Equipment 24:1807-1813. DOI: 10.2478/v10133-010-0042-2.

Kolev S., Vassilev D., Kostov K., Todorovska E. (2011) Allele variation in loci for adaptive response in Bulgarian wheat cultivars and landraces and its effect on heading date. Plant Genetic Resources-Characterization and Utilization 9:251-255. DOI: 10.1017/s1479262111000475.

Kumar S., Sharma V., Chaudhary S., Tyagi A., Mishra P., Priyadarshini A., Singh A. (2012) Genetics of flowering time in bread wheat Triticum aestivum: complementary interaction between vernalization-insensitive and photoperiod-insensitive mutations imparts very early flowering habit to spring wheat. Journal of Genetics 91:33-47.

Law C.N., Snape J.W., Worland A.J. (1978a) Genetic relationship between height and yield in wheat. Heredity 40:133-151. DOI: 10.1038/hdy.1978.13.

Law C.N., Sutka J., Worland A.J. (1978b) Genetic study of day-length response in wheat. Heredity 41:185-191. DOI: 10.1038/hdy.1978.87.

Ling H.-Q., Zhao S., Liu D., Wang J., Sun H., Zhang C., Fan H., Li D., Dong L., Tao Y., Gao C., Wu H., Li Y., Cui Y., Guo X., Zheng S., Wang B., Yu K., Liang Q., Yang W., Lou X., Chen J., Feng M., Jian J., Zhang X., Luo G., Jiang Y., Liu J., Wang Z., Sha Y., Zhang B., Wu H., Tang D., Shen Q., Xue P., Zou S., Wang X., Liu X., Wang F., Yang Y., An X., Dong Z., Zhang K., Zhang X., Luo M.-C., Dvorak J., Tong Y., Wang J., Yang H., Li Z., Wang D., Zhang A., Wang J. (2013) Draft genome of the wheat A-genome progenitor Triticum urartu. Nature 496:87-90. DOI: 10.1038/nature11997.

Lorenzana R.E., Bernardo R. (2009) Accuracy of genotypic value predictions for marker-based selection in biparental plant populations. Theor Appl Genet 120:151-161.

Mason RE, Addison CK, Babar A, Acuna A, Lozada D, Subramanian N, Arguello MN, Miller RG, Brown-Guedira G, Guedira M, Johnson J (2017) Diagnostic Markers for Vernalization and Photoperiod Loci Improve Genomic Selection for Grain Yield and Spectral Reflectance in Wheat. Crop Science. doi:10.2135/cropsci2017.06.0348.

Mayer K.F.X., Rogers J., Dolezel J., Pozniak C., Eversole K., Feuillet C., Gill B., Friebe B., Lukaszewski A.J., Sourdille P., Endo T.R., *et al.* (2014) A chromosome-based draft sequence of the hexaploid bread wheat (Triticum aestivum) genome. Science 345. DOI: 10.1126/science.1251788.

McIntosh R., Yamazaki Y., Dubcovsky J., Rogers J., Morris C., Somers D., Appels R., Devos K. (2008) Catalogue of Gene Symbols in Wheat, 11th International Wheat Genetics Symposium, Brisbane, Australia.

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

Muranty H., Troggio M., Sadok I.B., Rifaï M.A., Auwerkerken A., Banchi E., Velasco R., Stevanato P., van de Weg W.E., Di Guardo M., Kumar S., Laurens F., Bink M.C.A.M. (2015) Accuracy and responses of genomic selection on key traits in apple breeding. Hort Res 2:15060. DOI: 10.1038/hortres.2015.60.

Mohan M., Nair S., Bhagwat A., Krishna T., Yano M., Bhatia C., Sasaki T. (1997) Genome mapping, molecular markers and marker-assisted selection in crop plants. Molecular Breeding 3:87-103.

Moose S.P., Mumm R.H. (2008) Molecular Plant Breeding as the Foundation for 21st Century Crop Improvement. Plant Physiology 147:969-977. DOI: 10.1104/pp.108.118232.

Morgan G., Bean B., Baughman T. (2008) Wheat Newsletter, AgriLife Extension Texas A & M System, College Station, TX.

Myles S., Peiffer J., Brown P.J., Ersoz E.S., Zhang Z., Costich D.E., Buckler E.S. (2009) Association Mapping: Critical Considerations Shift from Genotyping to Experimental Design. The Plant Cell 21:2194-2202. DOI: 10.1105/tpc.109.068437.

Nakaya A., and S.N. Isobe (2012) Will genomic selection be a practical method for plant breeding? Ann Bot:mcs109.

Neumann K., Kobiljski B., Denčić S., Varshney R.K., Börner A. (2011) Genome-wide association mapping: a case study in bread wheat (Triticum aestivum L.). Molecular Breeding 27:37-58. DOI: 10.1007/s11032-010-9411-7.

Nocente F., Gazza L., Pasquini M. (2007) Evaluation of leaf rust resistance genes Lr1, Lr9, Lr24, Lr47 and their introgression into common wheat cultivars by marker-assisted selection. Euphytica 155:329-336. DOI: 10.1007/s10681-006-9334-x.

Olsen K.M., Halldorsdottir S.S., Stinchcombe J.R., Weinig C., Schmitt J., Purugganan M.D. (2004) Linkage disequilibrium mapping of Arabidopsis CRY2 flowering time alleles. Genetics 167:1361-1369. DOI: 10.1534/genetics.103.024950.

Onogi A., Watanabe M., Mochizuki T., Hayashi T., Nakagawa H., Hasegawa T., Iwata H. (2016) Toward integration of genomic selection with crop modelling: the development of an integrated approach to predicting rice heading dates. Theor Appl Genet 129:805-817.

Pearce S., Saville R., Vaughan S.P., Chandler P.M., Wilhelm E.P., Sparks C.A., Al-Kaff N., Korolev A., Boulton M.I., Phillips A.L., Hedden P., Nicholson P., Thomas S.G. (2011) Molecular Characterization of Rht-1 Dwarfing Genes in Hexaploid Wheat. Plant Physiology 157:1820-1831. DOI: 10.1104/pp.111.183657.

Peng J.R., Richards D.E., Hartley N.M., Murphy G.P., Devos K.M., Flintham J.E., Beales J., Fish L.J., Worland A.J., Pelica F., Sudhakar D., Christou P., Snape J.W., Gale M.D., Harberd N.P. (1999) 'Green revolution' genes encode mutant gibberellin response modulators. Nature 400:256-261.

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:e32253.

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

Raman H., Stodart B., Ryan P.R., Delhaize E., Emebiri L., Raman R., Coombes N., Milgate A. (2010) Genome-wide association analyses of common wheat (Triticum aestivum L.) germplasm identifies multiple loci for aluminium resistance. Genome 53:957-966. DOI: 10.1139/g10-058.

Rebetzke G.J., Ellis M.H., Bonnett D.G., Condon A.G., Falk D., Richards R.A. (2011) The Rht13 dwarfing gene reduces peduncle length and plant height to increase grain number and yield of wheat. Field Crops Research 124:323-331. DOI: 10.1016/j.fcr.2011.06.022.

Rebetzke G.J., Ellis M.H., Bonnett D.G., Mickelson B., Condon A.G., Richards R.A. (2012) Height reduction and agronomic performance for selected gibberellin-responsive dwarfing genes in bread wheat (Triticum aestivum L.). Field Crops Research 126:87-96. DOI: 10.1016/j.fcr.2011.09.022.

Reynolds M., Foulkes J., Furbank R., Griffiths S., King J., Murchie E., Parry M., Slafer G. (2012) Achieving yield gains in wheat. Plant, Cell and Environment 35:1799–1823

Roder M., Korzun V., Wendehake K., Plaschke J., Tixier M., Leroy P., Ganal M. (1998) A Microsatellite Map of Wheat. Genetics 149:2007–2023.

Rousset M., Bonnin I., Remoue C., Falque M., Rhone B., Veyrieras J.-B., Madur D., Murigneux A., Balfourier F., Le Gouis J., Santoni S., Goldringer I. (2011) Deciphering the genetics of flowering time by an association study on candidate genes in bread wheat (Triticum aestivum L.). Theoretical and Applied Genetics 123:907-926. DOI: 10.1007/s00122-011-1636-2.

Roy J.K., Bandopadhyay R., Rustgi S., Balyan H.S., Gupta P.K. (2006) Association analysis of agronomically important traits using SSR, SAMPL and AFLP markers in bread wheat. Current Science 90:683-689.

Rutkoski J., Singh R., Huerta-Espino J., Bhavani S., Poland J., Jannink J.-L., Sorrells M. (2015) Efficient use of historical data for genomic selection: a case study of stem rust resistance in wheat. Plant Genome 8. Doi: 10.3835/plantgenome2014.09.0046.

Saintenac C., Jiang D., Wang S., Akhunov E. (2013) Sequence-Based Mapping of the Polyploid Wheat Genome. G3-Genes Genomes Genetics 3:1105-1114. DOI: 10.1534/g3.113.005819.

Salvi S., Sponza G., Morgante M., Tomes D., Niu X., Fengler K.A., Meeley R., Ananiev E.V., Svitashev S., Bruggemann E., Li B., Hainey C.F., Radovic S., Zaina G., Rafalski J.A., Tingey S.V., Miao G.-H., Phillips R.L., Tuberosa R. (2007) Conserved noncoding genomic sequences associated with a flowering-time quantitative trait locus in maize. Proceedings of the National Academy of Sciences of the United States of America 104:11376-11381. DOI: 10.1073/pnas.0704145104.

Santra D.K., Santra M., Allan R.E., Campbell K.G., Kidwell K.K. (2009) Genetic and Molecular Characterization of Vernalization Genes Vrn-A1, Vrn-B1, and Vrn-D1 in Spring Wheat Germplasm from the Pacific Northwest Region of the USA. Plant Breeding 128:576-584. DOI: 10.1111/j.1439-0523.2009.01681.x.

Scarth R., Law C.N. (1983) The location of the photoperiod gene, ppd2 and an additional genetic-factor for ear-emergence time on chromosome-2b of wheat. Heredity 51:607-619. DOI: 10.1038/hdy.1983.73.

Schnable P.S., Liu S., Wu W. (2013) Genotyping by next-generation sequencing, Google Patents.

Seki M., Chono M., Nishimura T., Sato M., Yoshimura Y., Matsunaka H., Fujita M., Oda S., Kubo K., Kiribuchi-Otobe C., Kojima H., Nishida H., Kato K. (2013) Distribution of photoperiod-insensitive allele Ppd-A1a and its effect on heading time in Japanese wheat cultivars. Breeding Science 63:309-316. DOI: 10.1270/jsbbs.63.309.

Semagn K., Babu R., Hearne S., Olsen M. (2013) Single nucleotide polymorphism genotyping using Kompetitive Allele Specific PCR (KASP): overview of the technology and its application in crop improvement. Molecular Breeding 33:1-14. DOI: 10.1007/s11032-013-9917-x.

Shaw L.M., Turner A.S., Herry L., Griffiths S., Laurie D.A. (2013) Mutant Alleles of Photoperiod-1 in Wheat (Triticum aestivum L.) That Confer a Late Flowering Phenotype in Long Days. Plos One 8. DOI: 10.1371/journal.pone.0079459.

Shcherban A.B., Emtseva M.V., Efremova T.T. (2012) Molecular Genetical Characterization of Vernalization Genes Vrn-A1, Vrn-B1 and Vrn-D1 in Spring Wheat Germplasm from Russia and Adjacent Regions. Cereal Research Communications 40:351-361. DOI: 10.1556/crc.40.2012.3.4.

Shewry P.R. (2009) Wheat. Journal of Experimental Botany 60:1537-1553. DOI: 10.1093/jxb/erp058.

Shindo C., Tsujimoto H., Sasakuma T. (2003) Segregation analysis of heading traits in hexaploid wheat utilizing recombinant inbred lines. Heredity 90:56-63. DOI: 10.1038/sj.hdy.6800178.

Singh S.K., Singh A.M., Jain N., Singh G.P., Ahlawat K., Ravi I. (2013) Molecular Characterization of Vernalization and Photoperiod Genes in Wheat Varieties from Different Agro-climatic Zones of India. Cereal Research Communications 41:376-387. DOI: 10.1556/crc.2013.0017.

Slafer G.A. (2003) Genetic basis of yield as viewed from a crop physiologist's perspective. Annals of Applied Biology 142:117-128. DOI: 10.1111/j.1744-7348.2003.tb00237.x.

Slafer G.A., Araus J.L. (2007) Physiological traits for improving wheat yield under a wide range of conditions.

Snape J.W., Butterworth K., Whitechurch E., Worland A.J. (2001) Waiting for fine times: genetics of flowering time in wheat. Euphytica 119:185-190. DOI: 10.1023/a:1017594422176.

Somers D.J., Kirkpatrick R., Moniwa M., Walsh A. (2003) Mining single-nucleotide polymorphisms from hexaploid wheat ESTs. Genome 46:431-437. DOI: 10.1139/g03-027.

Sonah H., O'Donoughue L., Cober E., Rajcan I., Belzile F. (2015) Identification of loci governing eight agronomic traits using a GBS-GWAS approach and validation by QTL mapping in soya bean. Plant Biotechnology Journal 13:211-221. DOI: 10.1111/pbi.12249.

Spindel J., Begum H., Akdemir D., Virk P., Collard B., Redoña E., Atlin G., Jannink J.-L., 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:e1004982.

Sun H., Guo Z., Gao L., Zhao G., Zhang W., Zhou R., Wu Y., Wang H., An H., Jia J. (2014) DNA methylation pattern of Photoperiod-B1 is associated with photoperiod insensitivity in wheat (Triticum aestivum). New Phytologist 204:682-692. DOI: 10.1111/nph.12948.

Sung S., Amasino R. (2004) Vernalization and epigenetics: how plants remember winter. Current Opinion in Plant Biology 7:4-10. DOI: 10.1016/j.pbi.2003.11.010.

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

Thornsberry J.M., Goodman M.M., Doebley J., Kresovich S., Nielsen D., Buckler E.S. (2001) Dwarf8 polymorphisms associate with variation in flowering time. Nature Genetics 28:286-289. DOI: 10.1038/90135.

Tucker D.M., Griffey C.A., Liu S., Maroof M.A.S. (2006) Potential for effective marker-assisted selection of three quantitative trait loci conferring adult plant resistance to powdery mildew in elite wheat breeding populations. Plant Breeding 125:430-436. DOI: 10.1111/j.1439-0523.2006.01233.x.

Uitdewilligen J.G., Wolters A.-M.A., Bjorn B., Borm T.J., Visser R.G., van Eck H.J. (2013) A next-generation sequencing method for genotyping-by-sequencing of highly heterozygous autotetraploid potato.

US Census B. (2013) International Data Base, World Population: 1950-2050. http://www.census.gov/population/international/data/idb/worldpopgraph.php.

USDA-ERS. (2014) Wheat data, http://www.ers.usda.gov/data-products/wheat-data.aspx#25171.

USDA. (2013) National Agricultural Statistics Service. Crop Production.

Vogel B. (2009) Marker-Assisted Selection: A non-invasive biotechnology alternative to genetic engineering of plant varieties, in: S. Erwood and N. Truchi (Eds.), Smart Breeding, Green Peace International. pp. 1-28.

Wang G., Leonard J.M., von Zitzewitz J., Peterson C.J., Ross A.S., Riera-Lizarazu O. (2014) Marker-trait association analysis of kernel hardness and related agronomic traits in a core collection of wheat lines. Molecular Breeding 34:177-184. DOI: 10.1007/s11032-014-0028-0.

Wang Y., Li J. (2011) Branching in rice. Current opinion in plant biology 14:94-99.

Whankaew S., Sraphet S., Thaikert R., Smith D.R., Triwitayakorn K. (2012) Characterization of microsatellite markers in cassava based on microsatellite-AFLP technique. Genetics and Molecular Research 11:1319-1326. DOI: 10.4238/2012.May.15.2.

William H.M., Trethowan R., Crosby-Galvan E.M. (2007) Wheat breeding assisted by markers: CIMMYT's experience. Euphytica 157:307-319. DOI: 10.1007/s10681-007-9405-7.

Wilson L.M., Whitt S.R., Ibanez A.M., Rocheford T.R., Goodman M.M., Buckler E.S. (2004) Dissection of maize kernel composition and starch production by candidate gene association. Plant Cell 16:2719-2733. DOI: 10.1105/tpc.104.025700.

Worland A.J. (1996) The influence of flowering time genes on environmental adaptability in European wheats. Euphytica 89:49-57. DOI: 10.1007/bf00015718.

Worland A.J., Borner A., Korzun V., Li W.M., Petrovic S., Sayers E.J. (1998) The influence of photoperiod genes on the adaptability of European winter wheats (Reprinted from Wheat: Prospects for global improvement, 1998). Euphytica 100:385-394. DOI: 10.1023/a:1018327700985.

Wurschum T., Langer S.M., Longin C.F., Korzun V., Akhunov E., Ebmeyer E., Schachschneider R., Schacht J., Kazman E., Reif J.C. (2013) Population structure, genetic diversity and linkage disequilibrium in elite winter wheat assessed with SNP and SSR markers. Theor Appl Genet 126:1477-86. DOI: 10.1007/s00122-013-2065-1.

Yan L., Helguera M., Kato K., Fukuyama S., Sherman J., Dubcovsky J. (2004) Allelic variation at the VRN-1 promoter region in polyploid wheat. Theoretical and Applied Genetics 109:1677-1686. DOI: 10.1007/s00122-004-1796-4.

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 of the United States of America 100:6263-6268. DOI: 10.1073/pnas.0937399100.

Zane L., Bargelloni L., Patarnello T. (2002) Strategies for microsatellite isolation: a review. Molecular Ecology 11:1-16. DOI: 10.1046/j.0962-1083.2001.01418.x.

Zhang D., Bowden R.L., Yu J., Carver B.F., Bai G. (2014a) Association analysis of stem rust resistance in U.S. winter wheat. PloS one 9:e103747-e103747. DOI: 10.1371/journal.pone.0103747.

Zhang J., Dell B., Biddulph B., Khan N., Xu Y., Luo H., Appels R. (2014b) Vernalization gene combination to maximize grain yield in bread wheat (Triticum aestivum L.) in diverse environments. Euphytica 198:439-454. DOI: 10.1007/s10681-014-1120-6.

Zhang X.K., Xiao Y.G., Zhang Y., Xia X.C., Dubcovsky J., He Z.H. (2008) Allelic variation at the vernalization genes Vrn-A1, Vrn-B1, Vrn-D1, and Vrn-B3 in Chinese wheat cultivars and their association with growth habit. Crop Science 48:458-470. DOI: 10.2135/cropsci2007.06.0355.

Zhao Y., Gowda M., Liu W., Würschum T., Maurer H.P., Longin F.H., Ranc N., Reif J.C. (2012) Accuracy of genomic selection in European maize elite breeding populations. Theor Appl Genet 124:769-776.

Zhao K., Tung C.-W., Eizenga G.C., Wright M.H., Ali M.L., Price A.H., Norton G.J., Islam M.R., Reynolds A., Mezey J., McClung A.M., Bustamante C.D., McCouch S.R. (2011) Genome-wide association mapping reveals a rich genetic architecture of complex traits in Oryza sativa. Nat Commun 2:467. DOI: http://www.nature.com/ncomms/journal/v2/n9/suppinfo/ncomms1467_S1.html.

Zhong S., Dekkers J.C., Fernando R.L., Jannink J.-L. (2009) Factors affecting accuracy from genomic selection in populations derived from multiple inbred lines: a barley case study. Genetics 182:355-364.

Zhou R.H., Zhu Z.D., Kong X.Y., Huo N.X., Tian Q.Z., Li P., Jin C.Y., Dong Y.C., Jia J.Z. (2005) Development of wheat near-isogenic lines for powdery mildew resistance. Theoretical and Applied Genetics 110:640-648. DOI: 10.1007/s00122-004-1889-0.

Zhu C., Gore M., Buckler E.S., Yu J. (2008) Status and Prospects of Association Mapping in Plants. Plant Genome 1:5-20. DOI: 10.3835/plantgenome2008.02.0089.

CHAPTER II

GENOME-WIDE ASSOCIATION STUDY FOR GRAIN YIELD AND AGRONOMIC TRAITS IN SOFT WINTER WHEAT

Abstract

Genome-wide association studies (GWAS) are useful to facilitate crop improvement via enhanced knowledge of marker-trait associations (MTA). A GWAS for grain yield (GY), yield components, and agronomic traits was conducted using a diverse panel of 239 soft red winter wheat (*Triticum aestivum*) genotypes evaluated across two growing seasons and eight site-years. Analysis of variance showed significant environment, genotype, and genotype-by-environment effects for GY and yield components. Broad sense heritability of GY (H^2 =0.48) was moderate compared to other traits including plant height (H^2 = 0.81) and kernel weight (H^2 =0.77). There were 112 significant MTA ($p < 0.0005$) detected for eight measured traits using compressed mixed linear models and 5,715 single nucleotide polymorphism markers. MTA for GY and agronomic traits coincided with previously reported QTL for winter and spring wheat. Highly significant marker trait associations for GY showed an overall negative allelic effect for the minor allele, indicating selection against these alleles by breeders. Markers associated with multiple traits observed on chromosomes 1A, 2D, 3B, and 4B with positive minor effects serve as potential targets for marker assisted breeding to select for improvement of GY and related traits. Following marker validation, these multi-trait loci have the potential to be utilized for MAS to improve GY and adaptation of soft red winter wheat.

Introduction

Identification of marker-trait associations (MTA) is a first step toward marker-assisted selection (MAS), which has become an important tool for accelerating varietal improvement and rate of genetic gain (Moose and Mumm 2008; Wang et al. 2014b). Whole-genome mapping approaches such as genome-wide association studies (GWAS) have recently become a popular alternative to bi-parental quantitative trait loci (QTL) mapping for identifying MTA in plant populations, due in large part to recent advances in high-throughput sequencing and genotyping platforms that have decreased cost and increased discovery of marker polymorphisms (Patel et al. 2015; Ruggieri et al. 2014; Thomson 2014).

GWAS use the concept of linkage disequilibrium (LD), the non-random co-segregation of alleles at multiple loci, to survey genomic regions that render significant variation to phenotypes (Breseghello and Sorrells 2006a; Flint-Garcia et al. 2003). A primary advantage of GWAS is exploitation of recombination events that have occurred over an individual's evolutionary history using a diverse population (Myles et al. 2009), consequently resulting in a higher mapping resolution compared to a bi-parental approach (Zhu et al. 2008). Additionally, GWAS allows for a much larger gene pool to be surveyed and screened for genetic variation in traits of interest (Neumann et al. 2011; Zhao et al. 2011).

Previous studies have established the usefulness of GWAS in identifying regions affecting variation for GY and adaptation traits in bread wheat (*Triticum aestivum*). Wang et al. (2014a) reported MTA for kernel hardness, kernel weight, grain protein concentration, grain volume, and plant height in a diverse set of 94 wheat lines. Prior to this, Neumann et al. (2011) conducted GWAS for 20 agronomic traits in a winter wheat core collection using diversity array technology (DArT) markers where significant MTA were detected for plant height, GY, and

disease resistance. Sukumaran et al. (2014) and Lopes et al. (2015) identified genomic regions associated with GY and yield-related traits in a wheat association mapping initiative (WAMI) panel consisting of 287 elite lines of spring wheat from CIMMYT, Mexico. Sehgal et al. (2017) recently identified regions affecting GY and yield stability and their epistatic interactions using a large elite panel of CIMMYT spring wheat genotypes under multiple environments.

Hoffstetter et al. (2016) identified important loci governing GY and other economic traits in an elite collection of soft red winter wheat (SRWW) lines adapted to the northeastern US, while Addison et al. (2016) determined genomic regions affecting GY potential utilizing a bi-parental approach in a population derived from two elite SRWW cultivars. Except for these studies, reports on MTA for GY and related traits for US soft winter wheat remain limited and hence there is a need to identify yield-related QTL in current soft red winter wheat germplasm. The objectives of this study were to perform GWAS for GY and agronomic traits and to examine population structure and linkage disequilibrium of a diverse panel of SRWW lines adapted to the southern region of the US using genome-wide SNP markers. Information from this research will serve as a valuable resource for genetic improvement of GY and related traits via marker-assisted selection approaches.

Materials and Methods

Plant material and experimental design

The association mapping panel (AMP) used for this study consisted of 239 inbred lines of SRWW, including cultivars from the SunGrains® (Southeastern University Grains) small grain breeding and genetics group, publicly and privately developed cultivars, and genotypes adapted to the southeastern region of the US. Trials were drill seeded in seven row plots (1.5m width x 4.5m length) at a rate of 118 kg of seed hectare^{-1}. The AMP was evaluated in a total of eight

high yield potential site-years that included two environments in the 2013-2014 season and six environments in the 2014-2015 season. Locations included Fayetteville (FAY14, FAY15), Marianna (MAR15), Stuttgart (STU14, STU15), Keiser (KEI15) and Rohwer (ROH15), in the state of Arkansas; and Okmulgee, in the state of Oklahoma (OKL15), US. All locations belong to the west south-central US region of SRWW commercial production.

The AMP was sown in an augmented incomplete block design (Federer and Raghavarao 1975; Federer and Crossa 2012), with two repeated check lines (Jamestown and Pioneer Brand 26R20) with unreplicated lines on each location. The random nature of the new treatments and blocking variables are considered in augmented designs resulting in a more efficient analysis (Federer et al. 2001). In all locations except for OKL15, the experimental field was divided into 24 incomplete blocks, each containing 10 different AMP genotypes and both checks. For OKL15, unequal incomplete block sizes, k, were used, where: k= 10 for IB 1-19; k=20 for IB 20-23 and k= 18 for IB 24. Planting and harvest dates and trial management varied based on recommendations at each location for maximizing yield potential but included routine fungicide applications to control foliar diseases.

Trait measurements

Grain yield (GY) in kg ha^{-1} was recorded by harvesting whole plots, weighing the grain, and adjusting values to 13% moisture content. Heading date (HD) was recorded as the date when 50% of plants from the whole plot had fully visible spikes and reported in Julian Days. Plant height (PH) was recorded from the soil surface to tip of the spike, excluding awns when present. Kernel weight (KW) was determined by counting 1000 seeds using a Seedburo® 801 seed counter (Chicago, IL, USA). Peduncle length (PL) was measured as the length of the uppermost internode, in cm, averaged across ten culms $plot^{-1}$. Spike length (SL) was taken as the

measurement from the base to tip of the spike (excluding awns), in cm, averaged across ten spikes plot^{-1}. Kernel number spike^{-1} (KNS) and kernel weight spike^{-1} (KWS) were estimated by hand-harvesting 50 spike-bearing culms from each plot at maturity prior to harvesting of whole plots.

Statistical analysis

Phenotypic data were analyzed following procedures described by Wolfinger et al. (1997) for analysis of augmented designs using PROC MIXED in SAS v.9.4 (SAS Institute 2011). Genotypes, incomplete blocks, environments, incomplete blocks nested within environments and genotype-by-environment interactions were regarded as random effects. Adjusted means represented as least square means (LSM) for each genotype were estimated using a restricted maximum likelihood (REML) approach for each site-year. Broad sense heritability (H^2) was calculated for each trait using TYPE3 sum of squares from the adjusted means, with the formula:

$$H^2 = \frac{\sigma_G^2}{\sigma_G^2 + \sigma_{\frac{GEI}{e}}^2 + \sigma_{\frac{E}{er}}^2},$$

where σ_G^2, σ_{GEI}^2 and σ_E^2 variances due to genotype, genotype-by-environment, and error, respectively; and *e* and *r* are the number of environments and replications. Associations between traits and environments were explored using principal component analysis (PCA) with the contribution of each variable to the first two principal components (PC) illustrated using bi-plots. The PROC CORR procedure in SAS v.9.4 was used to calculate correlation of normalized means of phenotypes across environments.

SNP marker genotyping

DNA was isolated from each sample following a CTAB extraction procedure modified from Pallotta et al. (2003). Samples were genotyped using the Illumina 9K *iSelect* assays for wheat previously described by Cavanagh et al. (2013) through the USDA-ARS Eastern Regional Small

Grain Genotyping Laboratory, Raleigh, NC. Marker data polymorphisms of 8,632 SNPs were scored using the GenomeStudio® software (Illumina, San Diego, USA). After filtering, 5,715 polymorphic markers with minor allele frequency (MAF) ≥0.04% and less than 10% missing data remained and were used to perform GWAS. SNPs with low MAF were included to capture rare allele variants (MAF < 0.01) which could potentially explain additional variability within the measured traits (Lee et al. 2014).

In addition to the 9K *iSelect* assay, the AMP was genotyped using KASP® allele-specific SNP markers (LGC Genomics, UK) diagnostic for height (*Rht-B1*, *Rht-D1*), vernalization (*Vrn-A1* and *Vrn-B1*) and photoperiod (*Ppd-B1*, *Ppd-D1*) loci (Guedira et al. 2014; Guedira et al. 2016). Reactions were performed in a total volume of 5 µL (2.5 µL KASP® mix and 2.5 µL DNA sample (50 ng)), following manufacturer's instructions with minor modifications. Conditions for thermal cycling were as follows: 94°C for 15 min; 94°C for 20 sec and 65-58°C (decrement of 0.8°C per cycle) for 9 cycles; 94°C 20 sec and 57°C for one minute for 25 cycles; 35°C for 3 min and a plate read step. An additional thermal cycling step (94°C for 20 seconds followed by 57.0°C for one minute for 2 cycles; and 35°C for one minute and a plate read step) was used as needed to improve accuracy and precision of clustering.

Linkage disequilibrium, population structure, and genetic diversity

Coefficients of linkage disequilibrium (LD), represented by the square of allele frequency correlations, r^2 (Weir and Cockerham 1996), were calculated using the program TASSEL 5.2.33 (Bradbury et al. 2007). Imputation for missing genotype data was done using a numeric, Euclidean-based distance method in TASSEL, with minimum and maximum allele frequencies set to 0.05 and 1.0, respectively. Pairwise r^2 values were plotted against genetic distance (in cM; based on genetic linkage map by Cavanagh et al (2013)) and a locally weighted polynomial

regression (LOESS) curve (Cleveland 1979) was fitted on the LD plot using RStudio® (R Development Core Team, 2010) using the 'loess' function. Critical values were estimated by performing a square root transformation of corresponding r^2 estimates for unlinked marker pairs (distance > 50 cM) and then taking the 95th percentile of this distribution (Breseghello and Sorrells 2006b). The intersection of LOESS line and r^2 critical value was regarded as the distance where LD starts to decay (Laido et al. 2014; Nielsen et al. 2014). A $p < 0.005$ was considered the significance threshold for marker pairs to be in LD with each other.

Population stratification was assessed using the program STRUCTURE (Pritchard et al. 2000) applying an admixture model, a burn-in of 10000 iterations followed by 10000 Monte Carlo Markov Chain (MCMC) replicates and number of clusters *(K)* set in the range 2-10, with number of replications per *K* equal to 10. The true number of clusters which best fit the data was inferred using the Evanno criterion, which uses an ad hoc statistic ΔK based on rate of change in the log probability of data between successive values of *K* (Evanno et al. 2005). Likelihood scores and results from STRUCTURE were collated and visualized using the program STRUCTURE Harvester (Earl 2012). Bar plots for membership coefficients, *Q* for the AMP were plotted using the 'pophelper' package (Francis 2016) in RStudio®.

Analysis of molecular variance (AMOVA; Excoffier et al. 1992) was conducted using a ploidy independent infinite allele model (ρ) tested under 999 permutations implemented in the software Genodive (Meirmans and Van Tienderen 2004). Rho (ρ) is an analogue of the population differentiation coefficient (Fixation index, *Fst*) and is independent of the organism's ploidy level (Meirmans and Van Tienderen 2004). Fixation indices and pairwise *Gst* values of subpopulations were calculated using STRUCTURE and Genodive programs, respectively. *Fst* estimates the correlation of alleles within the same subgroup relative to the entire population

(Chao et al. 2010) while G_{st} compares heterozygosity within and between populations, considering a correction for a bias resulting from sampling a limited number of populations (Nei 1987).

GWAS for grain yield and agronomic traits

Association analyses was performed employing several model selections for a compressed mixed linear model (CMLM) implemented in the Genome Association Prediction Integrated Tool (GAPIT) (Lipka et al. 2012) package in RStudio®. Models included: (1) a naïve model, where only the kinship, K information, and no correction for population structure were applied (*K* only model); (2) a K-PC model (Zhao et al. 2007) where kinship information together with the first three principal components (PC) were included for GWAS; and (3) a K-Q approach, where a centered IBS (Identical by State) kinship method (Endelman and Jannink 2012) in TASSEL 5.2.33 and a population structure matrix derived from STRUCTURE were included in the model as fixed effects to address population structure. In addition to these models, marker scores for *Rht* and *Vrn* loci were included under the K and K-PC as covariates to correct their effects in identifying GY related MTA (Lopes et al. 2015).

The mixed model used to account for genetic relatedness in the AMP was as follows:

$$y = \mu + x\beta + u + e$$

where y is the vector of observed phenotype; μ is the mean; x is the genotype of the SNP; β is the effect of the SNP; $\boldsymbol{u}$ is the random effects due to genetic relatedness with Var ($\boldsymbol{u}$) = $\sigma^2_g\mathbf{K}$ and Var (e) = σ^2_e; **K** is the kinship matrix across all genotypes (Kang et al. 2008; Lopes et al. 2015). CMLM tests one marker at a time and considers the **u** and **K** matrices as the mean additive genetic relatedness between individuals to model polygenetic effects (Lipka et al. 2012).

A total of five combined datasets were used for GWAS, namely BLUP trait values calculated from adjusted means across all environments (ABLUP); BLUP values derived from 2014 site-years (BLUP14); BLUP from the 2015 site-years (BLUP15); BLUP from northern environments across the two years (Fayetteville, Keiser, AR; Okmulgee, OK; NBLUP), and from southern environments across the two years (Stuttgart, Marianna, Rohwer, AR; SBLUP).

The most reliable model for GWAS was identified by performing a 10-fold cross validation (CV) under a ridge regression best linear unbiased prediction (rrBLUP) model (Endelman 2011) for the most heritable trait on an ABLUP dataset, where kinship, K represented as a marker relationship matrix and scores for Q and PC as covariates were fitted on the model. A value of $p < 0.0005$ was considered the threshold for defining significant SNP due to deviations of observed quantile-quantile (QQ) plots and to further reduce Type I errors (Hoffstetter et al. 2016; Lopes et al. 2015). Manhattan plots were visualized using the 'qqman' package (Turner 2014) in RStudio®

Results

Genotype-by-environment interactions and trait heritability

FAY15 had the highest mean GY, followed by ROH15, and OKL15, while STU14, STU15, and FAY14 had the lowest. Significant genotype effects were observed for all traits indicating differential performance (Table 1). Genotype-by-environment interaction was highly significant for all traits. Incomplete block treatments as well as incomplete blocks nested within environments did not show a significant effect for measured phenotypic traits. Broad sense heritability (H^2) estimates ranged from 0.30 to 0.81, with PH the most heritable ($H^2 = 0.81$), followed by KW ($H^2 = 0.71$) and HD ($H^2 = 0.63$). GY was moderately heritable ($H^2 = 0.48$) while SL was the least heritable trait ($H^2 = 0.30$).

Principal components analyses (PCA) and phenotypic correlations

Results from PCA showed PC1 to explain 36.4% of the total variation for phenotypic traits and was positively associated with PL and negatively associated with all other traits (Fig. 1). PC2 contributed 20.1% of the total variation and was in positive correlation with GY and KNS. The PCA biplot was divided into two trait clusters: (1) GY and its components including KNS, KWS, and KW; and (2) HD and agronomic traits including PH, SL, and PL. Pearson correlation coefficients (*r*) further supported these PCA groupings as GY was strongly correlated with KW (r=0.48), KNS (r=0.67) and KWS (r =0.73) (Table 2). PH was positively correlated with PL (r= 0.49) and HD (r=0.19). Neither HD nor PH was significantly correlated with GY.

PCA biplot analyses for GY across site-years revealed separation based on year, with the 2014 (FAY14 and STU14) and 2015 (excluding MAR15) clustering separately (Fig.1). PC1 explained 21.9% of the variation for GY and was positively correlated with MAR15. PC2 contributed 15.2% of variation for GY across environments, was positively correlated with OKL15, STU14, FAY14, and MAR15 and was negatively correlated with STU15, FAY15, ROH15, and KEI15.

Analysis of LD

A total of 74,822 intrachromosomal pairs were in significant LD ($p < 0.005$) at the whole genome level (Appendix 1). Average distance of markers in significant LD was ~14.40 cM, while markers in complete LD (r^2=1.0) had an average distance of 1.71 cM for the whole genome. Genome D had the highest average distance for pairs in complete LD (3.14 cM), followed by Genomes B (1.90 cM) and A (1.34 cM). Average r^2 value for significant pairs across the whole genome was 0.32. Among the subgenomes, genome D also had the highest mean r^2 for all significant pairs (0.37), followed by genomes A (0.32) and B (0.31). LD was

estimated to decay at ~7 cM for the whole genome, while genome D had the highest extent of LD among the subgenomes, estimated at ~10 cM, compared to genomes A and B (both estimated at ~7 cM) (Appendix 2).

Population structure

Genetic structure was evaluated using 5,661 genome-wide SNP markers where markers linked to major genes were designated as fixed effects. Inference for the true number of clusters (*K)* using the Evanno criterion (Evanno et al. 2005) revealed the optimum number of subpopulations for this panel at *K*=3 (Appendix 3). Each entry was assigned to one of three subpopulations based on its largest value for coefficient of membership (*Q*). Fifty-nine lines were assigned to the first subgroup (*Q1*), 54 lines were assigned to the second subgroup, *Q2*, and 126 lines to the third subgroup, *Q3* (Appendix 4). There was no observable clustering based on geographic origin for the lines across the different subgroups. Analysis of molecular variance (AMOVA) further revealed the presence of within population variation, which accounted for 89.1% of the total variance (Appendix 5). Mean value for *Fst* was highest for *Q1* (0.69), followed by the *Q2* (0.43) and *Q3* (0.23) subpopulations (Appendix 6).

Genetic diversity for developmental genes

A total of 207 (87%) lines were semi-dwarfs, having a dwarfing allele in combination with a tall allele for either *Rht-B1* or *Rht-D1*. Two of the lines were double dwarfs, while 26 lines possessed wild-type tall alleles for both loci. Subgroup *Q3* had the highest number of semi-dwarf entries for both the *Rht-B1a*/*Rht-D1b* and *Rht-B1b*/*Rht-D1a* (semi dwarf) allelic combinations (106; 51.2%), in addition to 17 wild-type lines. Majority of lines possessing the photoperiod insensitive *Ppd-D1a* allele also belonged to the *Q3* subpopulation (56; 57.7%). Forty-seven of the entries (19.7%) had a short vernalization allele at the winter *vrn-1A* locus (*vrn-A1b*,

M_vrn_A1_ex4 locus) with 23 of these lines belonging to subgroup Q3, while 40 of the lines (16.7%) had short vernalization at *vrn-B1* (*Vrn-B1a*, *Vrn-B1_AGS2000* locus) (Guedira et al. 2014).

Summary for marker-trait associations (MTA) identified

Predictability for PH (i.e. the most heritable trait) for the ABLUP dataset was highest for K-PC (0.25) under an rrBLUP model; hence this was regarded as the most reliable in identifying significant MTA. K-Q and K only models, performed similarly with prediction values equal to 0.18 and 0.16 (data not shown). GWAS identified 113 loci significantly associated with the eight measured traits at a threshold of $p < 0.0005$ (Appendix 7).

MTA were detected in all chromosomes except 1D, 3D, 5D, and 6D based on a significance threshold of $p < 0.0005$ (Appendix 7). SNPs associated with multiple traits included: SNP *wsnp_Ex_c12254_19574891* (1A) associated with HD and KNS; *Ppd-D1* 'Norstar' allele (2D) associated with both PH and HD (Table 3; Fig. 2). SNP *wsnp_Ex_c2500_4671165* (3B) associated with PH and KNS; *wsnp_Ex_c13849_21698240* (4B) with GY and KNS, and *wsnp_Ex_c48922_53681502* (4B), associated with GY and KWS.

MTA for GY and yield components

Fifteen markers significant for GY were distributed across eight chromosomes and responsible for 8-28% of the phenotypic variation. Highly significant GY MTA (*wsnp_Ex_c259_497455*; $p = 8.56E-05$) in chromosome 2B showed an overall negative allelic effect (-49.35) under a K-Q model. Using *Rht-B1* and *Vrn-A1* as covariates in a K and K-PC model identified nine SNPs associated with GY in four different datasets. There were 19 markers in 11chromosomes associated with KNS, explaining 6 to 16% of the phenotypic variation. MTA for KWS (19) were distributed across seven chromosomes and responsible for 8 to 26% of the phenotypic variance.

Markers associated with KW (9) which accounted for 10 to 29% of the variation were located in four chromosomes (1A, 2B, 3A, 6A).

MTA for agronomic traits

Fourteen trait-specific MTA for HD were detected in four chromosomes with KASP markers for the alleles of *Ppd-D1* 'Ciano 67' and *Ppd-D1* 'Norstar' being highly significant across four datasets. PH had the largest number of detected MTA (24) which included *Rht-D1* (4D) detected across all BLUP datasets and responsible for 17-34% of variation. *Rht-D1* was highly significant for PH, with p values ranging from 1.90E-08 to 1.80E-05. Spike length had the least number of detected MTA (8), which mapped to chromosomes 1A, 1B, 7B, and 7D. Significant markers for PL (10) were identified in four chromosomes and were responsible for 6-13% of trait variation.

Discussion

Rapid LD decay

Analysis of LD is a prerequisite for evaluating a collection of genotypes and determining adequate marker density for GWAS (Bellucci et al. 2015; ; Chen et al. 2012; Lopes et al. 2015). LD in the AMP was estimated at ~7 cM across the whole genome, with the low proportion of observed marker pairs in complete LD (3.96%) and significant LD (48.71%) leading to this rapid decay. The mean r^2 value for significant marker pairs was 0.32, comparable to a previous study on eastern US soft winter wheat (Cabrera et al. 2014). Other studies have shown LD in winter wheat to decay at distances from 2 to 5 cM (Chen et al. 2012; Hoffstetter et al. 2016; Tadesse et al. 2015) and up to >10 cM distances (Benson et al. 2012; Zhang et al. 2010). Higher LD in the D compared to the A and B genomes was consistent with previous reports (Chao et al. 2010; Sukumaran et al. 2014) and is a possible consequence of recent introgression and bottleneck accompanying the origin of hexaploid wheat (Chao et al. 2010).

The relatively rapid LD decay implies a higher number of markers required for GWAS, which can result in higher mapping resolution (Abdurakhmonov and Abdukarimov 2008). Next-generation sequencing (NGS) platforms such as genotyping by sequencing (GBS) (Elshire et al. 2011; Poland and Rife 2012) could help in generating a larger number of markers amenable to GWAS, particularly for the D genome where marker coverage was low. This low marker coverage in the genome D could also have led to higher mean r^2 values, average distance of pairs in significant LD, and markers in complete LD. Using a two-tailed *t*-test to compare the average r^2 values and cM distance revealed significant differences between values for genome D and genomes A and B ($p < 0.05$). Higher average r^2 value for the D genome, nevertheless, indicates that fewer markers are needed for association mapping (Sukumaran et al. 2014).

Moderate genetic stratification

The presence of population structure (PS) can lead to false positive discoveries in GWAS and thus relationships must be accounted for (Sorrells and Yu 2009; Sukumaran and Yu 2014). Moderate genetic stratification for the AMP was supported by a high within group genetic variance (89.1%) and by the optimum number of clusters produced through STRCUTURE (K=3; Appendix 8). This observation was similar with previous results in spring wheat (Edae et al. 2014) and wheat lines from US and Mexico (Chao et al. 2010) and reflects the impact of selection in maintaining allelic diversity in wheat breeding populations (Edae et al. 2014). The lack of clustering of entries from the same geographic origin within a subpopulation in this study further supported this large within group variation. Subgroup *Q1* was more genetically similar with *Q3*, reflected by a lower *Gst* value between these subgroups (0.13), compared to *Q1* and *Q2* (0.17). One possible explanation for this is the presence of more entries possessing the *Rht-B1b/Rht-D1b* allele combinations in the *Q1* (7) and the *Q3* (17) subgroups, compared to the *Q2*

(2) subgroup. *Q3* was the least differentiated among the subgroups, as reflected by having the lowest value for *Fst*. In contrast with the current observation, higher levels of population structure had been detected in Chinese wheat cultivars (Zhang et al. 2011), US elite winter wheat (Zhang et al. 2010), and CIMMYT elite spring wheat yield trial lines (Dreisigacker et al. 2012).

Genome location of identified MTA compared to previous studies

GY is a complex trait and its improvement is a primary objective for wheat breeding programs (Ain et al. 2015; Green et al. 2012). The distribution of MTA in multiple chromosomes confirms a complex genetic architecture for yield (Quarrie et al. 2005; Shi et al. 2009). In the present study, significant associations identified for GY and yield component MTA in chromosomes 1A, 2A, 2B, 3B, and 5A agreed with previous reports (Addison et al. 2016; Bennett et al. 2012; Bordes et al. 2014; Lopes et al. 2015). Markers in LD in chromosome 4B associated with GY (*wsnp_Ex_c13849_21698240*, *wsnp_Ex_c48922_53681502*, and *wsnp_CAP11_c84_120095*) were mapped in a region flanking the *Rht-B1* locus, which was previously associated with variation for GY in a CIMMYT spring wheat GWAS (Lopes et al. 2015). SNP *wsnp_Ex_c259_497455*, identified in the SBLUP dataset, coincided with a GY QTL mapped between 9 and 12.5 cM in chromosome 2B by Bordes et al. (2014). Additionally, GY-associated markers *wsnp_Ex_c2723_5047696*, mapped in ABLUP, BLUP15, and SBLUP datasets under a K-Q model, together with *wsnp_Ex_rep_c66331_64502363* and *wsnp_Ex_rep_c66331_64502558* co-localized with a QTL previously mapped in chromosome 3BS for yield under irrigated conditions (Bennett et al. 2012). The use of BLUP trait values from combined analyses increased the power in finding significant QTL as BLUPs are robust in identifying significant associations (Mason et al. 2013). Majority of the GY MTA observed in this study showed negative allelic effects with respect to the minor allele, indicating that breeders

have been successful in selecting alleles that improve yield and productivity in modern winter wheat cultivars. Validation of yield QTL in CIMMYT's WAMI panel (Lopes et al. 2015; Sukumaran et al. 2014) also showed that selections were made for the yield "enhancing" major allele (DN Lozada, unpublished data), suggesting that both winter and spring classes have undergone similar selection pressures to achieve optimum yield. Simultaneously capturing these favorable alleles into new germplasm would be beneficial for breeding higher yielding varieties of wheat.

Yield component traits are generally more heritable than GY itself and therefore have potential for genetic improvement. A SNP associated with KNS, *wsnp_Ex_c12254_19574891* (1A), was mapped within a 6 cM distance from marker *wPt6122*, previously associated with grain number and spike number m^{-2} in a winter wheat core collection (Neumann et al. 2011). The same marker was also located proximal to a KNS QTL (within 1 cM) region previously detected by Edae et al. (2014). SNP *wsnp_Ex_c1276_2445537* mapped at 172.32 cM in chromosome 6B coincided with a KWS-associated region reported by Neumann et al. (2011) at 175.9 cM. For KW, *wsnp_JD_c5699_6859527* (3A) co-located with a thousand grain weight "enhancing" locus *BARC0197_174* in a panel of European winter and spring wheat varieties (Zanke et al. 2015). The positive minor allele effect of this marker and its detection in three BLUP datasets (ABLUP, BLUP15, NBLUP) under a K-Q model, indicate that it could be a potential target for improving KW in existing germplasm.

Twenty-four markers distributed across 10 chromosomes were associated with variation in PH. Although influenced by many genes, PH is highly heritable and controlled in large part by *Rht-B1* and *Rht-D1* (Snape et al. 1977; Würschum et al. 2015; Zanke et al. 2014b). *Rht-D1* was highly significant for PH across all BLUP datasets and models used with the dwarfing allele

present in 64% of the lines. The positive allelic effect for this locus indicates that selection by breeders has favored the "height reducing" major allele, as shorter stature has been shown to reduce lodging and increase harvest index (Rebetzke et al. 2011). Despite this, PH was not correlated with GY, in agreement with a previous study by Sukumaran et al. (2014) and in contrast with Bellucci et al. (2015) where negative correlation between these traits was observed. No PH MTA were detected in chromosome 4B harboring the *Rht-B1* gene, consistent with other studies that have shown *Rht-D1* to have a larger genetic effect (Bellucci et al. 2015; Neumann et al. 2011; Würschum et al. 2015; Zanke et al. 2014b). It is also worth noting that PH did not share common significant loci with PL and SL, an unexpected result considering a high correlation observed between these traits and in contrast with previous studies (Heidari et al. 2012; Sukumaran et al. 2014).

The timing of anthesis is a critical trait for adaptation of wheat to diverse environments and is primarily affected by genes for vernalization and photoperiod response (Zanke et al. 2014a). In the present study, MTA for HD were identified in four chromosomes and did not include the *Ppd-B1* region on 2B. This result is likely due to both the stronger effect of the *Ppd-D1a* allele for conferring photoperiod insensitivity (Guedira et al. 2016; Kamran et al. 2014) and its higher frequency within the population (54.8%) compared to *Ppd-B1a* (14.6%) (Online Resource 7). *Ppd-D1* markers for 'Ciano 67' and 'Norstar' alleles were significantly associated with HD across four BLUP datasets and all GWAS models used, similar to previous observations (Zanke et al. 2014a). Major alleles for these loci had negative allelic effects for HD, indicating that insensitivity to photoperiod decreased days to HD, which plays a large role in the adaptation of wheat to the southern US growing areas.

Current and future genetic improvement of southern US winter wheat

The pleiotropic effect of photoperiod insensitivity conferred by *Ppd-D1a* on plant development has previously been shown (Snape et al. 2001; Zanke et al. 2014b) and has its importance for adaptation of southern US winter wheat (Addison et al. 2016; Guedira et al. 2016). In addition to HD, *Ppd-D1* 'Norstar' allele was associated with PH, with a positive minor allele effect indicating selection for reduced PH to improve grain yield. Bentley et al. (2014) and Wilhelm et al. (2013) noted a reduction in PH caused by *Ppd-D1a* among elite European lines and in a worldwide wheat germplasm panel. In this study, 66 of the 100 highest yielding lines possessed the *Ppd-D1b* allele for the *Ppd-D1* 'Norstar' allele, which was higher than expected based on allele frequency (Appendix 9), indicating its importance for yield and adaptation in the current germplasm. The *Rht-D1b* dwarfing allele was also present in 60 of the 100 highest yielding entries. Taken together, our results showed the interplay of reduced PH and photoperiod to produce higher yielding cultivars of soft winter wheat adapted to the southern US.

Several studies have previously reported multi-trait MTA associated with GY, yield components and agronomic traits using a GWAS approach in spring wheat (Edae et al. 2014; Sukumaran et al. 2014). GY shared common MTA (*wsnp_Ex_c13849_21698240* and *wsnp_Ex_c48922_53681502* (4B)) with KNS and KWS (Table 3), which explained 10-26% of trait variation (Table 2). To our knowledge, there has not been a report on multi-trait loci related with controlling variation for GY and yield components mapped in chromosome 4B. Edae et al. (2014) previously identified multi-trait markers associated with GY, spikes m^{-2}, KW, and TW in chromosome 5B while Wang et al. (2009) mapped loci in 1B, 2A, and 3B associated with grain filling rate, KWS, and KW. Our results here thus provide additional multi-trait loci associated with yield and yield components which can be targeted for future MAS to improve GY and

adaptation in soft winter wheat. The multi-trait markers identified in this study could ultimately be used to accelerate pyramiding of yield and adaptation-related QTL to develop southern US winter wheat varieties with increased GY potential and broader adaptations.

Conclusions

A GWAS for GY, yield components, and agronomic traits in soft winter wheat was conducted using genome-wide SNP markers. Multi-trait MTA in chromosomes 1A, 2D, 3B, and 4B were identified that could be potential targets of selection for marker-assisted breeding to capitalize on variation for GY, yield components, and adaptation traits in winter wheat. QTL validation and development of breeder-friendly assays for these multi-trait loci and their deployment to existing breeding programs could ultimately help accelerate MAS to improve GY and adaptation in soft winter wheat. Results from this study serve as valuable resources for molecular breeding towards varietal improvement of wheat. The utility of association mapping approach for determining genomic regions affecting variation for traits of agricultural and economic importance was demonstrated.

References

Abdurakhmonov IY, Abdukarimov A (2008) Application of Association Mapping to Understanding the Genetic Diversity of Plant Germplasm Resources. Int J Plant Genomics 2008:574927. doi:10.1155/2008/574927

Addison CK, Mason RE, Brown-Guedira G, Guedira M, Hao Y, Miller RG, Subramanian N, Lozada DN, Acuna A, Arguello MN, Johnson JW, Ibrahim AMH, Sutton R, Harrison SA (2016) QTL and major genes influencing grain yield potential in soft red winter wheat adapted to the southern United States. Euphytica:1-13. doi:10.1007/s10681-016-1650-1

Ain Q-u, Rasheed A, Anwar A, Mahmood T, Imtiaz M, Mahmood T, Xia X, He Z, Quraishi UM (2015) Genome-wide association for grain yield under rainfed conditions in historical wheat cultivars from Pakistan. Front Plant Sci 6:743. doi:10.3389/fpls.2015.00743

Bellucci A, Torp AM, Bruun S, Magid J, Andersen SB, Rasmussen SK (2015) Association Mapping in Scandinavian Winter Wheat for Yield, Plant Height, and Traits Important for Second-Generation Bioethanol Production. Front Plant Sci 6:1046. doi: 10.3389/fpls.2015.01046

Bennett D, Reynolds M, Mullan D, Izanloo A, Kuchel H, Langridge P, Schnurbusch T (2012) Detection of two major grain yield QTL in bread wheat (Triticum aestivum L.) under heat, drought and high yield potential environments. Theor Appl Genet 125 (7):1473-1485

Benson J, Brown-Guedira G, Paul Murphy J, Sneller C (2012) Population structure, linkage disequilibrium, and genetic diversity in soft winter wheat enriched for fusarium head blight resistance. The Plant Genome 5 (2):71-80

Bentley AR, Scutari M, Gosman N, Faure S, Bedford F, Howell P, Cockram J, Rose GA, Barber T, Irigoyen J, Horsnell R, Pumfrey C, Winnie E, Schacht J, Beauchêne K, Praud S, Greenland A, Balding D, Mackay IJ (2014) Applying association mapping and genomic selection to the dissection of key traits in elite European wheat. Theor Appl Genet 127 (12):2619-2633. doi:10.1007/s00122-014-2403-y

Bordes J, Goudemand E, Duchalais L, Chevarin L, Oury FX, Heumez E, Lapierre A, Perretant MR, Rolland B, Beghin D (2014) Genome-wide association mapping of three important traits using bread wheat elite breeding populations. Mol Breeding 33 (4):755-768

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23 (19):2633-2635. doi:10.1093/bioinformatics/btm308

Breseghello F, Sorrells ME (2006a) Association analysis as a strategy for improvement of quantitative traits in plants. Crop Sci. 46 (3):1323-1330. doi:10.2135/cropsci2005.09-0305

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

Cabrera A, Souza E, Guttieri M, Sturbaum A, Hoffstetter A, Sneller C (2014) Genetic Diversity, Linkage Disequilibrium, and Genome Evolution in Soft Winter Wheat. Crop Sci. 54 (6):2433-2448

Cavanagh C, Chao S, Wang S, Huang B, Stephen S, Kiani S, Forrest K, Saintenacc C, Brown-Guedira G, Akhunova A, et al. (2013) Genome-wide comparative diversity uncovers multiple targets of selection for improvement in hexaploid wheat landraces and cultivars. PNAS 110 (20):8057–8062. doi:www.pnas.org/cgi/doi/10.1073/pnas.1217133110

Chao S, Dubcovsky J, Dvorak J, Luo M-C, Baenziger SP, Matnyazov R, Clark DR, Talbert LE, Anderson JA, Dreisigacker S, Glover K, Chen J, Campbell K, Bruckner PL, Rudd JC, Haley S, Carver BF, Perry S, Sorrells ME, Akhunov ED (2010) Population- and genome-specific patterns of linkage disequilibrium and SNP variation in spring and winter wheat (Triticum aestivum L.). BMC Genomics 11 (1):1-17. doi:10.1186/1471-2164-11-727

Chen X, Min D, Yasir TA, Hu Y-G (2012) Genetic diversity, population structure and linkage disequilibrium in elite Chinese winter wheat investigated with SSR markers. PLOS One 7 (9):e44510

Cleveland WS (1979) Robust locally weighted regression and smoothing scatterplots. J Am Stat Assoc 74 (368):829-836

Dreisigacker S, Shewayrga H, Crossa J, Arief VN, DeLacy IH, Singh RP, Dieters MJ, Braun H-J (2012) Genetic structures of the CIMMYT international yield trial targeted to irrigated environments. Mol Breeding 29 (2):529-541

Earl DA (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conser Genet Resour 4 (2):359-361

Edae EA, Byrne PF, Haley SD, Lopes MS, Reynolds MP (2014) Genome-wide association mapping of yield and yield components of spring wheat under contrasting moisture regimes. Theor Appl Genet 127 (4):791-807

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLOS One 6 (5):e19379

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

Endelman JB, Jannink J-L (2012) Shrinkage estimation of the realized relationship matrix. G3 (Bethesda): 2 (11):1405-1413

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol 14 (8):2611-2620. doi:10.1111/j.1365-294X.2005.02553.x

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 131 (2):479-491

Federer WT, Crossa J (2012) I.4 Screening Experimental Designs for Quantitative Trait Loci, Association Mapping, Genotype-by Environment Interaction, and Other Investigations. Front Physiol 3:156-156. doi:10.3389/fphys.2012.00156

Federer WT, Raghavarao D (1975) Augmented Designs. Biometrics 31 (1):29-35. doi:10.2307/2529707

Federer WT, Reynolds M, Crossa J (2001) Combining results from augmented designs over sites. Agron J 93 (2):389-395

Flint-Garcia SA, Thornsberry JM, Buckler ES (2003) Structure of linkage disequilibrium in plants. Annu Rev Plant Biol 54:357-374. doi:10.1146/annurev.arplant.54.031902.134907

Francis R (2016) Pophelper: an R package and web app to analyse and visualize population structure. Mol Ecol Resour 17: 1 27-32 10.1111/1755-0998.12509

Green AJ, Berger G, Griffey CA, Pitman R, Thomason W, Balota M, Ahmed A (2012) Genetic Yield Improvement in Soft Red Winter Wheat in the Eastern United States from 1919 to 2009. Crop Sci. 52 (5):2097-2108. doi:10.2135/cropsci2012.01.0026

Guedira M, Maloney P, Xiong M, Petersen S, Murphy JP, Marshall D, Johnson J, Harrison S, Brown-Guedira G (2014) Vernalization Duration Requirement in Soft Winter Wheat is Associated with Variation at the Locus. Crop Sci. 54.5 (2014): 1960-1971

Guedira M, Xiong M, Hao YF, Johnson J, Harrison S, Marshall D, Brown-Guedira G (2016) Heading Date QTL in Winter Wheat (*Triticum aestivum*) Coincide with Major Developmental Genes *VERNALIZATION1* and *PHOTOPERIOD1*. PLOS One 11 (5):e0154242. doi:10.1371/journal.pone.0154242

Heidari B, Saeidi G, Sayed Tabatabaei BE, Suenaga K (2012) QTLs Involved in Plant Height, Peduncle Length and Heading Date of Wheat (Triticum aestivum L.). J Agr Sci Technol 14 (5):1090-1104

Hoffstetter A, Cabrera A, Sneller C (2016) Identifying Quantitative Trait Loci for Economic Traits in an Elite Soft Red Winter Wheat Population. Crop Sci 56:1-12

Kamran A, Iqbal M, Spaner D (2014) Flowering time in wheat (Triticum aestivum L.): a key factor for global adaptability. Euphytica 197 (1):1-26. doi:10.1007/s10681-014-1075-7

Kang HM, Zaitlen NA, Wade CM, Kirby A, Heckerman D, Daly MJ, Eskin E (2008) Efficient control of population structure in model organism association mapping. Genetics 178 (3):1709-1723. doi:10.1534/genetics.107.080101

Laido G, Marone D, Russo MA, Colecchia SA, Mastrangelo AM, De Vita P, Papa R (2014) Linkage disequilibrium and genome-wide association mapping in tetraploid wheat (Triticum turgidum L.). PLOS One 9 (4):e95211

Lee S, Abecasis G R, Boehnke, M, & Lin X. (2014). Rare-Variant Association Analysis: Study Designs and Statistical Tests. Am J of Hum Genet 95(1), 5–23. http://doi.org/10.1016/j.ajhg.2014.06.009

Lipka AE, Tian F, Wang Q, Peiffer J, Li M, Bradbury PJ, Gore MA, Buckler ES, Zhang Z (2012) GAPIT: genome association and prediction integrated tool. Bioinformatics 28 (18):2397-2399. doi:10.1093/bioinformatics/bts444

Lopes MS, Dreisigacker S, Pena RJ, Sukumaran S, Reynolds MP (2015) Genetic characterization of the wheat association mapping initiative (WAMI) panel for dissection of complex traits in spring wheat. Theor Appl Genet 128 (3):453-464. doi:10.1007/s00122-014-2444-2

Mason RE, Hays DB, Mondal S, Ibrahim AMH, Basnet BR (2013) QTL for yield, yield components and canopy temperature depression in wheat under late sown field conditions. Euphytica 194 (2):243-259. doi:10.1007/s10681-013-0951-x

Meirmans PG, Van Tienderen PH (2004) GENOTYPE and GENODIVE: two programs for the analysis of genetic diversity of asexual organisms. Mol Ecol Notes 4 (4):792-794

Moose SP, Mumm RH (2008) Molecular Plant Breeding as the Foundation for 21st Century Crop Improvement. Plant Physiol 147 (3):969-977. doi:10.1104/pp.108.118232

Myles S, Peiffer J, Brown PJ, Ersoz ES, Zhang Z, Costich DE, Buckler ES (2009) Association Mapping: Critical Considerations Shift from Genotyping to Experimental Design. The Plant Cell 21 (8):2194-2202. doi:10.1105/tpc.109.068437

Nei M (1987) Molecular evolutionary genetics. Columbia university press, 512 pp.

Neumann K, Kobiljski B, Denčić S, Varshney RK, Börner A (2011) Genome-wide association mapping: a case study in bread wheat (Triticum aestivum L.). Mol Breeding 27 (1):37-58. doi:10.1007/s11032-010-9411-7

Nielsen NH, Backes G, Stougaard J, Andersen SU, Jahoor A (2014) Genetic Diversity and Population Structure Analysis of European Hexaploid Bread Wheat (*Triticum aestivum* L.) Varieties. PLOS One 9 (4):e94000. doi:10.1371/journal.pone.0094000

Pallotta M, Warner P, Fox R, Kuchel H, Jefferies S, Langridge P Marker assisted wheat breeding in the southern region of Australia. In: Proc. 10th Int. Wheat Genet. Symp., Paestum, Italy, 2003. pp 1-6

Patel D, Zander M, Dalton-Morgan J, Batley J (2015) Advances in Plant Genotyping: Where the Future Will Take Us. In: Batley J (ed) Plant Genotyping, vol 1245. Methods Mol Biol. Springer New York, pp 1-11. doi:10.1007/978-1-4939-1966-6_1

Poland JA, Rife TW (2012) Genotyping-by-sequencing for plant breeding and genetics. The Plant Genome 5 (3):92-102

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155 (2):945-959

Quarrie SA, Steed A, Calestani C, Semikhodskii A, Lebreton C, Chinoy C, Steele N, Pljevljakusić D, Waterman E, Weyen J, Schondelmaier J, Habash DZ, Farmer P, Saker L, Clarkson DT, Abugalieva A, Yessimbekova M, Turuspekov Y, Abugalieva S, Tuberosa R, Sanguineti M-C, Hollington PA, Aragués R, Royo A, Dodig D (2005) A high-density genetic map of hexaploid wheat (Triticum aestivum L.) from the cross Chinese Spring × SQ1 and its use to compare QTLs for grain yield across a range of environments. Theor Appl Genet 110 (5):865-880. doi:10.1007/s00122-004-1902-7

R Development Core Team (2010) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/

Rebetzke GJ, Ellis MH, Bonnett DG, Condon AG, Falk D, Richards RA (2011) The Rht13 dwarfing gene reduces peduncle length and plant height to increase grain number and yield of wheat. Field Crops Res 124 (3):323-331. doi:http://dx.doi.org/10.1016/j.fcr.2011.06.022

Ruggieri V, Francese G, Sacco A, D'Alessandro A, Rigano MM, Parisi M, Milone M, Cardi T, Mennella G, Barone A (2014) An association mapping approach to identify favourable alleles for tomato fruit quality breeding. BMC Plant Biol 14. doi:10.1186/s12870-014-0337-9

SAS Institute (2011) SAS System Options: Reference, 2nd ed. SAS Institute, Cary, NC

Sehgal D, Autrique E, Singh R, Ellis M, Singh S, Dreisigacker S (2017) Identification of genomic regions for grain yield and yield stability and their epistatic interactions. 7:41578. doi:10.1038/srep41578 Sci Rep https://www.nature.com/articles/srep41578

Shi J, Li R, Qiu D, Jiang C, Long Y, Morgan C, Bancroft I, Zhao J, Meng J (2009) Unraveling the Complex Trait of Crop Yield With Quantitative Trait Loci Mapping in Brassica napus. Genetics 182 (3):851-861. doi:10.1534/genetics.109.101642.

Snape JW, Butterworth K, Whitechurch E, Worland AJ (2001) Waiting for fine times: genetics of flowering time in wheat. Euphytica 119 (1-2):185-190. doi:10.1023/a:1017594422176

Snape JW, Law CN, Worland AJ (1977) Whole chromosome analysis of height in wheat. Heredity 38 (1):25-36

Sorrells ME, Yu J (2009) Linkage disequilibrium and association mapping in the Triticeae. In: Genetics and genomics of the Triticeae. Springer, pp 655-683

Sukumaran S, Dreisigacker S, Lopes M, Chavez P, Reynolds M (2014) Genome-wide association study for grain yield and related traits in an elite spring wheat population grown in temperate irrigated environments. Theor Appl Genet:1-11. doi:10.1007/s00122-014-2435-3

Sukumaran S, Yu J (2014) Association Mapping of Genetic Resources: Achievements and Future Perspectives. In: Tuberosa R, Graner A, Frison E (eds) Genomics of Plant Genetic Resources: Volume 1. Managing, sequencing and mining genetic resources. Springer Netherlands, Dordrecht, pp 207-235. doi:10.1007/978-94-007-7572-5_9

Tadesse W, Ogbonnaya FC, Jighly A, Sanchez-Garcia M, Sohail Q, Rajaram S, Baum M (2015) Genome-Wide Association Mapping of Yield and Grain Quality Traits in Winter Wheat Genotypes. PLOS One 10 (10). doi:10.1371/journal.pone.0141339

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

Turner SD (2014) qqman: an R package for visualizing GWAS results using QQ and manhattan plots. bioRxiv:005165

Wang G, Leonard JM, von Zitzewitz J, Peterson CJ, Ross AS, Riera-Lizarazu O (2014a) Marker-trait association analysis of kernel hardness and related agronomic traits in a core collection of wheat lines. Mol Breeding 34 (1):177-184. doi:10.1007/s11032-014-0028-0

Wang RX, Hai L, Zhang XY, You GX, Yan CS, Xiao SH (2009) QTL mapping for grain filling rate and yield-related traits in RILs of the Chinese winter wheat population Heshangmai × Yu8679. Theor Appl Genet 118 (2):313-325. doi:10.1007/s00122-008-0901-5

Wang S, Wong D, Forrest K, Allen A, Chao S, Huang BE, Maccaferri M, Salvi S, Milner SG, et al. (2014b) Characterization of polyploid wheat genomic diversity using a high-density 90 000 single nucleotide polymorphism array. Plant Biotechnol J 12 (6):787-796. doi:10.1111/pbi.12183

Weir BS, Cockerham C (1996) Genetic data analysis II: Methods for discrete population genetic data. Sinauer Assoc. Inc, Sunderland, MA, USA

Wilhelm EP, Boulton MI, Al-Kaff N, Balfourier F, Bordes J, Greenland AJ, Powell W, Mackay IJ (2013) Rht-1 and Ppd-D1 associations with height, GA sensitivity, and days to heading in a worldwide bread wheat collection. Theor Appl Genet 126 (9):2233-2243. doi:10.1007/s00122-013-2130-9

Wolfinger R, Federer WT, Cordero-Brana O (1997) Recovering information in augmented designs, using SAS PROC GLM and PROC MIXED. Agron J 89 (6):856-859

Würschum T, Langer SM, Longin CFH (2015) Genetic control of plant height in European winter wheat cultivars. Theor Appl Genet 128 (5):865-874. doi:10.1007/s00122-015-2476-2

Zanke C, Ling J, Plieske J, Kollers S, Ebmeyer E, Korzun V, Argillier O, Stiewe G, Hinze M, Beier S, Ganal MW, Röder MS (2014a) Genetic architecture of main effect QTL for heading date in European winter wheat. Front Plant Sci 5:217. doi:10.3389/fpls.2014.00217

Zanke CD, Ling J, Plieske J, Kollers S, Ebmeyer E, Korzun V, Argillier O, Stiewe G, Hinze M, Neumann F, Eichhorn A, Polley A, Jaenecke C, Ganal MW, Röder MS (2015) Analysis of main effect QTL for thousand grain weight in European winter wheat (Triticum aestivum L.) by genome-wide association mapping. Front Plant Sci 6:644. doi:10.3389/fpls.2015.00644

Zanke CD, Ling J, Plieske J, Kollers S, Ebmeyer E, Korzun V, Argillier O, Stiewe G, Hinze M, Neumann K (2014b) Whole genome association mapping of plant height in winter wheat (Triticum aestivum L.). PLOS One 9 (11):e113287

Zhang D, Bai G, Zhu C, Yu J, Carver BF (2010) Genetic diversity, population structure, and linkage disequilibrium in US elite winter wheat. The Plant Genome 3 (2):117-127

Zhang L, Liu D, Guo X, Yang W, Sun J, Wang D, Sourdille P, Zhang A (2011) Investigation of genetic diversity and population structure of common wheat cultivars in northern China using DArT markers. BMC Genetics 12 (1):42

Zhao K, Aranzana MJ, Kim S, Lister C, Shindo C, Tang C, Toomajian C, Zheng H, Dean C, Marjoram P, Nordborg M (2007) An Arabidopsis example of association mapping in structured samples. PLOS Genetics 3 (1). doi:10.1371/journal.pgen.0030004

Zhao K, Tung C-W, Eizenga GC, Wright MH, Ali ML, Price AH, Norton GJ, Islam MR, Reynolds A, Mezey J, McClung AM, Bustamante CD, McCouch SR (2011) Genome-wide association mapping reveals a rich genetic architecture of complex traits in Oryza sativa. Nat Commun 2:467. doi:http://www.nature.com/ncomms/journal/v2/n9/suppinfo/ncomms1467_S1.html

Zhu C, Gore M, Buckler ES, Yu J (2008) Status and Prospects of Association Mapping in Plants. Plant Genome 1 (1):5-20. doi:10.3835/plantgenome2008.02.00

Table 1. Adjusted means and analysis of variance (ANOVA) of the measured traits for the soft winter wheat association mapping panel

Trait	Mean	Range	h^2 [a] (Adj.)	ANOVA (Mean Squares) Block (B)	Env. (E)	B(E)[b]	Genotype (G)	GEI [c]
Grain yield (kg ha^{-1})	3172	604-7184	0.48	NS	$1.42x10^7$***	NS	$1.37x10^6$***	710921***
Heading date (days)	116.3	57-136	0.63	NS	535.67***	NS	30.80***	11.37***
Kernel number $spike^{-1}$	27.4	9.5-43.9	0.37	NS	933.04	NS	30.46***	19.27***
Kernel weight (mg/kernel)	33.2	11.4-51.0	0.77	NS	316.33***	NS	51.54***	11.70***
Kernel weight $spike^{-1}$ (mg/spike)	0.85	0.3-1.9	0.47	NS	0.85***	NS	0.08***	0.043***
Plant height (cm)	80.0	62.4-106.1	0.81	NS	3940.59***	NS	309.75***	58.82***
Peduncle length (cm)	34.1	24.4-47.2	0.33	NS	3183.30***	NS	74.11***	49.52***
Spike length (cm)	18.0	12.8-27.8	0.30	NS	394.34***	NS	19.32***	13.65***

[a] Broad sense heritability estimates for adjusted means; calculated as $H^2 = \frac{\sigma^2_G}{\sigma^2_G + \frac{\sigma^2_{GEI}}{e} + \frac{\sigma^2_E}{er}}$

[b] Incomplete blocks
[c] Blocks nested within environments
[d] GEI genotype by environment interaction
***Significant at $p < 0.0001$ level

Table 2. Phenotypic correlations (*r*) of the measured traits for the soft winter wheat association mapping panel

Trait	GY	HD	KNS	KW	KWS	PH	PL	SL
GY	-							
HD	0.10	-						
KNS	0.67***	0.13*	-					
KW	0.48***	-0.07	0.26***	-				
KWS	0.73***	0.06	0.87***	0.70***	-			
PH	-0.07	0.19*	0.02	0.18*	0.10	-		
PL	-0.15*	-0.04	-0.10	0.06	-0.07	0.49***	-	
SL	0.01	0.09	0.08	0.062	0.08	0.12	0.11	-

GY grain yield, *HD* heading date, *KNS* kernel number spike^{-1}, *KW* kernel weight, *KWS* kernel weight spike^{-1}, *PH* plant height, *PL* peduncle length, *SL* spike length

*Correlation is significant at $p < 0.05$ level

** Correlation is significant at $p < 0.001$ level

***Correlation is significant at $p < 0.0001$ level

Table 3. Markers associated with multiple traits identified for yield and agronomic traits for the soft winter wheat association mapping panel

Marker	Model	Traits	Dataset [a]	Chr	Position (cM)	*p* value	Allelic effects [b]	R^2 [c]
wsnp_Ex_c12254_19574891	*K-Q*	HD	NBLUP	1A	12.43	0.00027	2.136	0.08
	K-Q	KNS	NBLUP	1A	12.43	0.00012	-7.331	0.09
Ppd-D1[d]	*K-Q*	HD	BLUP14	2D		4.86×10^{-5}	0.314	0.13
	K-Q	PH	SBLUP	2D		0.000375	1.067	0.28
wsnp_Ex_c2500_4671165	*K-PC*	PH	ABLUP	3B	263.71	7.10×10^{-5}	0.113	0.18
	K	KNS	NBLUP	3B	263.71	0.000433	-6.877	0.06
wsnp_Ex_c13849_21698240	*K-PC-Rht-Vrn*	GY	BLUP15	4B	85.15	2.03×10^{-5}	128.3	0.26
	K-Q	KNS	NBLUP	4B	85.15	3.33×10^{-5}	-2.034	0.10
wsnp_Ex_c48922_53681502	*K-PC*	GY	SBLUP	4B	100.86	0.000288	-43.347	0.27
	K-PC	KWS	BLUP15	4B	100.86	0.000270	-0.012	0.24

GY grain yield*;* *HD* Heading date; *KNS* kernel number spike^{-1}; *KWS* kernel weight spike^{-1}; *PH* plant height

[a] Phenotypic dataset generated from combining adjusted means from BLUP across all environments (ABLUP); across the northern locations (NBLUP); southern locations (SBLUP); 2014 site-years (BLUP14); and 2015 site-years (BLUP15)

[b] Allelic effects with respect to the minor allele

[c] Reflect the phenotypic variation explained by the marker, R^2 of the model with SNP calculated in GAPIT package in R

[d] *Ppd-D1* ‘Norstar’ allele

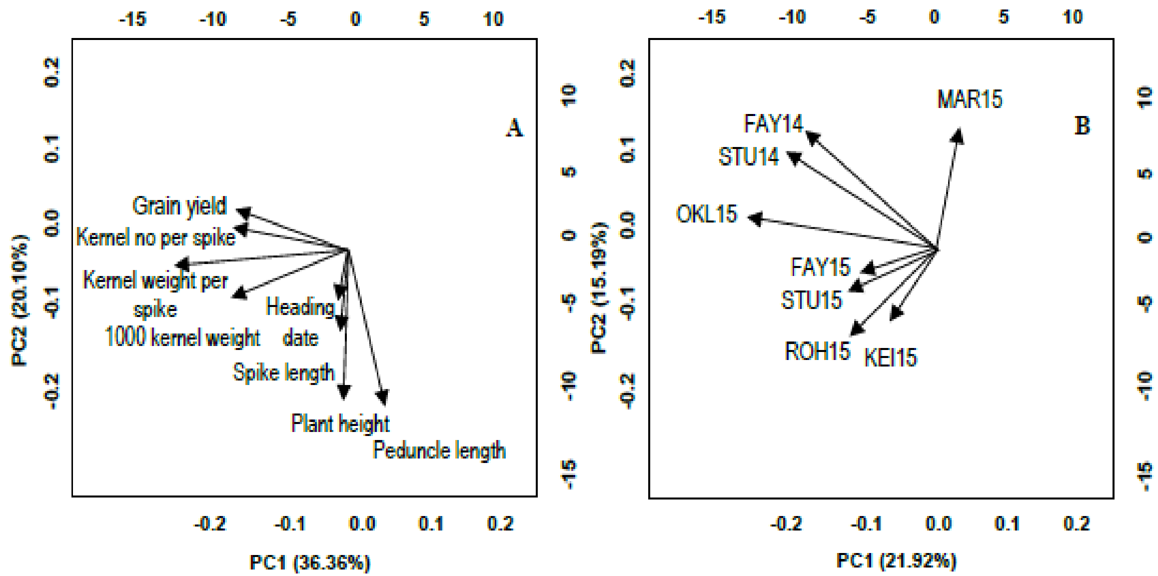


Figure 1. PCA biplots for the (A) measured traits and (B) adjusted grain yield across different site-years for the soft winter wheat AMP. Site-years: *FAY14*- Fayetteville14; *FAY15*- Fayetteville15; *KEI15*- Keiser15; *MAR15*- Marianna15; *OKL15*- Oklahoma15; *ROH15*- Rohwer15; *STU14*- Stuttgart14; *STU15*- Stuttgart15

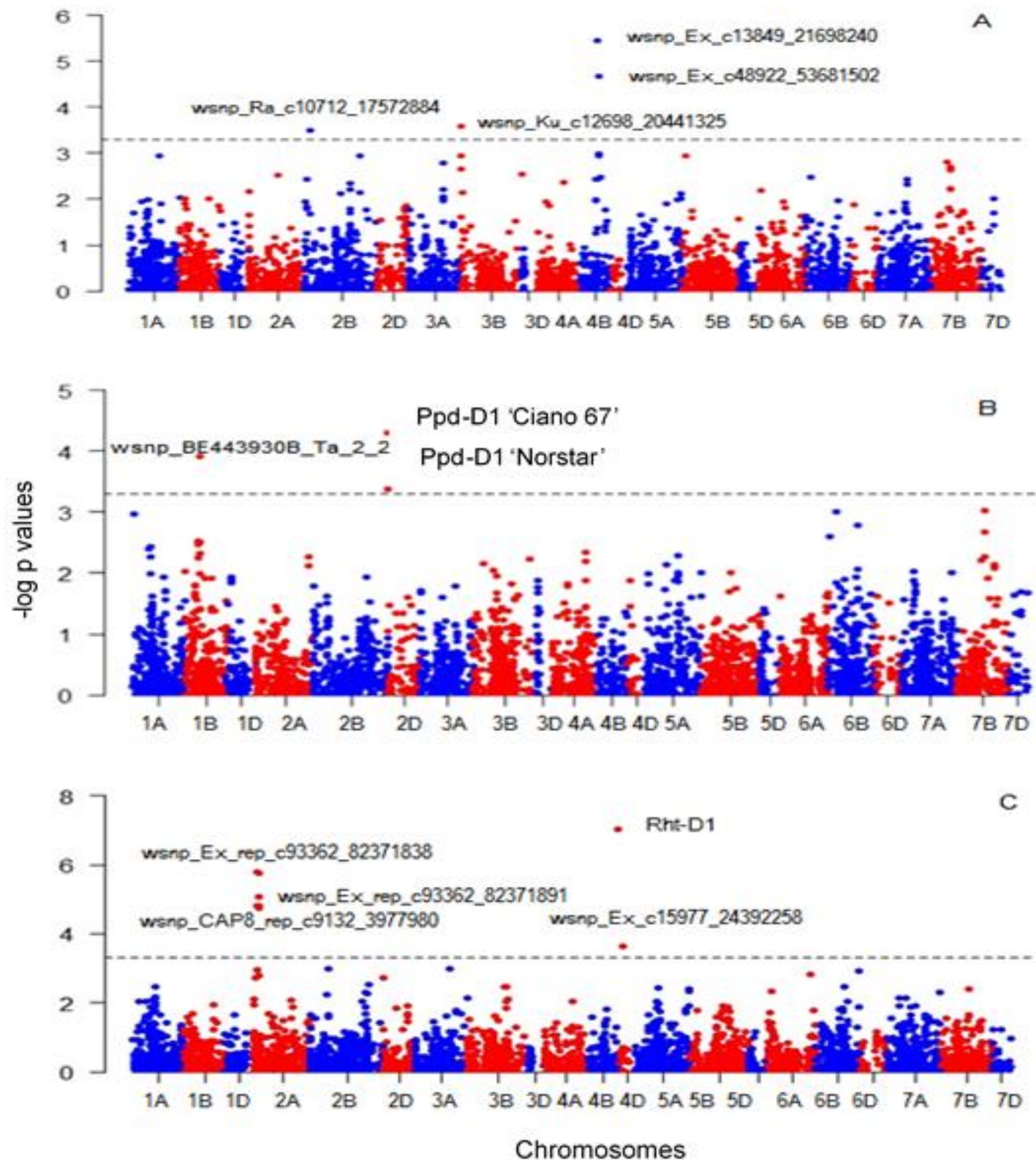


Figure 2. Manhattan plot showing genome-wide SNP loci associated with grain yield, heading date, and plant height. Horizontal line represents the significant threshold by which markers were considered associated with a trait ($p <$ 0.0005; ~3.30); (A) Plot of genome-wide markers associated with GY under a *K-PC* model observed using the BLUP values for the 2015 site-years (B) Plot of genome-wide markers associated with HD under a *K-Q* model observed using BLUP values across all environments (ABLUP); (C) Plot of genome-wide SNPs associated with PH observed using the BLUP trait values for northern environments (NBLUP), *K-Q* model

CHAPTER III

VALIDATION OF YIELD QTL FROM SOFT WINTER WHEAT USING A CIMMYT SPRING WHEAT PANEL

Abstract

Validation of quantitative trait loci (QTL) is an essential step in marker-assisted breeding. The objective of this study was to validate grain yield (GY) QTL previously identified in soft red winter wheat (SRWW) through biparental and association mapping using the spring wheat association mapping initiative (WAMI) panel from International Wheat and Maize Improvement Center (CIMMYT), Mexico. Linked SNP for *IWA3560* (3A), *IWA1818* (4B), and *IWA755* (6B) were significantly associated ($p < 0.001$) with GY, grain number, and thousand grain weight in the WAMI. Lines possessing the favorable allele for the QTL at the 3A, 4B, and 6B loci (A-C-G allele combination) validated on the WAMI had the highest mean GY at 4.55 t ha^{-1}. Predicted candidate gene functions for the validated loci at chromosomes 3A, 4B, and 6B included repressor of RNA pol III transcription, regulation of ubiquitin activity, and a transcription factor. BLAST analyses against a recently developed exome capture platform revealed that putative single nucleotide mutations at the hit region for the 3A and 6B loci could result either to missense or no amino changes (synonymous) for the corresponding proteins. These results validate GY QTL across winter and spring wheat through genome-wide association analysis and demonstrate the potential for pyramiding favorable alleles for the genetic improvement of wheat breeding populations.

Introduction

It is necessary to validate quantitative trait loci (QTL) across different genetic backgrounds for more efficient implementation of marker-assisted selection (MAS) (Dao et al. 2017; Sallam et al. 2016; Su et al. 2016). This is particularly true of QTL for grain yield (GY), as it is a quantitative trait influenced by many loci with mostly small effects, making the identification and validation of significant marker-trait associations (MTA) a challenge. While improvement of GY has remained the top priority of wheat breeding programs (Green et al. 2012), validation and utilization of GY QTL lags the progress seen in more qualitatively inherited traits such as resistance to biotic stresses (Bokore et al. 2017; Petersen et al. 2017; Prat et al. 2017).

Despite the complexity, some studies have reported QTL validation and marker development for quantitative traits. Wang et al. (2016) recently validated a novel low-tillering QTL, *Qltn.sicau-2D*, using multiple recombinant inbred line (RIL) populations derived from the genotype 'H461'. QTL for quality traits previously identified in a double haploid (DH) population were validated using a recombinant inbred and $BC_3F_{2:3}$ mapping panel, with markers *Bx7-MAR* (1B), *Xwmc182a* (6B), and *Xwmc182b* (7B) recommended for further investigation (Dao et al. 2017). Loci controlling variation for thousand kernel weight (TGW) were also identified and two breeder-friendly Kompetitive allele specific primers (KASP) were developed for MAS of a major QTL in 7AL (Su et al. 2016). Rasheed et al (2016) validated markers for genes controlling TGW in chromosomes 2B (*TaSus2-2B*), 3A (*TaTGW6-3A*), 3D (*TaCKX-D1*), 5D (*TaCwi-5D*), 7D (*TaGS-D1*), and spike number in 7A (*TaMoc1-7A*) in a diverse population of wheat from China and 13 other countries. Even with these findings, QTL validation for GY and yield components remains limited and there are no reports of stable QTL across both spring and winter wheat.

Based on previous meta-analyses showing co-localization of QTL detected across wheat classes (Acuña-Galindo et al., 2015, Zhang et al., 2010), it was hypothesized that stable GY related QTL could be identified across winter and spring wheat. For the current study, loci previously reported by Addison et al. (2016) and Lozada et al. (2017; published, Chapter II) for GY and yield components in soft red winter wheat (SRWW) were selected for validation in spring wheat. The specific objectives were to 1) develop KASP markers for selected QTL and validate these loci in spring wheat through genome-wide association mapping; 2) identify allelic haplotypes resulting in the highest GY, and 3) determine candidate genes and effects of single nucleotide mutations at the validated SRWW QTL based on the flanking sequences of single nucleotide polymorphism (SNP). Results are anticipated to facilitate GY improvement in wheat by identifying target loci for MAS and future gene cloning.

Materials and Methods

Wheat association mapping initiative (WAMI)

The WAMI panel consists of 287 advanced lines of spring wheat with a narrow range of variation for phenology and plant height from different elite spring wheat yield trial (ESWYT) nurseries of CIMMYT, Mexico (Lopes et al. 2012; Lopes et al. 2015). The population has been characterized for GY, yield components, and physiological traits (Edae et al. 2014; Sukumaran et al. 2015a), earliness per se (Sukumaran et al. 2016), adaptation to plant density (Sukumaran et al. 2015b), and spike ehtylene production (Valluru et. al 2017) under different high yield potential and stressed environments across mutiple international testing locations. Population structure for the WAMI is primarily based on the presence or absence of the 1B.1R wheat-rye chromosome translocation (Lopes et. al 2015).

Selection of QTL for validation, primer design, and KASP genotyping

SNPs associated with GY, grain number (GNO) and TGW earlier identified in a SRWW panel (winter wheat association mapping panel, AMP; Lozada et al. 2017) were selected for marker design and validation. Markers were chosen based on parameters that included significant associations across multiple phenotypic datasets, *p*-values, allele effects, and association with GY and component traits. Phenotypic datasets in the SRWW panel were derived from combining environments across years (2014 and 2015) and geographic locations (North and South) and calculating BLUPs from these combinations (Lozada et al. 2017; published, Chapter II). From 64 SNPs associated with GY and yield components at $p < 0.0005$, six SNP markers from five chromosomes were selected and converted to KASP assays for validation on the WAMI panel. Five additional KASP assays of SNPs from chromosomes 1A, 3A, and 6B previously reported in association with GY QTL in a bi-parental soft winter wheat population derived from the cross 'Pioneer Brand 26R61 × 'AGS 2000' (Addison et al. 2016; named PA hereafter) were also selected, resulting to a total of 11 winter wheat GY related QTL for validation (Table 1).

Marker design was conducted using the Polymarker primer design pipeline (http://polymarker.tgac.ac.uk/) (Ramirez-Gonzalez et al. 2015) from a database of pre-designed primer sequences derived from the iSelect 90K SNP chip (Wang et al. 2014) and BatchPrimer3 (http://batchprimer3.bioinformatics.ucdavis.edu/) (You et al. 2008). Marker genotyping was done using the KASP assay genotyping protocol described in the CIMMYT Wheat Molecular Genetics Laboratory manual (Dreisigacker et al. 2016), with minor modifications, under a 5µL reaction volume. KASP assays were run under the following thermal conditions: 94°C for 15 mins (hot start enzyme activation); 94°C for 30 sec, 65°C for 1 min (touchdown over 65-57°C for 60 sec, 10 cycles (dropping 0.8°C per cycle), and 72°C for 30 sec (11 cycles); 94°C for 30

sec, 57°C for 1 min, and 72°C for 30 sec (26 cycles); 72°C for 5 mins and 20°C (final). Plates were read on a BMG Pherastar Plus® (Ortenberg, Germany) fluorescent plate reader. Analysis of marker data was carried out using the Klustercaller® software (LGC Genomics, UK). Plates were read on a BMG Pherastar Plus® (Ortenberg, Germany) fluorescent plate reader. Analysis of marker data was carried out using the Klustercaller® software (LGC Genomics, UK).

WAMI phenotypic data analyses

Data for GY, GNO, and TGW were collected from 29 international locations distributed across Asia (Bangladesh, India, Iran, Nepal, Pakistan), Africa (Egypt, Sudan), and North America (Mexico) (Lopes et al. 2012; Sukumaran et al. 2015b) (Appendix 11). Heritability values (H^2) for GY, GNO, and TGW for each environment were calculated using the multi-environment trial analysis in R (META-R) v.6.0 (Alvarado et al. 2016) for an alpha lattice design. Broad sense heritability for individual environments was calculated with the formula:

$$H^2 = \sigma^2_G / \sigma^2_G + \sigma^2_{E/r}$$

where σ^2_G and σ^2_E are variances due to genotype and environment, respectively; and *r* is the number of replications (Sukumaran et. al 2018). Best linear unbiased predictions (BLUP) for each line of the WAMI panel were calculated through combined analyses for all locations and for the environments with H^2 values > 0.50 using the MIXED procedure in SAS v 9.4 (SAS Institute, Cary, NC), where genotypes were considered random. Principal components analysis (PCA) bi-plots were visualized using the META-R program. Mean GY of lines from WAMI having different allele combinations (haplotypes) of the SRWW QTL were compared using a *t*-test with least significant difference (LSD) in a PROC GLM procedure in SAS v 9.4. Pearson correlation coefficient, *r* between measured traits were calculated using PROC CORR in SAS v.

9.4. Phenotypic data for each environment and trait, as well as the genotype data, is available from the link http://hdl.handle.net/11529/10714.

Genome-wide association study and candidate gene analysis

The Illumina® 90K SNP array (Wang et al. 2014) was used for genotyping the WAMI panel. After filtering for a minor allele frequency (MAF) cut-off of 5% and quality control, 26,814 high-quality SNP markers remained for genome-wide association study (GWAS). Out of the 11 markers designed for QTL validation, seven that were segregating in the WAMI were included in this dataset for association mapping. The remaining four markers were excluded from further analysis. GWAS was conducted using the Genome Association Prediction Integrated Tool (GAPIT) package (Lipka et al. 2012) in RStudio®, where the first three principal components (PC) were included for analysis (K-PC model). This model was selected as it was previously identified to be the most reliable in identifying significant MTA in a SRWW mapping panel (Lozada et. al 2017).

To test the effects of the SRWW QTL across single and multiple environments for the WAMI, different best linear unbiased prediction (BLUP) and best linear unbiased estimate (BLUE) datasets were used. Phenotypic datasets for association mapping included BLUP trait values across geographic regions (Asia, Africa, North America; and by individual countries) and by growing season (2010 and 2011) and BLUE calculated for each individual environment. A value of $p < 0.001$ was used as the threshold for defining significant SNPs as the measured traits have generally low to moderate heritability (Arguello et al. 2016; Mwadzingeni et al. 2017). Polymorphic SNPs for the winter wheat loci were separated based on allele calls and the mean GY for these were compared using *t*-test at $p < 0.05$. Results from association mapping for the validated winter wheat QTL were confirmed using generalized linear and mixed models in

TASSEL v.4.0 (Bradbury et al. 2007) and single marker ANOVA in SAS v. 9.4. Manhattan plots were visualized using the 'qqman' package in RStudio (Turner 2014).

Sequences of validated SNPs mapped were BLASTn searched in EnsemblPlants (http://plants.ensembl.org/index.html) against the genomes of the wild diploid D genome ancestor *Aegilops tauschii*, the A genome progenitor *Triticum urartu,* hexaploid wheat (*T. aestivum*), and other plant genomes including *Arabidopsis thaliana, Brachypodium distachyon, Hordeum vulgare,* and *Oryza sativa* (*indica* and *japonica*) to identify putative candidate genes or proteins for these loci. In addition, these sequences were also searched against a recently developed exome capture database derived from sequencing a population of tetraploid ('Kronos'; *T. durum*) and a hexaploid ('Cadenza'; *T. aestivum*) wheat mutants (Krasileva et al. 2017) to identify putative effects of single nucleotide mutations on these regions.

Results

Heritability across environments and trait correlations

A total of 15 environments for GY, 17 for GNO, and 21 for TGW had $H^2 > 0.50$ (Appendix 10). Heritability for these locations ranged between 0.55 and 0.91 (for GY), 0.52 and 0.86 (GNO), and 0.52 and 0.96 (TGW). Environment IH11 (Dharwad, India; 2011 season) had the highest heritability for GY (0.91) and GNO (0.86) while MI10 (Obregon, Mexico; Irrigated; 2011) had the highest H^2 for TGW (0.96). GNO was positively correlated with GY (r= 0.54, $p < 0.0001$), and was in negative correlation with TGW (r=-0.66, $p < 0.0001$). GY also was positively correlated with TGW (r=0.27, $p < 0.0001$) (Appendix 11).

Principal components analysis (PCA) for measured traits across environments

PCA bi-plots for GY, GNO, and TGW across environments (with $H^2 > 0.50$) revealed clustering of locations into different groups (Appendix 12). For GY, PC1 and PC2 explained 44.1 and

18.5% of the variation, respectively, with the African (EE10 and SD10) and Mexican (MD10 and MI10) environments clustering together. Locations from India (ID10, ID11, IL10, IL11, IV10) grouped with environments from Bangladesh (BJ11) and Nepal (NB11). For GNO, PC1 and PC2 explained 57.3 and 25.0% of the variation among environments, respectively. IH11, the environment with the highest heritability for GNO, did not cluster with any other locations. Environments from Bangladesh (BJ11), Nepal (NB10, NB11), and India (ID10, ID11, IL10, IL11, IV10) formed a group while locations from Mexico (MD10, MH10, MHD10, MI10) and Africa (EE10, SH10, SW10) clustered together. For TGW, PC1 and PC2 explained 68.7 and 9.1% of the variation, respectively with two environments from India (IH11 and IL11) grouping together. As with GY and GNO, other environments from Asia including the ones from Bangladesh, Nepal, Pakistan, and rest of India locations also formed a cluster while African locations (EE10, SW10, SH10) grouped with Mexican environments (MD10, MH10, MHD10, MI10).

KASP analysis and association mapping

Fig. 1 illustrates the comparisons for mean GY across seven SRWW SNP loci that were segregating in the WAMI panel. The respective major allele was positive for GY for three out of the seven winter wheat QTL, whereas significant differences among means were observed only at the 4B (*IWA1818*; additive effect: 0.09 t ha^{-1}) and the 6B (*IWA755*; 0.08 t ha^{-1}) loci. The minor allele was favorable of GY for two winter wheat QTL, where only the *IWA3560* (3A) showed significant differences among GY means.

GWAS further identified these three winter wheat QTL to be associated ($p < 0.001$) with GY and component traits across different BLUP and BLUE datasets in the WAMI (Table 2; Fig. 2), confirming the results from the mean comparisons. KASP genotyping for these SNP markers

showed *IWA3560* to produce three clusters (i.e. with homozygous and heterozygous calls) while both the *IWA1818* and *IWA755* produced two groups (no heterozygous genotypes) (Fig. 3).

IWA1818 (4B) was significantly associated with GY and GNO for the ABLUP, BLUPIND, and BLUPH10 datasets (see Appendix 13 for the full description of the datasets used for GWAS). The marker-trait association was responsible for 7-11% of trait variation and showed negative minor allele effects for GNO (-420.74 and -316.68) and GY (-0.10 and 0.06 t ha^{-1}). Marker *IWA3560* (3A; A/G; favorable allele, 'A') was associated with GNO in seven datasets, including MI10 BLUE and MHD10 BLUE for GNO and a single dataset for TGW (AFRBLUPH). This locus was responsible for 4-15% of phenotypic variation for GNO and for 21% of variation for TGW. Marker *IWA755* (6B; A/G; favorable allele, 'G') was associated with GY and GNO in four datasets (ABLUP, ABLUPH, ASIABLUPH, and BLUPH10) explaining 5-12% of the variability for these traits. Consistent allele effects for the 3A, 4B, and 6B loci were observed even for the non-significant environments (data not shown).

The association of *IWA1818* and *IWA755* with GY and GNO were also confirmed using generalized linear and mixed models in TASSEL 4.0 (data not shown). Likewise, single marker ANOVA under a GLM demonstrated the association of these loci with GY ($p < 0.05$, data not shown). Loci other than these winter wheat QTL associated with GY and component traits for the WAMI have been reported recently (Sukumaran et al., 2018).

Haplotype analysis of validated winter wheat SNPs

For GY, the 'A', 'C', and 'G' alleles for the 3A, 4B, and 6B loci, respectively were favorable for higher yield. 'G' was the major and favorable allele for the 3A and 6B loci, with a frequency of 68.4 (197 lines) and 66.7% (192 lines). 'C' was the major allele for the 4B locus (91.0%; 262 lines). The minor allele 'A' had a positive additive effect for *IWA3560* and was present in 56 lines

(19.4%). The A-C-G allele combination for the 3A, 4B, and 6B loci, respectively was present in 44 lines and resulted in the highest mean GY (4.55 t ha^{-1}) across all locations, significantly higher ($p < 0.05$) with haplotypes G-T-A (4.25 t $ha-^{1}$; 5 lines) and G-T-G (3.76 t $ha-^{1}$; 7 lines) (Fig. 4). Likewise, the A-C-G haplotype also had the highest mean GY (4.74 t ha^{-1}) across environments with $H^2 > 0.70$ (data not shown). No significant differences among lines with A-C-G were found with those having the allele haplotypes G-C-G (4.46 t ha^{-1}; 137 lines), A-C-A (4.36 t ha^{-1}; 6 lines) and G-C-A (4.29 t ha^{-1}; 33 lines). The A-T-A haplotype, with GY at 3.98 t ha^{-1}, was present in only one line. No entries possessed the A-T-G allele combination.

Candidate genes associated with validated SNP loci

BLAST against the genomes of nine different crops revealed fifty-five candidate sequences for the validated loci from chromosomes 3A, 4B, and 6B (Appendix 14). Putative gene functions included protein coding (27 sequences), putative E3 ubiquitin-protein ligase (4), TATA-box binding proteins (2), cytochrome b561 (2), heat stress transcription factor A (1), putative WRKY transcription factors (1), repressor of RNA pol III (1), auxin response factor (1), and zinc finger CCCH domain containing protein (1), among others. Two of the candidate genes have uncharacterized functions.

Analyses using a wheat exome-capture database also showed sequences with significant hits. There were three significant sequences (one per locus) in for the 3A (*IWGSC_CSS_3AL_scaff_3069047*), 4B (*IWGSC_CSS_4BL_scaff_7026506*), and 6B loci (*IWGSC_CSS_6BL_scaff_4224574*), with expected (E) values from 3.0E-95 to 1.0E-93 and representing a 198-200 bp region. Single nucleotide variants (SNVs) identified on the hit regions were present in 26 different mutant lines (corresponding to 16 ‘Cadenza’ and 10 ‘Kronos’ mutants; Appendix 15). Mutations identified were either a C → T or a G→A, both of which can

either lead to synonymous (no amino acid change), missense, intron, or upstream gene variants. Reciprocal transitions (i.e. A→G or T→C) were not present among the SNVs. Moreover, base pair changes leading to "stop" codons were not identified.

Discussion

Winter wheat QTL on 3A, 4B, and 6B were validated on WAMI spring wheat panel

Validation of loci across genetic backgrounds is an essential step for MAS. To the best of our knowledge, this is the first report validating winter wheat GY related QTL and SNP markers, identified through bi-parental and association mapping, in spring wheat. Of the 11 winter wheat QTL selected for validation, three were significantly associated with GY or yield components in the WAMI panel and may have utility for MAS. *IWA1818* (4B) and *IWA755* (6B) were both associated with GY. Lozada et al. (2017; published, Chapter II) showed *IWA1818* to be associated with GY and kernel number spike^{-1} (KNS) in a SRWW panel evaluated across southern U.S. environments. *IWA755* (6B) was associated with multiple traits including GY, TGW, and spike density in the PA bi-parental population described earlier (Addison et al. 2016), with additive effects of 0.073 and 0.012 t ha^{-1} for GY in two environments in 2014. The effect of the minor allele for GY was negative for the 4B locus (associated with GY and KNS in the winter wheat AMP) in both SRWW and WAMI panels, suggesting that similar selection pressures were placed on selecting for the major 'C' allele for yield improvement.

IWA3560 (3AL) was reported by Addison et al. (2016) to be associated with GY in the PA bi-parental mapping population (phenotypic variation explained, R^2 = 4.6%) in five southern US environments with an additive effect of 0.049 t ha^{-1}. In the WAMI panel, *IWA3560* was significant for both GNO and TGW and was stable across seven different datasets including MI10BLUE and MHD10BLUE for GNO. Chromosome 3A is a major determinant of GY related

traits in wheat, with several QTL and genes identified, and characterized (e.g. Ma et al. 2016; Mengistu et al. 2012; Rustgi et al. 2013). The negative correlation between GNO and TGW confirmed the relationship observed by Griffiths et al. (2015) using a RIL population derived from crossing lines with high GW and GNO. The allele effects for *IWA3560* were positive for GNO and negative for TGW, demonstrating a genetic trade-off between these yield components (Sukumaran et al. 2018). The results presented herein supported our previous hypothesis that winter and spring wheat share common QTL regions that control GY and component traits.

Positive alleles at all three validated loci resulted in the highest mean grain yield

The highest mean GY was observed with the A-C-G haplotype (4.55 t ha^{-1}), with favorable effects at all loci. This confirmed the initial hypothesis that an additive response to the number of favorable alleles would be observed resulting to the highest GY. Nevertheless, it should be noted that the mean GY for this haplotype did not differ significantly with other allele combinations (A-C-A, G-C-A, and G-C-G). One potential explanation for this is the marginal values for allele effects (~0.05 to 0.08); thus substituting one allele might not be sufficient to capture significant differences for GY among haplotypes. The fact that two of these validated loci for the A-C-G haplotype (*IWA1818* and *IWA755*) had major alleles with positive effects also suggests that favorable SNP have already been selected for in the WAMI population to improve GY. While selecting lines through MAS containing this haplotype may lead to an improvement in mean GY within breeding populations in some environments, it is not an absolute as some low yielding lines had the A-C-G combination. This is likely due to both lack of linkage between the favorable allele and the functional gene in some lines and of the additional favorable alleles for GY present in the population. Ellis et al (2007) noted that even when tight linkage is observed, evolutionary processes and the large number of generations in plant breeding provide the opportunity for loss

of association between alleles and traits in different wheat populations. This was observed recently by Emebiri et al. (2017) where some lines possessing desirable alleles for Sunn resistance were phenotypically susceptible to the disease. Testing these assays on additional genetic backgrounds could further confirm the effects of these haplotypes.

Candidate genes associated with the validated SNPs reflect genetic complexity of grain yield

The quantitative inheritance of GY makes identifying the genes underlying QTL a challenge. Predicted putative gene functions at the 4B locus included a repressor of RNA Pol III transcription (*Oryza sativa* japonica) and a transcription factor (*Arabidopsis thaliana*). Transcription factors have been associated with plant adaptation to abiotic stresses, including the APETALA7 (*AP7*) and an inducible *T. aestivum* nuclear factor Y (A subunit)-B1 (*TaNFYA-B1*), which increased GY in rice under drought stress (Kim and Kim, 2009) and in wheat under low P and N (Qu et al. 2015) respectively when overexpressed at different developmental stages. A positive regulator of ubiquitin protein ligase activity meanwhile was identified for the 6B locus when compared to the *A. thaliana* genome. Several studies on the ubiquitin mediated control of seed size in Arabidopsis and rice have been reported (reviewed in Li and Li, 2014). Song et al. (2007) found the rice grain weight 2 (*GW2*) QTL to encode a RING-type protein with ubiquitin ligase activity and loss of function resulted in increased grain width, length, and yield. Similarly, seed size in Arabidopsis is influenced by the *DA1* gene family encoding a predicted ubiquitin receptor (Li et al. 2008). The gene functions associated with the validated loci demonstrate the complex regulation of GY, as many of the putative genes have roles associated with developmental and biochemical processes across different crop species. However, further work is necessary to confirm the functions of these genes in wheat.

Sequence hits against a recently developed exome capture platform (Krasileva et al. 2017) for the validated QTL revealed these loci to be within the coding regions of the wheat genome. Analyses demonstrated that mutations (i.e. SNV) on these sites could either lead to synonymous (i.e. no change) or changes (missense) in the amino acid composition for the corresponding protein. The most common SNV for the hit regions for the three loci was a change from C→T, causing mostly synonymous mutations. It is interesting to note that BLAST analysis for the 4B loci did not show SNVs causing either type of mutations, but only intron variants, a possible consequence of low gene density on this region. No SNV causing "stop" codons were identified, which could be attributed to the low frequency of these variants (<1.50% for both the 'Cadenza' and 'Kronos' mutant lines).

Mutagenesis through either physical, chemical, or TILLING strategies have been primarily used in plant breeding programs for forward genetic screening, generating genetic diversity, and studying important traits in cereals (Uauy et al. 2009; Rakszegi et al. 2010). Exploiting these mutations could help understand gene functions, identify novel alleles, and reveal the hidden variations in polyploid wheat (Krasileva et al. 2017). Mutant populations from these genetic stocks could ultimately be incorporated into existing wheat breeding programs to examine the effects of these specific mutations on GY and component traits.

Conclusions

Grain yield QTL on chromosomes 3A, 4B, and 6B previously identified in winter wheat were validated on a CIMMYT spring wheat panel. In spring wheat, these loci were significantly associated with GY and yield components across international testing environments, demonstrating their robust potential for use in MAS. Candidate genes underlying these validated QTL had putative functions reportedly involved in the regulation of GY. Mutations on the

validated loci were identified to either result in no changes or some modifications in the amino acid sequences of the coded proteins. Identifying and validating loci responsible for complex traits such as GY from diverse unrelated panels represents a first step in bridging the gap for molecular breeding for both classes of wheat. Results of this study will enable MAS for these QTL in spring and winter wheat and serve as a resource for future map based cloning and functional genomic studies for these important loci.

References

Acuña-Galindo, M.A., R.E. Mason, N.K. Subramanian, and D.B. Hays. 2015. Meta-analysis of wheat QTL regions associated with adaptation to drought and heat stress. Crop Sci. 55:477-492.

Addison C.K., R.E. Mason., G. Brown-Guedira, M., Y. Hao, R.G. Miller, N. Subramanian, D.N. Lozada, A. Acuna, M.N. Arguello, J.W. Johnson, A.M.H. Ibrahim, R. Sutton, and S.A. Harrison. 2016. QTL and major genes influencing grain yield potential in soft red winter wheat adapted to the southern United States. Euphytica:1-13.

Alvarado G., M. López, M. Vargas, Á. Pacheco, F. Rodríguez, J. Burgueño, and J. Crossa. 2016. META-R (Multi Environment Trail Analysis with R for Windows) Version 6.0 http://hdl.handle.net/11529/10201. International Maize and Wheat Improvement Center.

Arguello M.N., R.E. Mason, T.L. Roberts, N. Subramanian, A. Acuña, C.K. Addison, D.N. Lozada, R.G. Miller, and E. Gbur. 2016. Performance of soft red winter wheat subjected to field soil waterlogging: Grain yield and yield components. Field Crops Res. 194:57-64.

Bokore F.E., R.E. Knox, R.M. DePauw, F. Clarke, R.D. Cuthbert, H. Campbell, A. Brûlé-Babel J. Gilbert., and Y. Ruan. 2017. Validation of Molecular Markers for Use With Adapted Sources of Fusarium Head Blight Resistance in Wheat. Plant Dis. 101:1292-1299.

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

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

Dao H.Q., P.F. Byrne, S.D. Reid, and S.D. Haley. 2017. Validation of quantitative trait loci for grain quality-related traits in a winter wheat mapping population. Euphytica 213:5.

Dreisigacker S., D. Sehgal, A. Reyes Jaimez, B. Luna Garrido, C. Nunez Rios., J. Mollins, and S. Mall. 2016. CIMMYT Wheat Molecular Genetics: Laboratory Protocols and Applications to Wheat Breeding. Mexico, D.F.: CIMMYT.

Ellis M.H., D.G. Bonnett, and G.J. Rebetzke. 2007. A 192 bp allele at the Xgwm261 locus is not always associated with the Rht8 dwarfing gene in wheat (*Triticum aestivum* L.). Euphytica 157:209–214.

Edae E.A., P.F. Byrne, S.D. Haley, M.S. Lopes, and M.P. Reynolds. 2014. Genome-wide association mapping of yield and yield components of spring wheat under contrasting moisture regimes. Theor. Appl. Genet. 127:791-807.

Emebiri LC, M.K. Tan, M. El-Bouhssini, O. Wildman, A. Jighly, W. Tadesse , and F.C. Ogbonnaya. 2017. QTL mapping identifies a major locus for resistance in wheat to Sunn pest (Eurygaster integriceps) feeding at the vegetative growth stage. Theor. Appl. Genet. 130 (2):309-318.

Green A.J., G. Berger, C.A. Griffey, R. Pitman, W. Thomason, M. Balota, and A. Ahmed. 2012. Genetic Yield Improvement in Soft Red Winter Wheat in the Eastern United States from 1919 to 2009. Crop. Sci. 52:2097-2108.

Griffiths S, Wingen L, Pietragalla J, Garcia G, Hasan A, et al. 2015. Genetic Dissection of Grain Size and Grain Number Trade-Offs in CIMMYT Wheat Germplasm. PLoS ONE 10(3): e0118847. https://doi.org/10.1371/journal.pone.0118847

Kim, Y. S., and J. K. Kim. 2009. Rice transcription factor AP37 involved in grain yield increase under drought stress. Plant Signal. Behav. 4: 735–736.

Krasileva KV, H.A. Vasquez-Gross, T. Howell , P. Bailey P, F. Paraiso, L. Clissold, J. Simmonds , R.H. Ramirez-Gonzalez, X. Wang, P. Borrill , C. Fosker, S. Ayling, A.L. Phillips, C. Uauy, and J. Dubcovsky. 2017. Uncovering hidden variation in polyploid wheat. Proc. Natl. Acad. Sci. 114:913-921.

Li N. and Y. Li. 2014. Ubiquitin-mediated control of seed size in plants. Front. Plant. Sci. 5:332.

Li Y., L. Zheng, F. Corke, C. Smith, and M.W. Bevan. 2008. Control of final seed and organ size by the DA1 gene family in Arabidopsis thaliana. Genes. Dev. 22:1331-1336.

Lipka A.E., F. Tian, Q. Wang, J. Peiffer, M. Li, P.J. Bradbury, M.A. Gore, E.S. Buckler, and Z. Zhang. 2012. GAPIT: genome association and prediction integrated tool. Bioinformatics 28:2397-2399.

Lopes M.S., M.P. Reynolds, M.R. Jalal-Kamali, M. Moussa, Y. Feltaous, I.S.A. Tahir, N. Barma, M. Vargas, Y. Mannes, and M. Baum. 2012. The yield correlations of selectable physiological traits in a population of advanced spring wheat lines grown in warm and drought environments. Field Crops Res. 128:129-136.

Lopes M.S., S. Dreisigacker, R.J. Pena., S. Sukumaran, M.P. Reynolds. 2015. Genetic characterization of the wheat association mapping initiative (WAMI) panel for dissection of complex traits in spring wheat. Theor. Appl. Genet. 128:453-464.

Lozada, D.N., R.E. Mason, M.A. Babar, B. Carver, G. Brown-Guedira, et al. 2017. Association mapping reveals loci associated with multiple traits that affect grain yield and adaptation in soft winter wheat. Euphytica 213:213-222.

Ma, L., T. Li, C. Hao, Y. Wang, X. Chen, and X. Zhang. 2016. *TaGS5-3A,* a grain size gene selected during wheat improvement for larger kernel and yield. Plant Biotechnol. J. 14:1269–1280.

Mengistu N., P. Baenziger, K. Eskridge, I. Dweikat, S. Wegulo, K. Gill, and A. Mujeeb-Kazi. 2012. Validation of QTL for grain yield-related traits on wheat chromosome 3A using recombinant inbred chromosome lines. Crop. Sci. 52:1622-1632.

Mwadzingeni L, H. Shimelis, D.J.G. Rees, and T.J. Tsilo. 2017. Genome-wide association analysis of agronomic traits in wheat under drought-stressed and non-stressed conditions. PLoS One 12(2):e0171692.

Petersen S., J.H. Lyerly, A.L. McKendry, M.S. Islam, G. Brown-Guedira, C. Cowger, Y. Dong, and J.P. Murphy. 2017. Validation of Fusarium Head Blight Resistance QTL in US Winter Wheat. Crop. Sci. 57:1-12

Prat N., C. Guilbert, U. Prah, E. Wachter, B. Steiner, T. Langin, O. Robert, and H. Buerstmayr. 2017. QTL mapping of Fusarium head blight resistance in three related durum wheat populations. Theor. Appl. Genet. 130:13-27

Qu B., X. He, J. Wang, Y. Zhao, W. Teng, A. Shao, et al. 2015. A Wheat CCAAT Box-Binding Transcription Factor Increases the Grain Yield of Wheat with Less Fertilizer Input. Plant Physiol. 167:411-423.

Rasheed A, W. Wen, F. Gao, S. Zhai, H. Jin, J. Liu, Q. Guo, Y. Zhang, S. Dreisigacker, and X. Xia. 2016. Development and validation of KASP assays for genes underpinning key economic traits in bread wheat. Theor. Appl. Gen. 129:1843-1860.

Ramirez-Gonzalez R.H., C. Uauy, and M. Caccamo. 2015. PolyMarker: a fast polyploid primer design pipeline. Bioinformatics 31:2038-2039.

Rakszegi M., B.N. Kisgyörgy, K. Tearall, P.R. Shewry, L. Láng, A. Phillips, and Z. Bedő. 2010. Diversity of agronomic and morphological traits in a mutant population of bread wheat studied in the Healthgrain program. Euphytica 174:409-421.

Rustgi S., M.N. Shafqat, N. Kumar, P.S. Baenziger, M.L. Ali, I. Dweikat I, T. Campbell, and K.S. Gill. 2013. Genetic Dissection of Yield and Its Component Traits Using High-Density Composite Map of Wheat Chromosome 3A: Bridging Gaps between QTLs and Underlying Genes. PLoS One 8(7): e70526.

SAS Institute. 2011. SAS System Options: Reference, 2nd ed. SAS Institute, Cary, NC.

Sallam A., M. Arbaoui, M. El-Esawi, N. Abshire, and R. Martsch. 2016. Identification and verification of QTL associated with frost tolerance using linkage mapping and GWAS in winter faba bean. Front. Plant. Sci. 7: 7:1098.

Song X.J., W. Huang, M. Shi, M.Z. Zhu, and H.X. Lin. 2007. A QTL for rice grain width and weight encodes a previously unknown RING-type E3 ubiquitin ligase. Nat. Genet. 39:623-630.

Su Z., S. Jin, Y. Lu, G. Zhang, S. Chao, and G. Bai. 2016. Single nucleotide polymorphism tightly linked to a major QTL on chromosome 7A for both kernel length and kernel weight in wheat. Mol. Breed. 36:15. DOI 10.1007/s11032-016-0436-4.

Sukumaran S., S. Dreisigacker, M. Lopes, P. Chavez, and M. Reynolds. 2015a. Genome-wide association study for grain yield and related traits in an elite spring wheat population grown in temperate irrigated environments. Theor. Appl. Genet. 128:353-363.

Sukumaran S., M.P. Reynolds, M.S. Lopes, and J. Crossa. 2015b. Genome-wide association study for adaptation to agronomic plant density: A component of high yield potential in spring wheat. Crop Sci. 55:2609-2619.

Sukumaran, S., M.S. Lopes, S. Dreisigacker, L.E. Dixon, M. Zikhali, et al. 2016. Identification of Earliness Per Se Flowering Time Locus in Spring Wheat through a Genome-Wide Association Study. Crop. Sci. 56:2962-2672.

Sukumaran S., M. Lopes M, S. Dreisigacker, and M. Reynolds. 2018. Genetic analysis of multi-environmental spring wheat trials identifies genomic regions for locus-specific trade-offs for grain weight and grain number. Theor. Appl. Gen. doi:10.1007/s00122-017-3037-7.

Turner S.D. 2014. qqman: an R package for visualizing GWAS results using QQ and manhattan plots. bioRxiv:005165.

Uauy C., F. Paraiso, P. Colasuonno, R.K. Tran, H. Tsai, S. Berardi, L. Comai, and J. Dubcovsky. 2009. A modified TILLING approach to detect induced mutations in tetraploid and hexaploid wheat. BMC Plant. Biol. 9:115.

Valluru, R., M.P. Reynolds, W.J. Davies, and S. Sukumaran. 2017. Phenotypic and genome-wide association analysis of spike ethylene in diverse wheat genotypes under heat stress. New Phytol. 214: 271–283.

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

Wang Z., Y. Liu, H. Shi, H. Mo, F. Wu, Y. Lin, S. Gao, J. Wang, Y Wei, and C. Liu. 2016 Identification and validation of novel low-tiller number QTL in common wheat. Theor. Appl. Genet. 129:603-612.

You F.M., N. Huo, Y. Q. Gu, M. Luo, Y. Ma, D. Hane, G.R. Lazo, J. Dvorak, O.D. and Anderson. 2008. BatchPrimer3: a high throughput web application for PCR and sequencing primer design. BMC Bioinformatics 9:253.

Zhang, L.Y., D.C. Liu, X.L. Guo, W.L. Yang, J.Z. Sun, D.W. Wang, and A.M. Zhang. 2010. Genomic Distribution of Quantitative Trait Loci for Yield and Yield-related Traits in Common Wheat. J. Integr. Plant Biol. 52:996-1007.

Table 1. List of QTL from winter wheat for validation in the WAMI spring wheat panel from CIMMYT, Mexico.

SNP	SNP ID	Chr	Position (cM) [a]	Population [b]	Trait [c]	Allele	Variation explained (%)
wsnp_Ex_c7252_12452995	IWA4643	1A	31.31	AMP	TGW	T/C	14.0-17.0
wsnp_Ku_c557_1166684	IWA7173	1A	76.14	PA	GY	T/C	2.0
wsnp_Ex_rep_c102538_87682273	IWA5068	2A	108.46	AMP	GY	T/C	26.0-27.0
wsnp_Ex_c361_708712	IWA3560	3A	177.24	PA	GY	A/G	4.6
wsnp_Ex_rep_c104141_88935451	IWA5112	3A	268.79	PA	GY	A/C	3.5
wsnp_Ex_c13849_21698240	IWA1818	4B	71.29	AMP	GY	T/C	8.0-26.0
wsnp_Ex_c48922_53681502	IWA4041	4B	77.72	AMP	GY, KWS	T/C	21.0-28.0
wsnp_Ex_c1276_2445537	IWA1679	6B	73.93	AMP	KWS	T/C	10.0
wsnp_Ku_c11690_19042937	IWA6428	6B	71.97	PA	GY	T/C	3.5
wsnp_CAP11_c3599_1741800	IWA755	6B	66.76	PA	GY	A/G	5.2
wsnp_CAP7_c1860_917952	IWA1053	7A	212.37	AMP	GY	A/C	9.0

[a] Position based on consensus map by Cavanagh et al., (2013) and Wang et al., (2014)
[b] *AMP* Association mapping panel; *PA* 'Pioneer'/ 'AGS' biparental mapping population
[c] *GY* grain yield; *KWS* kernel weight per spike; *TGW* thousand grain weight

Table 2. SNPs from winter wheat that were validated and showed significant associations on the WAMI spring wheat panel

SNP	Chr	Pos (cM) [a]	Allele	Winter wheat population (Associated trait)	Dataset [b]	*p*-value	Maf [c]	R^2 [d]	Allele effects [e]	Trait
wsnp_Ex_c361_708712	3A	177.24	A/G	PA (GY) [f]	BLUPINDH	4.40x10^{-4}	0.26	0.04	306.22	GNO
					ABLUP	3.51x10^{-6}	0.26	0.11	312.92	GNO
					BLUPMEXH	4.91x-10^{-5}	0.26	0.12	368.12	GNO
					MHD10BLUE	3.81x10^{-5}	0.26	0.12	616.05	GNO
					BLUPH10	6.44x10^{-6}	0.26	0.15	359.61	GNO
					MI10BLUE	1.17x10^{-5}	0.26	0.15	390.00	GNO
					AFRBLUPH	8.60x10^{-4}	0.26	0.21	-0.75	TGW
wsnp_Ex_c13849_21698240	4B	71.29	T/C	AMP [g] (GY, KNS)	ABLUP	8.38x10^{-5}	0.08	0.09	-420.74	GNO
					ABLUP	9.38x10^{-5}	0.08	0.11	-0.10	GY
					BLUPINDH	4.40x10^{-4}	0.08	0.04	-316.68	GNO
					BLUPH10	2.50x10^{-4}	0.08	0.07	-0.06	GY
wsnp_CAP11_c3599_1741800	6B	66.76	A/G	PA[g] (GY)	ABLUPH	5.25x10^{-5}	0.23	0.08	-295.26	GNO
					ABLUPH	2.10x10^{-4}	0.23	0.09	-0.08	GY
					ABLUP	1.03x10^{-5}	0.23	0.11	-307.90	GNO
					ABLUP	4.00x10^{-5}	0.23	0.12	-0.07	GY
					BLUPH10	2.51x10^{-5}	0.23	0.12	-345.64	GNO
					ASIABLUPH	6.01x10^{-5}	0.23	0.05	-319.29	GNO

[a] Position based on map by Wang et al. (2014)

[b] Dataset with 'H' indicates that only locations with heritability $H^2 > 0.50$ were used for analysis; refer to Appendix 13 for the full description of phenotypic datasets used for GWAS

[c] Minor allele frequency

[d] Phenotypic variation explained (R^2) for model with SNP

[e] Allele effects with respect to the minor allele

[f] *PA*- Pioneer/AGS biparental mapping population; *GY*- grain yield

[g] *AMP*- Winter wheat association mapping panel; *KNS*- kernel no. spike^{-1}

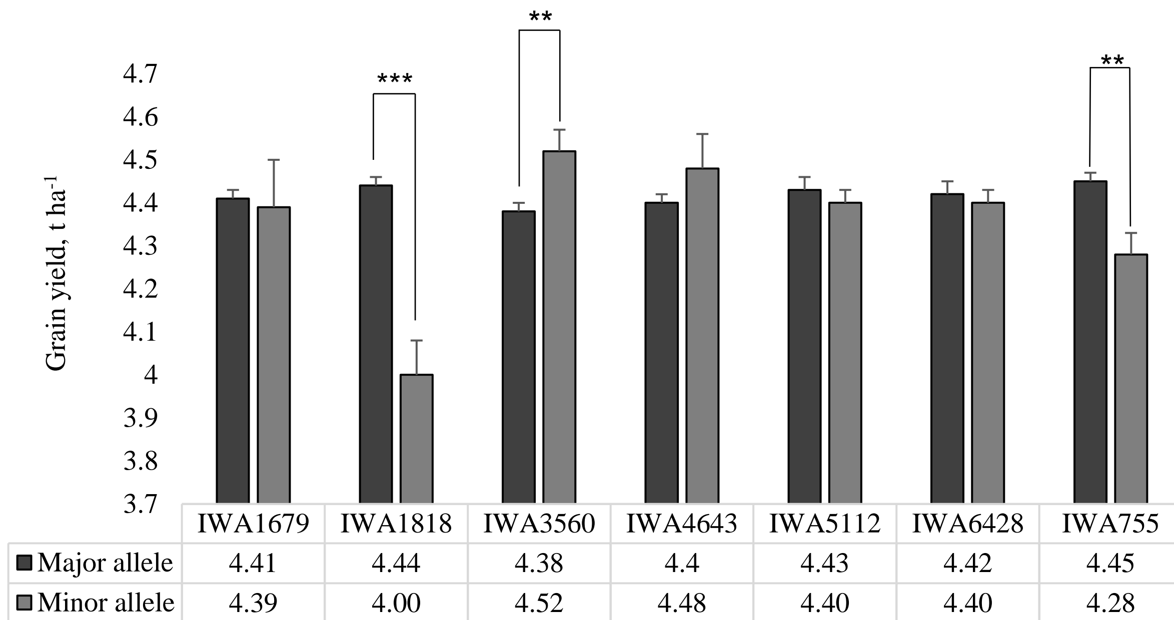


Figure 1. Mean and standard error comparisons of seven segregating SNP markers on the WAMI spring wheat panel for grain yield, mean across 29 environments. *IWA1679* (6B; **T**/C); *IWA1818* (4B; T/**C**); *IWA3560* (3A; A/**G**); *IWA4643* (1A; **T**/C); *IWA5112* (3A; **A**/C); *IWA6428* (6B; **T**/C); *IWA755* (6B; A/**G**). **- *t*-test LSD significant at $P < 0.01$, ***- significant at $P < 0.001$. Error bars indicate standard error. Major alleles are in bold and underscored.

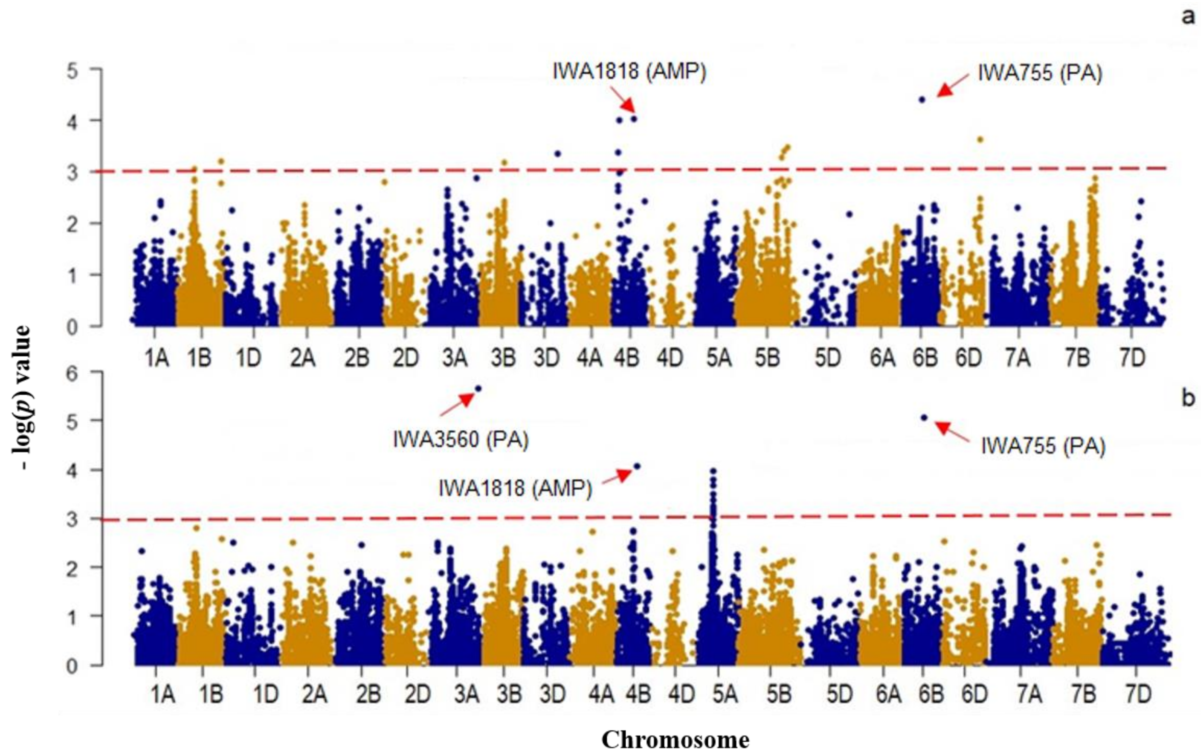


Figure 2. Manhattan plots showing association of grain yield (GY) and grain number (GNO) QTL from winter wheat on the WAMI spring wheat panel across all environments (ABLUP dataset) for (a) GY and (b) GNO. *PA*- 'Pioneer 26R61'/'AGS 2000' biparental mapping population; *AMP*- winter wheat association mapping panel. Horizontal line represents the threshold by which a marker was considered to be significantly associated with a trait ($p < 0.001$; $-\log(p) \geq 3.0$).

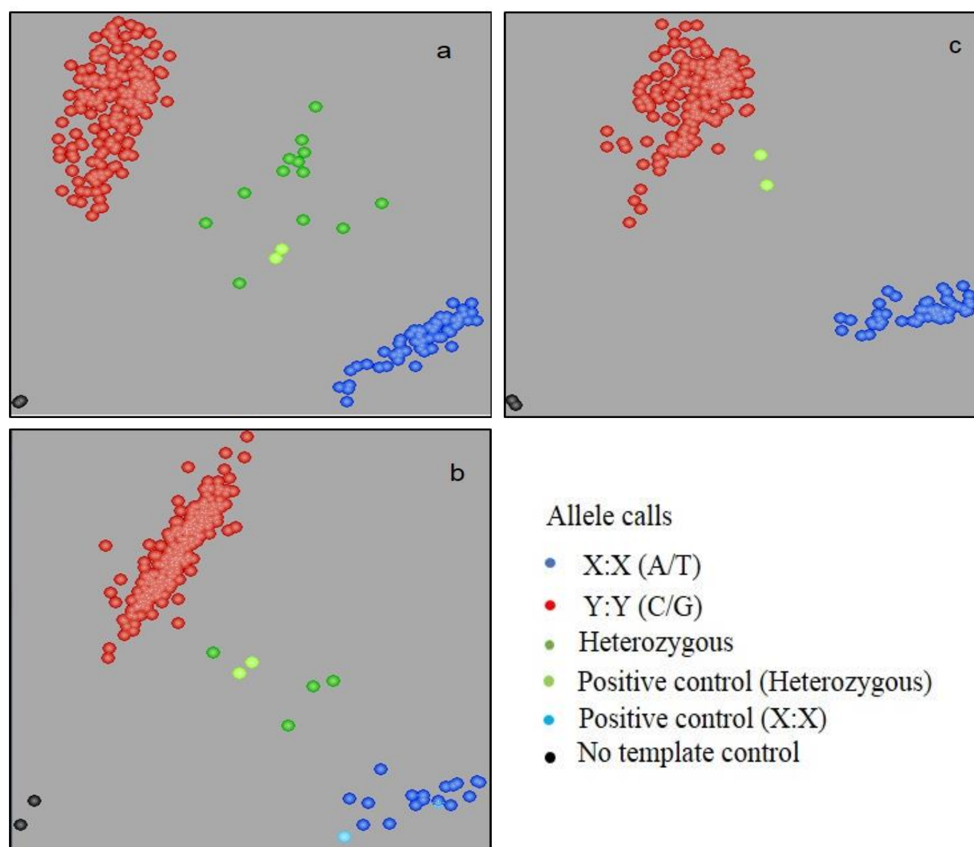


Figure 3. . Marker segregation for the designed assays: a) *IWA3560* (3A; *PA*- 'Pioneer 26R61'/'AGS 2000' biparental mapping population); b) *IWA1818* (4B; *AMP*- winter wheat association mapping panel); and c) *IWA755* (6B; PA).

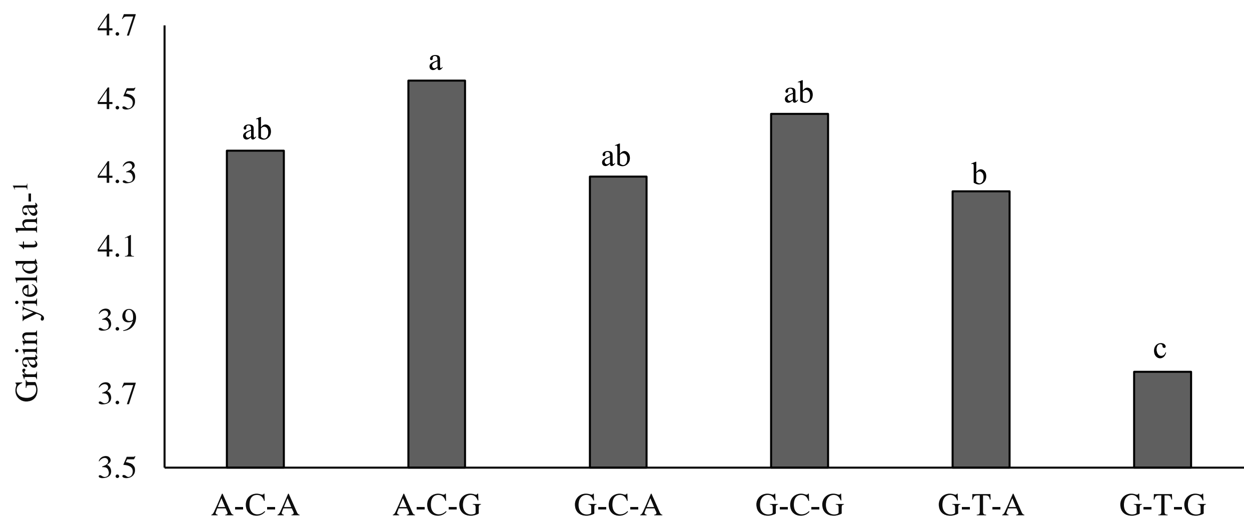


Fig. 4. Average grain yield across all environments (t ha^{-1}) for the different combinations of alleles for the winter wheat QTL validated at 3A (*IWA3560,* **A**/G), 4B (*IWA1818,* **C**/T), and 6B (*IWA755,* A/**G**) (allele combination in that order) loci. Means followed by the same letter do not differ significantly (*t*-test LSD; $p < 0.05$). Favorable alleles are underlined and written in bold.

CHAPTER IV

GENOMIC SELECTION FOR GRAIN YIELD AND AGRONOMIC TRAITS IN SOFT WINTER WHEAT

Abstract

Genomic selection (GS) has the potential to increase genetic gain by using molecular markers as predictors of breeding values. The effects of training population (TP) size, marker number, relatedness, and covariates on the accuracy of genomic predictions (*r*) for grain yield (GY) and agronomic traits were evaluated under a cross validation (CV) scheme using a population consisting of 239 soft red winter wheat (SRWW) cultivars and breeding lines. Increasing TP size resulted in an increase in *r*, with maximum prediction accuracies reached when ~80% of the lines were used as TP. Using subsets of associated markers increased accuracies by 64-70% for GY but resulted in lower *r* for traits with high heritability such as plant height. Inclusion of major growth habit genes as covariates generally increased GY predictability under a single population CV procedure. GS was "superior" to marker-assisted selection in terms of response to selection (*R)* and complementing phenotypic selection (PS) with GS resulted in the highest *R* for GY, leading to 10% gain compared to using PS alone. Forward prediction using the TP to predict GY of two biparental populations (*N*=100 and *N*=156 lines) derived from parents present in the TP resulted in *r* ranging from -0.14 and 0.43 dependent on the grouping of site-year data for the training and validation populations. Taken together, our results showed the effects of different factors on GS accuracies in SRWW and that complementing traditional PS with GS should increase the rate of genetic gains in wheat breeding programs.

Introduction

High-throughput genotyping technologies that generate large sets of DNA marker data at low-cost have accelerated the adoption of genomic selection (GS) in plant breeding programs (Patel et al., 2015). GS is a marker-assisted selection (MAS) tool that aims to predict and perform selection based on genomic estimated breeding values (GEBV) of individuals that are generated using genome-wide marker data through training and validation of a prediction model (Meuwissen et al., 2001). GS is a complement to traditional breeding strategies, potentially reducing the need for large-scale phenotyping and accelerating genetic gain through shorter breeding cycles (Heffner et al., 2010; Muranty et al., 2015; Nakaya and Isobe, 2012).

Pioneering studies on GS in animal breeding, particularly of cattle (Hayes et al., 2009; Meuwissen et al., 2001) have now been extended to crops, including rice (Onogi et al., 2016; Spindel et al., 2015), tomato (Duangjit et al., 2016; Hernández-Bautista et al., 2016), maize (Zhao et al., 2012), soybean (Bao et al., 2014), and barley (Lorenzana and Bernardo, 2009). In soft winter wheat, GS studies have been conducted for Fusarium head blight resistance (Arruda et al., 2016), grain yield and stability traits (Huang et al., 2016), yield, softness equivalence, and flour yield (Hoffstetter et al., 2016), grain yield, plant height, heading date, and flour quality traits (Heffner et al., 2011b), and normalized difference vegetative index (NDVI) (Mason et al., 2017).

The performance of GS depends primarily on prediction accuracy, defined as the Pearson's correlation between the selection criterion and the true breeding value to select individuals with unknown phenotypes (Desta and Ortiz, 2014). Factors affecting GS accuracy include gene effects, genetic composition of the training population (TP), level of LD, marker density, model performance, QTL number, relationship between TP and the validation

population (VP) or selection candidates, TP size, and trait heritability (Desta and Ortiz, 2014; Rutkoski et al., 2015; Zhong et al., 2009).

Muleta et al. (2017) recently evaluated the effects of trait architecture, size of TP, and different marker densities on GS accuracies for stripe rust in a diverse collection of spring wheat. Currently however, there is no report on the effects of TP size, marker number, and relatedness in GS accuracy for a population of soft winter wheat (SWW) lines having different genetic backgrounds and pedigrees (i.e. "diverse") and are adapted to the southeastern region of the US. Our objectives are to (1) evaluate the effects of TP size, marker number, covariates, and relatedness on genomic prediction accuracy in a TP consisting of SWW breeding lines and cultivars; (2) validate prediction models in two bi-parental populations related to the TP, and; (3) compare phenotypic (PS), genomic (GS), and marker-assisted (MAS) selection strategies in terms of response to selection (*R*) for GY.

Materials and Methods

Plant Material

The population used for training and cross-validation in this study consisted of a panel of soft winter wheat cultivars and breeding lines previously utilized for an association study (referred to as TP; *N*=239 lines) and comprised of genotypes from the SunGrains® (Southeastern University Grains) Breeding Cooperative and others adapted to the southeastern region of the US (Lozada et al., 2017; published, Chapter II). Two bi-parental populations were used for forward validation of prediction models, including: (1) A recombinant inbred line (RIL) population derived from a cross between soft winter wheat cultivars 'Pioneer Brand 26R61' and 'AGS 2000' (PI612956);" (PA-RIL; *N*=156, Addison et al., 2016), and; (2) a double haploid (DH) population derived from a cross between 'NC-Neuse' (PI633037) and 'Bess' (PI 642794) (NB; *N*=100) (Petersen et al.,

2017). Development of the PA-RIL was described previously by Hao et al. (2011) and QTL analyses for yield and component traits and genomic predictions for GY and spectral reflectance were previously reported (Addison et al. 2016; Mason et al., 2017). 'NC-Neuse' was released in 2003 by North Carolina State University (Murphy et al., 2004) and 'Bess' was released in 2005 from the University of Missouri (McKendry et al., 2007). Development and evaluation for Fusarium head blight (FHB) resistance traits of the NB mapping panel was previously reported by Petersen et al. (2017).

Genotype data

The TP and PA-RIL were genotyped using the Illumina® 9K single nucleotide polymorphism (SNP) chip (Cavanagh et al., 2013) through the USDA-ARS Eastern Regional Genotyping Laboratory in Raleigh, NC while NB was genotyped with the 90K *iSelect* assay (Wang et al., 2014). After filtering and quality control, 5,661, 1,188, and 2,780 SNP markers remained for the TP, NB, and PA-RIL, respectively. Genotype data were converted into a numeric format (0,1,2) for GS using the 'GAPIT' package (Lipka et al., 2012) in RStudio (R Development Core Team, 2010). Imputation for missing data was done using a kinship-based "expectation maximization" (EM) algorithm (Poland et al., 2012). A total of 1,089 and 1,632 common SNP markers were used for forward validation with the NB and PA-RIL as VP, respectively.

Phenotype data

Data consisted of BLUP values derived from adjusted means evaluated based on an augmented design for the TP. Adjusted (least square) means for each genotype were estimated using a restricted maximum likelihood (REML) approach using the PROC MIXED function in SAS v.9.4 (SAS Institute, 2011). Measured traits for the TP included GY, plant height (PH), heading date (HD), and yield components kernel number spike^{-1} (KNS), kernel weight spike^{-1} (KWS),

and thousand kernel weight (TKW) collected in eight total site-years in Arkansas and Oklahoma (Okmulgee), U.S. for 2014 and 2015 planting seasons. Collection and analyses of the phenotypic data were described previously by Lozada et al. (2017; published, Chapter II).

Data for GY of the PA-RIL was similar with those used for genomic predictions by Mason et al. (2017). The PA-RILs were grown in three growing seasons (2012-2014) over twelve site-years in Arkansas (Fayetteville, (FAY12, FAY13, FAY14); Stuttgart (STU13; STU14); and Marianna (MAR13, MAR14), Georgia (Plains; GA12, GA13), Louisiana (Baton Rouge; LA13), and Texas (Farmersville; TX12, TX13) in a randomized complete block design (RCBD) with two replications per site-year. Collection and analyses of the phenotypic data were described previously by Addison et al. (2016) and Mason et al. (2017). GY data for NB was collected in a total of five site-years, including in Fayetteville (FAY15, FAY16, and FAY17), and Newport, AR (NPT16 and NPT17) in an RCBD with two reps per site year except FAY15 (single replication) due to limited seed. GY were recorded by harvesting whole plots, weighing the grains, and adjusting for 13% moisture.

Genomic selection model

A ridge regression best linear unbiased prediction (RRBLUP) model was used for genomic selection (GS). All analyses were done in R using the 'rrBLUP' package (Endelman, 2011). RRBLUP considers additive marker effects and is based on the infinitesimal model with all markers sharing a common variance and all effects are shrunken toward zero but allows for markers to have uneven effects (Arruda et al., 2016; He et al., 2016; Heffner et al., 2011a; Meuwissen et al., 2001). The basic RRBLUP model is

$$Y = \mathbf{WGu} + \acute{\varepsilon},$$

where **u** ~ N (0, $I\sigma^2_u$) is a vector of marker effects, **G** is the genotype matrix (e.g. (aa, Aa, AA) = (-1,0,1) for bialleleic SNP under an additive model), and **W** is the design matrix relating lines to observations (Endelman, 2011).

Genomic selection scenarios

Two GS scenarios were evaluated in this study: (1) a standard single population CV scheme where the effects of different factors namely, training population size, marker number, relatedness, and covariates on prediction accuracy were evaluated and (2) forward predictions, where the TP was used to predict GY in NB and PA-RIL bi-parental populations using 1,089 and 1,632 SNP markers, respectively. All scenarios used RRBLUP model for genomic predictions.

Different factors affecting prediction accuracy for the TP

Size of the training population

To test the effect of training population (TP) size on prediction accuracy for GY, yield components, and agronomic traits, 50 different subsets of 25, 50, 75, 100, 125, and 150 lines were sampled as TP at a constant VP size of 60. This analysis used the BLUP values across all environments (ABLUP) for the TP.

Number of markers

Subsets of markers with varying levels of significance, namely, subset $SS_{0.15}$ ($p < 0.15$), $SS_{0.10}$ ($p < 0.10$) and $SS_{0.05}$ ($p < 0.05$) derived from genome-wide association analysis were used to perform predictions to examine the effects of marker number (N_M) on GS accuracy. To determine the marker subsets, a total of 10 different TP (N=219) and VP (N=20) sets were generated, and an independent association analyses using the GAPIT package (Lipka et al. 2012) in R under a *K-PC* model (with number of PC = 3) was performed with each TP only and ABLUP dataset. This was done to prevent "inside trading effect," which occurs when prediction

accuracies are evaluated using QTL that were previously identified in the same group of lines, potentially resulting to overestimated accuracies (Arruda et al., 2016). Whole genotype data were filtered for p-values corresponding to marker $SS_{0.15}$, $SS_{0.10}$, and $SS_{0.05}$ from each cycle of GWAS. Mean accuracy for each round of GWAS-GS (total of 10) for each marker SS was recorded.

Relatedness and population structure

The effects of relatedness between the TP and VP were evaluated by grouping the lines based on corresponding membership coefficient, Q values derived from STRUCTURE (Lozada et al. 2017; published, Chapter II) and performing predictions where each subpopulation was used to predict the GY and component traits of other subgroups. Given that there was an uneven number of lines on each of the subgroups, a subset of 50 and 30 lines were used as TP and VP, respectively, to perform predictions under a 10-fold CV for the measured traits.

Covariates in the model

Covariates including growth habit genes for photoperiod (*Ppd-D1*) and vernalization requirement (*vrn-A1*) were included in the model as fixed effects, either individually or in combination. GS accuracies with or without the presence of these covariates were compared under a 10-fold CV for TP size= 144.

Response to selection for grain yield using the TP

Response to selection, R for mean GY across eight site-years was calculated using the formula $R = h^2S$ (Falconer and Mackay, 1996), where h^2 is the heritability for GY in Lozada et. al (2017), equal to 0.48; and S is the selection differential calculated as the difference between the population mean and mean of population with selection strategy applied $S = \mu_S - \mu_P$, under a selection intensity of 10% (i.e. selecting the top 25 lines based on average GY and GEBV). Selection strategies included phenotypic selection (PS), marker assisted selection (MAS),

genomic selection (GS), random selection (RS), and a combination of PS and GS (PS+GS). Mean for GY under PS (μ_{PS}) was calculated based on the top 25 highest yielding lines; μ_{MAS} was equal to the mean GY of the lines having the favorable alleles for three loci, *wsnp_Ex_c2723_5047696* (3B), *wsnp_Ex_c13849_21698240* (4B), and *wsnp_Ex_c48922_53681502* (4B), significantly associated with GY and previously reported (Lozada et al., 2017; published, Chapter II); μ_{GS} was equal to the mean of lines having the top 25 GEBV in 10 different rounds of GS under a 10-fold CV in RRBLUP, with TP size =144; μ_{RS} was computed based on a function to generate 25 random selections, 10 different times and calculating the mean for these selections; μ_{GS+PS} was equal to the mean of the top entries based on average GY and GEBV.

Forward validation of genomic selection model using biparental populations

The TP (*N*=239) was used to predict GY in the PA-RIL (*N*=157) and NB (*N*=100) mapping populations using RRBLUP model. Datasets used for the training set were BLUP across all environments (ABLUP), across northern locations (Fayetteville and Keiser, AR; Okmulgee, OK; NBLUP), and SBLUP (Marianna, Stuttgart, and Rohwer, AR). BLUP across all locations (NB_ALL), across Fayetteville (NB_FAY) and Newport (NB_NPT) were used as VP sets for NB. Site-year groupings based from previous site-regression analyses (Addison et al., 2016) were used for PA-RIL as VP. Simple matching coefficients between the TP and VPs were calculated using the nominal clustering ‘nomclust’ package and ‘sm’ function in R to evaluate relatedness between populations.

Results

Phenotypic data

The TP (N=239) consists of cultivars and breeding lines of SRWW adapted to the southeastern region of the US while VP included biparental populations derived from cross between SRWW cultivars ‘Pioneer26R61’ and ‘AGS2000’ (PA; N=156) and ‘NC-Neuse’ and ‘Bess’ (NB; N=100). Phenotypic data for the TP were reported previously by Lozada et al., (2017; published, Chapter II). Broad sense heritability (h^2) values of traits measured in the TP were 0.48 (GY), 0.63 (HD), 0.47 (KWS), 0.37 (KNS), 0.77 (TKW), and 0.81 (PH). Values of h^2 for GY datasets across the three populations ranged between 0.33 (PA_ALL) and 0.85 (PA_Cluster3), with mean GY between 2.82 (NB_NPT) and 5.56 t ha^{-1} (PA_Cluster3) (Table 1).

Effect of TP size

Increasing TP size increased r across all the measured traits when VP size was held constant and reached a maximum at TP150 (Fig. 1; Appendix 16). Comparing TP25 to TP150, prediction accuracies increased from 0.18 to 0.46 for GY, from 0.27 to 0.73 PH (the most heritable trait) and from 0.19 to 0.47 for HD. For yield components, r increased from 0.12 to 0.40 for KNS, 0.19 to 0.59 for KWS and 0.28 to 0.58 for TKW. A minimal increase in r was observed (between 4.6% and 20.5%) from TP125 to TP150 as accuracy values hit a plateau.

Effect of marker number

Average number of markers for each subset were 820 ($SS_{0.15}$), 540 ($SS_{0.10}$), and 270 ($SS_{0.05}$) SNPs. Prediction accuracies for GY increased when these subsets of significant markers were used for GS (Fig. 2; Appendix 17). For GY and compared to using the entire marker dataset, $SS_{0.05}$, SS_{10} and $SS_{0.15}$ resulted in 64, 70 and 64% increases in r, respectively. For PH, no change in r was observed for $SS_{0.10}$ and $SS_{0.15}$, with a 19% decrease observed for $SS_{0.05}$. For HD,

significant decreases in *r* were observed for all the marker SS. For yield components (KNS, KWS, and TKW) there was a 14-39% decrease in *r* for using the marker SS.

Effect of relatedness

Previous STRUCTURE analyses (Lozada et al. 2017; published, Chapter II) identified three subpopulations in the TP, *Q1* (*N*=59 lines), Q2 (*N*=54) and Q3 (*N*=126), with *Q2* and *Q3* the most related. On the average, using *Q2* to predict *Q3* (and vice versa) resulted to the highest accuracies, while using *Q1* to predict *Q2* resulted to the lowest accuracies for GY and yield components. For GY, there were no significant differences among GS accuracies when *Q2* was used in predicting *Q3* (and vice versa). Prediction accuracies of *r* = 0.09 and 0.10 were observed when *Q1* was used as a TP to predict *Q2* and *Q3,* respectively (Fig. 3; Appendix 18). Higher accuracies were observed when *Q2* was used to predict *Q1* (*r* = 0.22) and *Q3* (*r* = 0.30). Using *Q3* to predict *Q1* and *Q2* resulted to prediction accuracies of 0.09 and 0.26, respectively. Accuracies for KNS ranged between 0.07 (*Q1*/*Q2;* TP/VP) and 0.25 (*Q3/Q2*). For KWS, accuracies ranged between 0.04 (*Q1/Q2*) and 0.21 (*Q3*/*Q1*) while for TKW, accuracy values ranged between 0.08 (*Q1*/*Q2*) and 0.37 (*Q3*/*Q2*).

Effect of covariates

In general, genomic prediction accuracy for GY increased when *Ppd* and *vrn* were included in the model (Fig. 4; Appendix 19). For the ABLUP, there was an increase in *r* from 0.33 to 0.37 (12% increase) with the addition of *Ppd-D1*, while no increase was observed when *vrn-A1* was added. Using both *Ppd-D1* and *vrn-A1* as covariates simultaneously in the model had a greater effect on prediction accuracy for the ABLUP, BLUP14, and BLUP15 datasets compared to using only either gene as a covariate. Using *Ppd-D1* as a covariate increased GS accuracy for NBLUP

(from 0.09 to 0.13). No significant differences in accuracy were observed for the SBLUP when covariates were used.

Response to selection for grain yield

Response to selection *R* for GY was highest for PS+GS (0.34 t ha^{-1}), followed by PS (0.31 t ha^{-1}) and GS (0.21 t ha^{-1}) (Table 1), equal to a 22, 20, and 14% increase above the population mean, respectively. *R* for MAS was 0.08 t ha^{-1} and for RS was 0.01 t ha^{-1}, corresponding to a 3.8 and 0.63% increase above the population mean. Variance (σ^2) was highest for RS and MAS (both at 0.13) followed by GS (0.12), whereas PS and PS + GS exhibited the lowest σ^2 at 0.03.

Forward predictions in bi-parental populations

Accuracy of the TP to predict two related bi-parental populations ranged from r = -0.14 to 0.43 (Fig. 5; Appendix 20). Using NB as a VP resulted in prediction accuracies ranging from r = 0.06 to 0.22 while using PA-RIL as a VP resulted in prediction accuracies between r=-0.14 and 0.43. Grouping of site-years in both the TP and VP significantly affected accuracy. For example, PA_Cluster4 was the most predictable (mean r= 0.40) of the PA-RIL site-year groupings, compared to r = 0.23 in PA_ALL, where all VP site-years were included. Simple matching coefficients reveal a low to moderate similarity between the TP and the PA-RIL (0.48) and between the TP and NB (0.45). Overall, using major growth habit genes as covariates in the model did not improve the reliability of forward predictions (data not shown).

Discussion

This study reports the effects of different factors to GS accuracy for GY and agronomic traits in SRWW adapted to the southeastern region of the US. Among the parameters evaluated include number of markers, size of the TP, covariates and relatedness between the training and selection

candidates. Additionally, a panel comprised of SRWW cultivars having different pedigrees and genetic backgrounds were used to train a model to predict related biparental populations.

Marker number, training population size, and relatedness affect the accuracy of genomic prediction

Using subsets of markers for genome-wide prediction had varying effects on the accuracy of genomic selection. GY (h^2= 0.48) had higher prediction accuracies (an increase in *r* from 0.33 to 0.56) when subsets of associated markers were used. These results were consistent with a previous study in soft winter wheat where the highest accuracies were observed when subsets of statically associated markers ($p < 0.05$) were used (Hoffstetter et al. 2016). In other crops such as rice (Spindel et al., 2015) and soybean (Xavier et al., 2016), however, prediction accuracies decreased marginally when marker subsets were used for predicting GY. The use of evenly distributed markers was suggested in performing predictions for GY and related traits in rice, with the SNP position regarded as the most important factor for accuracy (Spindel et al., 2015). For traits other than GY, using marker subsets decreased GS accuracy, irrespective of heritability, in agreement with reports that showed the conservation of marker-QTL associations under higher marker datasets for increased prediction accuracy (Desta and Ortiz, 2014, Heffner et al., 2009). High marker number is of particular importance in diverse panels where there are many generations of recombination (Rutkoski et al., 2011).

By performing association analyses exclusively on the TP and using the significant loci identified from these as our marker subsets for predictions, we disregarded the “inside trading” effect that results when prediction accuracies are evaluated using QTL identified in the same group of lines (Arruda et al., 2016). In winter wheat, Arruda et al. (2016) previously demonstrated that “inside trading” can lead to inflated values (i.e. ~32% overall increase) for GS accuracies for FHB-

related traits when significant QTL were treated as fixed effects in the model. We thus showed here that even without "inside trading," it is still possible improve prediction accuracy for GY, which reached a maximum of 0.56 when $SS_{0.10}$ was used.

Increasing TP size increased prediction accuracies across all measured traits but tended to plateau between TP125 and TP150. In spring wheat, Muleta et al. (2017) noted that accuracy values either plateaued at the largest TP size or showed no sign of reaching a plateau depending on the environment and trait. A positive correlation between TP size and prediction accuracy has been observed for biparental and multifamily wheat populations (Heffner et al., 2011a, b), a soybean nested association mapping (NAM) population (Xavier et al., 2016), and elite breeding populations of oats (Asoro et al., 2011). Overall, increasing TP size increases prediction accuracy by improving the estimation of marker effects (Heffner et al., 2011b).

Aside from TP size, the composition and relatedness of the TP to VP significantly affected prediction accuracy. Using *Q2* to predict *Q3* (and vice versa) for GY and component traits gave an 85% advantage over using the less related subgroup *Q1*. These results agree with previous studies that showed higher prediction accuracies for more related populations (Heffner et al., 2011b; Xavier et al., 2016). In barley, the inclusion of unrelated individuals in a TP reduced accuracy compared to a TP consisting of only highly related individuals (Lorenz and Smith, 2015). Close relatives share long haplotype and linkage blocks resulting in minimal statistical bias in estimating breeding values and more accurate predictions (Hickey et al., 2014). In contrast, inconsistent QTL effects of distantly related TP and VP can result in lower predictions (Bassi et al., 2016).

Using markers for major genes as covariates improves predictability for grain yield in the TP

Including *Ppd-D1* and *vrn-A1* covariates in the RRBLUP model resulted in a general increase in the predictability of GY in the TP. Mason et al. (2017) reported the same trend when using major genes as covariates to predict GY using CV in the PA-RIL, particularly for site-year groupings with low heritability. The same study also reported that inclusion of multiple loci as fixed effects did not significantly improve prediction accuracies, which was thought to be due to a limited population size. On the other hand, no significant improvements or decreased in prediction accuracies were observed when TP was used to predict GY in the PA-RIL and NB (forward validation), even when covariates were included in the model. Adding covariates might not have been sufficient to capture genetic effects in the populations used, especially given that there is a limited relatedness between the two. Thus, inclusion of covariates may only be effective in improving accuracies under a standard single population CV scheme.

Incorporating markers linked to QTL in genomic prediction models was shown to improve accuracies for adult plant stem rust resistance (Rutkoski et al., 2014). Daetwyler et al. (2014) on the other hand showed that inclusion of marker scores for known rust resistant genes (*Lr34*/*Sr57*/*Yr18*) had more substantial effects on increasing genomic prediction accuracy than using markers linked to QTL. Overall, our results indicate that including loci influencing wheat phenology in the genomic prediction model can increase the accuracy of genomic selection for grain yield and other traits in diverse wheat germplasm under a CV procedure.

Combining genome-wide prediction with phenotypic selection resulted to the highest response to selection for GY

GS is a tool to complement PS in selecting "better" genotypes and cannot replace phenotypic selection (Desta and Ortiz, 2014). Within the parameters of this study, *R* for GS could only approach the level of PS and therefore showed a lower *R* (-32% change relative to PS). However, the highest accuracy was observed when GS was coupled with PS, resulting to a 10% increase in *R* compared to PS alone, demonstrating the potential for an integrated approach to increase genetic gain. In the current study, GS was also superior to MAS for three significant loci in terms of *R*, while using four or more significant QTL for MAS might not be beneficial as there would be lower number of individuals being selected. Arruda et al. (2016) observed higher selection differentials for GS compared to MAS using a maximum of five QTL associated with FHB-related traits in SRWW. In the same study, it was shown that decreasing selection intensity (i.e. selecting for fewer lines) resulted to an increased selection differential and hence increased *R*. Using simulations in maize double haploid populations, Bernardo and Yu (2007) demonstrated that across different QTL number and trait heritability, the response to GS was 18-43% greater than response to MAS, with an increase in *R* observed as heritability and the number of QTL increased.

Forward validations showed potential for predictive breeding of complex traits in winter wheat

The goal of GS is to predict the performance of new lines before testing them in the field. With this, we were interested in evaluating prediction accuracies using a TP (*N*=239 lines) to predict GY of biparental populations derived from the cross between parents belonging to our TP. Lower accuracies for GY resulted when NB (0.06-0.22) and PA-RIL (-0.14-0.43) were used as VP compared to when predicting within the TP alone through CV. Previously, Charmet et al. (2014)

reported low accuracies for GY, HD, and test weight using different sets of wheat DH and RIL populations for predictions (*r* ranged between -0.12 and 0.24). In forward prediction using the PA-RIL as VP, highest mean predictions were observed for Cluster 4, the site-year grouping with highest heritability, consistent with results from Mason et al. (2017). Within this cluster, using NBLUP dataset which had the highest heritability also resulted to the highest accuracies for GY, demonstrating the importance of heritability in obtaining higher predictions for complex traits.

Most of previous GS studies in wheat focused on single population CV of biparental (Heffner et al.,2011a, b) and diversity panels (Muleta et al., 2017), while previous reports in other crops such as rice (Ben Hassen et al., 2018) and sugar beet (Würschum et al., 2013) used diverse mapping populations to predict biparental families. While this approach is not yet widely implemented in wheat breeding, prediction accuracies for GY presented here (max. *r*= 0.43) demonstrated the potential of using diverse lines to predict complex traits in related biparental populations. In rice, it was recently shown that prediction models can be trained from a diverse reference population to predict performance among advanced progenies of biparental crosses, with reported prediction accuracies reaching a maximum value of 0.54 (Ben Hassen et al., 2018).

Conclusions

Of the factors studied, training population size had the greatest impact on prediction accuracy. Inclusion of covariates in prediction model increased accuracy for GY under a single population CV, but not when using multiple populations. Using the TP to predict new biparental populations showed promise Ultimately, GS could be exploited further with traditional PS to increase response to selection towards GY improvement and increasing genetic gains in wheat breeding programs. The effects of the evaluated parameters should be thoroughly considered when implementing genomic prediction strategies in winter wheat.

References

Addison, C.K., R.E. Mason, G. Brown-Guedira, M. Guedira, Y. Hao, R.G. Miller, et al. 2016. QTL and major genes influencing grain yield potential in soft red winter wheat adapted to the southern United States. Euphytica: 1-13. doi:10.1007/s10681-016-1650-1.

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

Asoro FG, Newell MA, Beavis WD, Scott MP, Jannink J-L (2011) Accuracy and training population design for genomic selection on quantitative traits in elite North American oats. Plant Genome 4 (2):132-144.

Bassi FM, Bentley AR, Charmet G, Ortiz R, Crossa J (2016) Breeding schemes for the implementation of genomic selection in wheat (Triticum spp.). Plant Sci 242:23-36. doi:https://doi.org/10.1016/j.plantsci.2015.08.021

Bao Y., Vuong T., Meinhardt C., Tiffin P., Denny R., Chen S., Nguyen H.T., Orf J.H., Young N.D. (2014) Potential of association mapping and genomic selection to explore PI 88788 derived soybean cyst nematode resistance. Plant Genome 7. Doi:10.3835/plantgenome2013.11.0039.

Ben Hassen M, Cao TV, Bartholomé J, Orasen G, Colombi C, Rakotomalala J, Razafinimpiasa L, Bertone C, Biselli C, Volante A, Desiderio F, Jacquin L, Valè G, Ahmadi N (2018) Rice diversity panel provides accurate genomic predictions for complex traits in the progenies of biparental crosses involving members of the panel. Theor Appl Genet 131 (2):417-435. doi:10.1007/s00122-017-3011-4.

Bernardo R, Yu J (2007) Prospects for Genomewide Selection for Quantitative Traits in Maize. Crop Sci 47 (3):1082-1090. doi:10.2135/cropsci2006.11.0690.

Burstin J, Salloignon P, Chabert-Martinello M, Magnin-Robert J-B, Siol M, Jacquin F, Chauveau A, Pont C, Aubert G, Delaitre C, Truntzer C, Duc G (2015) Genetic diversity and trait genomic prediction in a pea diversity panel. BMC Genomics 16 (1):105. doi:10.1186/s12864-015-1266-1

Cavanagh C, Chao S, Wang S, Huang B, Stephen S, Kiani S, Forrest K, Saintenacc C, Brown-Guedira G, et al. (2013) Genome-wide comparative diversity uncovers multiple targets of selection for improvement in hexaploid wheat landraces and cultivars. Proc Natl Acad Sci 110 (20):8057–8062. doi:www.pnas.org/cgi/doi/10.1073/pnas.1217133110

Charmet, G., E. Storlie, F.X. Oury, V. Laurent, D. Beghin, L. Chevarin, et al. (2014). Genome-wide prediction of three important traits in bread wheat. Mol Breed 34: 1843-1852. doi:10.1007/s11032-014-0143-y.

Cruz, C., L. Bhering and L. Peixoto (2016) Determination of the optimal number of markers and individuals in a training population necessary for maximum prediction accuracy in F 2 populations by using genomic selection models. Genet Mol Res. 15(4):gmr15048874

Daetwyler HD, Bansal UK, Bariana HS, Hayden MJ, Hayes BJ (2014) Genomic prediction for rust resistance in diverse wheat landraces. Theor Appl Genet 127 (8):1795-1803. doi:10.1007/s00122-014-2341-8.

Desta Z.A., Ortiz R. (2014) Genomic selection: genome-wide prediction in plant improvement. Trends Plant Sci 19:592-601.

Duangjit J., Causse M., Sauvage C. (2016) Efficiency of genomic selection for tomato fruit quality. Mol Breed 36:1-16. DOI: 10.1007/s11032-016-0453-3.

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

Falconer, D. S., & Mackay, T. F. (1996). Introduction to Quantitative Genetics, 4.

Hao, Y., Z. Chen, Y. Wang, D. Bland, J. Buck, G. Brown-Guedira, et al. 2011. Characterization of a major QTL for adult plant resistance to stripe rust in US soft red winter wheat. Theor Appl Gen 123: 1401-1411. doi:10.1007/s00122-011-1675-8.

Hayes B.J., Bowman P.J., Chamberlain A.C., Verbyla K., Goddard M.E. (2009) Accuracy of genomic breeding values in multi-breed dairy cattle populations. Genet Select Evol 41:1.

He S., Schulthess A.W., Mirdita V., Zhao Y., Korzun V., Bothe R., Ebmeyer E., Reif J.C., Jiang Y. (2016) Genomic selection in a commercial winter wheat population. Theor Appl Genet 129:641-651.

Heffner E.L., Jannink J.-L., Iwata H., Souza E., Sorrells M.E. (2011a) Genomic selection accuracy for grain quality traits in biparental wheat populations. Crop Sci 51:2597-2606.

Heffner EL, Jannink J-L, Sorrells ME (2011b) Genomic Selection Accuracy using Multifamily Prediction Models in a Wheat Breeding Program. Plant Genome 4 (1):65-75. doi:10.3835/plantgenome2010.12.0029.

Heffner EL, Lorenz AJ, Jannink J-L, Sorrells ME (2010) Plant breeding with genomic selection: gain per unit time and cost. Crop science 50 (5):1681-1690

Heffner E.L., Sorrells M.E., Jannink J.-L. (2009) Genomic selection for crop improvement. Crop Sci 49:1-12.

Hernández-Bautista A., Lobato-Ortiz R., García-Zavala J.J., Parra-Gómez M.A., Cadeza-Espinosa M., Canela-Doñan D., Cruz-Izquierdo S., Chávez-Servia J.L. (2016) Implications of genomic selection for obtaining F 2: 3 families of tomato. Sci Hort 207:7-13.

Hickey JM, Dreisigacker S, Crossa J, Hearne S, Babu R, Prasanna BM, Grondona M, Zambelli A, Windhausen VS, Mathews K (2014) Evaluation of genomic selection training population designs and genotyping strategies in plant breeding programs using simulation. Crop Sci 54 (4):1476-1488.

Hoffstetter A., Cabrera A., Huang M., Sneller C. (2016) Optimizing Training Population Data and Validation of Genomic Selection for Economic Traits in Soft Winter Wheat. G3 (Bethesda). 116.032532.

Huang M., Cabrera A., Hoffstetter A., Griffey C., Van Sanford D., Costa J., McKendry A., Chao S., Sneller C. (2016) Genomic selection for wheat traits and trait stability. Theor Appl Genet. 1-14.

Lipka A.E., Tian F., Wang Q., Peiffer J., Li M., Bradbury P.J., Gore M.A., Buckler E.S., Zhang Z. (2012) GAPIT: genome association and prediction integrated tool. Bioinformatics 28:2397-2399. DOI: 10.1093/bioinformatics/bts444.

Lorenz, A.J. and K.P. Smith. (2015) Adding genetically distant individuals to training populations reduces genomic prediction accuracy in barley. Crop Sci 55: 2657-2667.

Lorenzana R.E., Bernardo R. (2009) Accuracy of genotypic value predictions for marker-based selection in biparental plant populations. Theor Appl Genet 120:151-161.

Lozada, D.N., Mason, E. Babar, M. A., Carver, B., Brown-Guedira, G., Merrill K., Arguello, N., Acuna, A., Vieira, L., Holder, A., Addison, C., Moon, D., Miller, R., Dreisigacker, S. (2017) Association mapping reveals loci associated with multiple traits that affect grain yield and adaptation in soft winter wheat. Euphytica 213:213-222. DOI: 10.1007/s10681-017-2005-2

Mason RE, Addison CK, Babar A, Acuna A, Lozada D, Subramanian N, Arguello MN, Miller RG, Brown-Guedira G, Guedira M, Johnson J (2017) Diagnostic Markers for Vernalization and Photoperiod Loci Improve Genomic Selection for Grain Yield and Spectral Reflectance in Wheat. Crop Sci. doi:10.2135/cropsci2017.06.0348.

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

McKendry, A.L., D.N. Tague, R.L. Wright, and J.A. Tremain. 2007. Registration of ‘Bess’ wheat. J Plant Reg. 1:21–23. doi:10.3198/jpr2006.06.0405crc.

Muleta KT, Bulli P, Zhang Z, Chen X, Pumphrey M (2017) Unlocking Diversity in Germplasm Collections via Genomic Selection: A Case Study Based on Quantitative Adult Plant Resistance to Stripe Rust in Spring Wheat. The Plant Genome 10 (3). doi:10.3835/plantgenome2016.12.0124.

Muranty H., Troggio M., Sadok I.B., Rifaï M.A., Auwerkerken A., Banchi E., Velasco R., Stevanato P., van de Weg W.E., Di Guardo M., Kumar S., Laurens F., Bink M.C.A.M. (2015) Accuracy and responses of genomic selection on key traits in apple breeding. Hort Res 2:15060. DOI: 10.1038/hortres.2015.60.

Murphy, J.P., R.A. Navarro, S. Leath, D.T. Bowman, P.R. Weisz, L.G. Ambrose, et al. 2004. Registration of 'NC-Neuse' wheat. Crop Sci. 44:1479–1480. doi:10.2135/cropsci2004.1479

Nakaya A., and S.N. Isobe (2012) Will genomic selection be a practical method for plant breeding? Ann Bot:mcs109.

Petersen S, Lyerly JH, McKendry AL, Islam MS, Brown-Guedira G, Cowger C, Dong Y, Murphy JP (2017) Validation of Fusarium Head Blight Resistance QTL in US Winter Wheat. Crop Sci 57 (1):1-12

Onogi A., Watanabe M., Mochizuki T., Hayashi T., Nakagawa H., Hasegawa T., Iwata H. (2016) Toward integration of genomic selection with crop modelling: the development of an integrated approach to predicting rice heading dates. Theor Appl Genet 129:805-817.

Patel D., Zander M., Dalton-Morgan J., Batley J. (2015) Advances in Plant Genotyping: Where the Future Will Take Us, in: J. Batley (Ed.), Plant Genotyping, Springer New York. 1-11.

Poland, J., J. Endelman, J. Dawson, J. Rutkoski, S. Wu, Y. Manes, Dreisigacker S, Crossa J, Sachex-Villeda, H, Sorrells M, Jannink J-L. (2012). Genomic Selection in Wheat Breeding using Genotyping-by-Sequencing. Plant Genome 5: 103-113. doi:10.3835/plantgenome2012.06.0006.

R Development Core Team. (2010) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Rutkoski J., Singh R., Huerta-Espino J., Bhavani S., Poland J., Jannink J.-L., Sorrells M. (2015) Efficient use of historical data for genomic selection: a case study of stem rust resistance in wheat. Plant Genome 8. Doi: 10.3835/plantgenome2014.09.0046.

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

Rutkoski JE, Poland JA, Singh RP, Huerta-Espino J, Bhavani S, Barbier H, Rouse MN, Jannink J-L, Sorrells ME (2014) Genomic Selection for Quantitative Adult Plant Stem Rust Resistance in Wheat. Plant Genome 7 (3). doi:10.3835/plantgenome2014.02.0006.

Sallam AH, Smith KP (2016) Genomic Selection Performs Similarly to Phenotypic Selection in Barley. Crop Sci 56 (6):2871-2881. doi:10.2135/cropsci2015.09.0557

SAS Institute. (2011) SAS System Options: Reference, 2nd ed. SAS Institute, Cary, NC.

Spindel J., Begum H., Akdemir D., Virk P., Collard B., Redoña E., Atlin G., Jannink J.-L., 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:e1004982.

Wang S, Wong D, Forrest K, Allen A, Chao S, Huang BE, Maccaferri M, Salvi S, Milner SG, Cattivelli L, et al. (2014) Characterization of polyploid wheat genomic diversity using a high-density 90 000 single nucleotide polymorphism array. Plant Biotech J 12 (6):787-796. doi:10.1111/pbi.12183

Würschum T, Reif JC, Kraft T, Janssen G, Zhao Y (2013) Genomic selection in sugar beet breeding populations. BMC Genetics 14 (1):85

Xavier A, Muir WM, Rainey KM (2016) Assessing Predictive Properties of Genome-Wide Selection in Soybeans. G3. 6 (8):2611-2616. doi:10.1534/g3.116.032268.

Zhao Y., Gowda M., Liu W., Würschum T., Maurer H.P., Longin F.H., Ranc N., Reif J.C. (2012) Accuracy of genomic selection in European maize elite breeding populations. Theor Appl Genet 124:769-776.

Zhong S., Dekkers J.C., Fernando R.L., Jannink J.-L. (2009) Factors affecting accuracy from genomic selection in populations derived from multiple inbred lines: a barley case study. Genetics 182:355-364.

Table 1. Descriptive statistics and heritability values for each of the grain yield datasets used for genome-wide predictions.

Population	No. of lines	Dataset [a]	No. of environments	Mean (t ha^{-1})	Min	Max	h^2 [b]
TP	239	ABLUP	8	3.10	0.07	7.14	0.48
		BLUP14	2	2.91	0.37	6.49	0.40
		BLUP15	6	3.31	0.07	7.60	0.80
		NBLUP	4	3.32	0.07	7.14	0.61
		SBLUP	4	2.88	0.37	5.66	0.60
NB	100	NB_ALL	5	3.63	0.03	7.49	0.70
		NB_FAY	3	4.38	1.04	7.49	0.70
		NB_NPT	2	2.82	0.03	5.91	0.45
PA-RIL[c]	156	PA_ALL	12	4.40	1.86	6.25	0.33
		PA_Cluster1	3	4.09	3.34	4.81	0.50
		PA_Cluster2	2	4.69	3.34	5.69	0.63
		PA_Cluster3	2	5.56	1.47	7.41	0.85
		PA_Cluster4	5	4.00	2.81	4.98	0.66

[a]*ABLUP*- BLUP across all environments for the CBL; *BLUP14*- BLUP across 2014 site-years; *BLUP15*- BLUP across 2015 site-years; *NBLUP*- BLUP across northern environments; *SBLUP*- BLUP across southern environments; *NB_ALL*- BLUP across all site-years for the NB; *NB_FAY*- BLUP across Fayetteville site-years (FAY15, FAY16, FAY17); *NB_NPT*- BLUP across Newport site-years (NPT16, NPT17); *PA_ALL* represents 12 site-years for the PA-RIL; *PA_Cluster1* includes site-years FAY12, STU12, and FAY14; *PA_Cluster2* includes FAY13 and MAR14; *PA_Cluster3* includes GA12 and GA13; *PA_Cluster4* includes TX12, TX13, MAR13, and STU13

[b] Broad sense heritability; calculated using the formula: $h^2 = \frac{\sigma_G^2}{\sigma_G^2 + \frac{\sigma_{GEI}^2}{e} + \frac{\sigma_E^2}{er}}$

[c] Results adapted from Mason et al. (2017)

Table 2. Response to selection, *R* for grain yield in the training population across different selection strategies.

Selection strategy	**GY (t ha^{-1}) ± SD**	**σ^2**	**Selection differential, *S*** [a]	**Response to selection, *R*** [b]	**% change relative to PS**
GS	3.61 ± 0.34	0.12	0.44	0.21	-32.3
MAS	3.34 ± 0.36	0.13	0.17	0.08	-74.2
PS	3.82 ± 0.16	0.03	0.65	0.31	-
RS	3.19 ± 0.36	0.13	0.02	0.01	-96.8
PS + GS	3.88 ± 0.18	0.03	0.71	0.34	9.7

GS- genomic selection; *MAS*- marker assisted selection; *PS*- phenotypic selection; *RS*- random selection

[a] $S = \mu_{sel} - \mu_{pop}$; $\mu_{pop} = 3.17$ t ha^{-1}

[b] Calculated as $R = h^2 S$ where h^2 is heritability for GY based on published value in Lozada et al., (2017; published, Chapter II); equal to 0.48

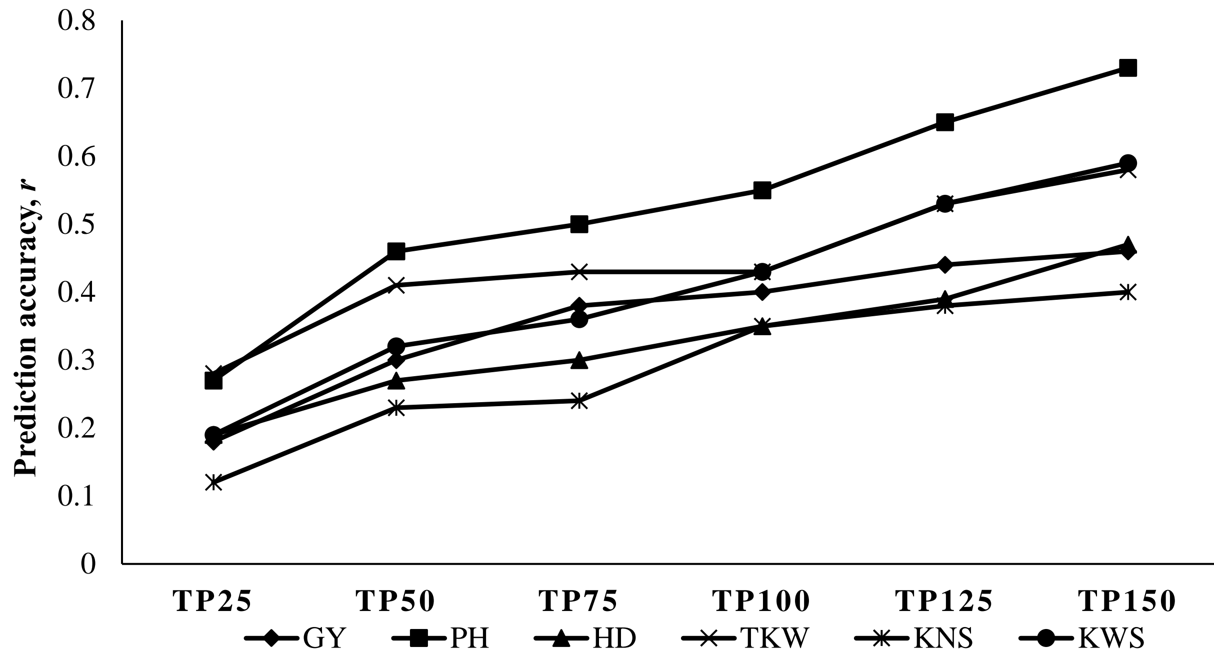


Figure 1. Effect of training population (TP) size on accuracy of genomic selection for GY, yield components and agronomic traits; rrBLUP model, 10-fold CV for BLUP across all environments (ABLUP) at a constant validation population (VP) size (N=60). *GY*- grain yield; *PH*- plant height; *HD*- heading date; *TKW*- thousand kernel weight; *KNS*- kernel number per spike; *KWS*- kernel weight per spike

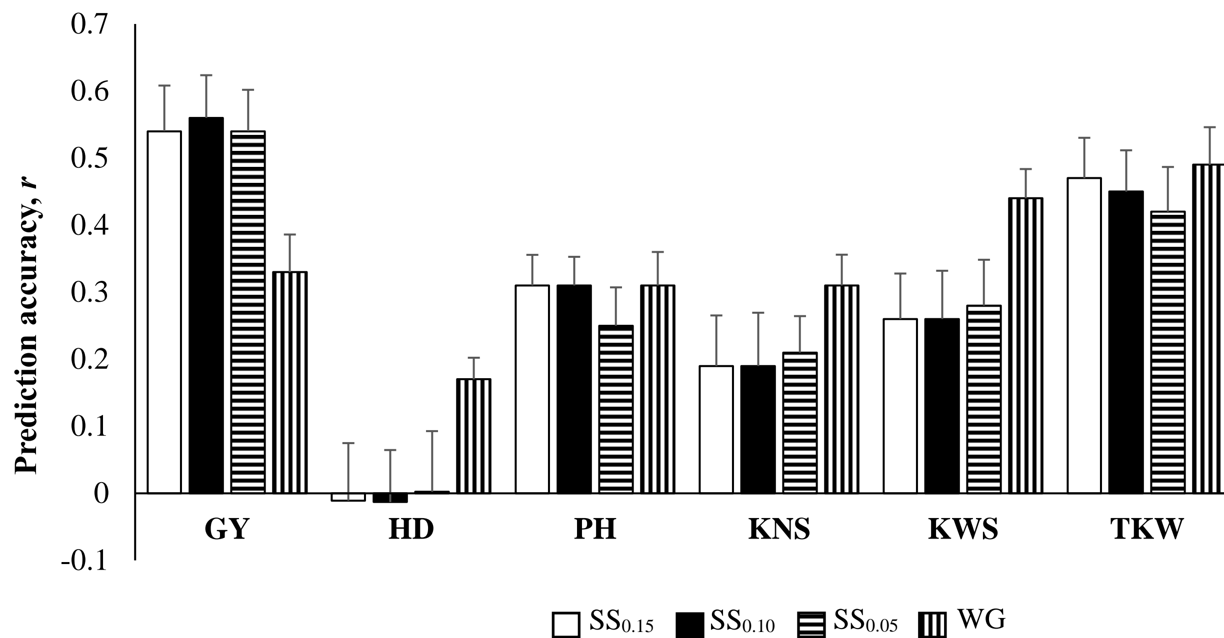


Figure 2. Effect of marker number on the accuracy of genome-wide prediction for GY, yield components and agronomic traits under an rrBLUP model, 10-fold CV for BLUP across all environments (ABLUP); *GY*- grain yield; *PH*- plant height; *HD*- heading date; *TKW*- thousand kernel weight; *KNS*- kernel number per spike; *KWS*- kernel weight per spike. $SS_{0.15}$- marker subset based on significance level $p < 0.15$ (~820 SNPs); $SS_{0.10}$- marker subset based on significance level $p < 0.10$ (~540 SNPs); $SS_{0.05}$- marker subset based on significance level $p < 0.05$ (~270 SNPs); *WG*- whole genome marker data (~5,600 SNPs). Bars indicate standard errors

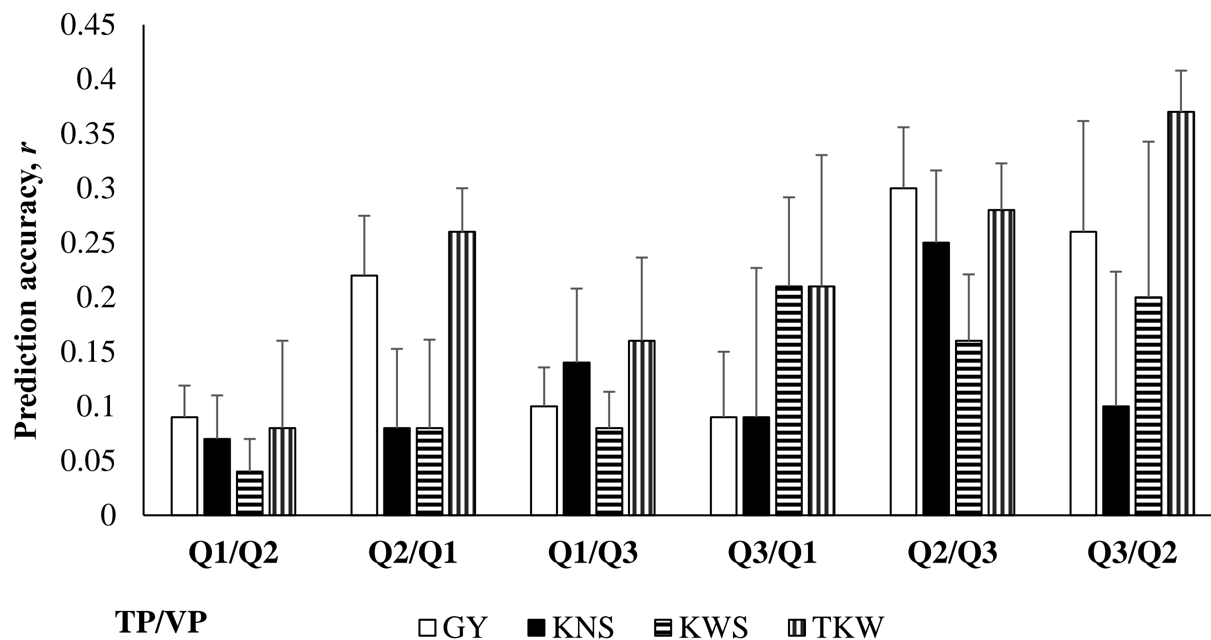


Figure 3. Effect of using different subgroups, *Q* as training population to predict grain yield and yield components for other subgroups. *Q* groupings based on STRUCTURE analyses. Predictions performed using a constant TP and VP sizes of 50 and 30, respectively under a 10-fold CV in RRBLUP

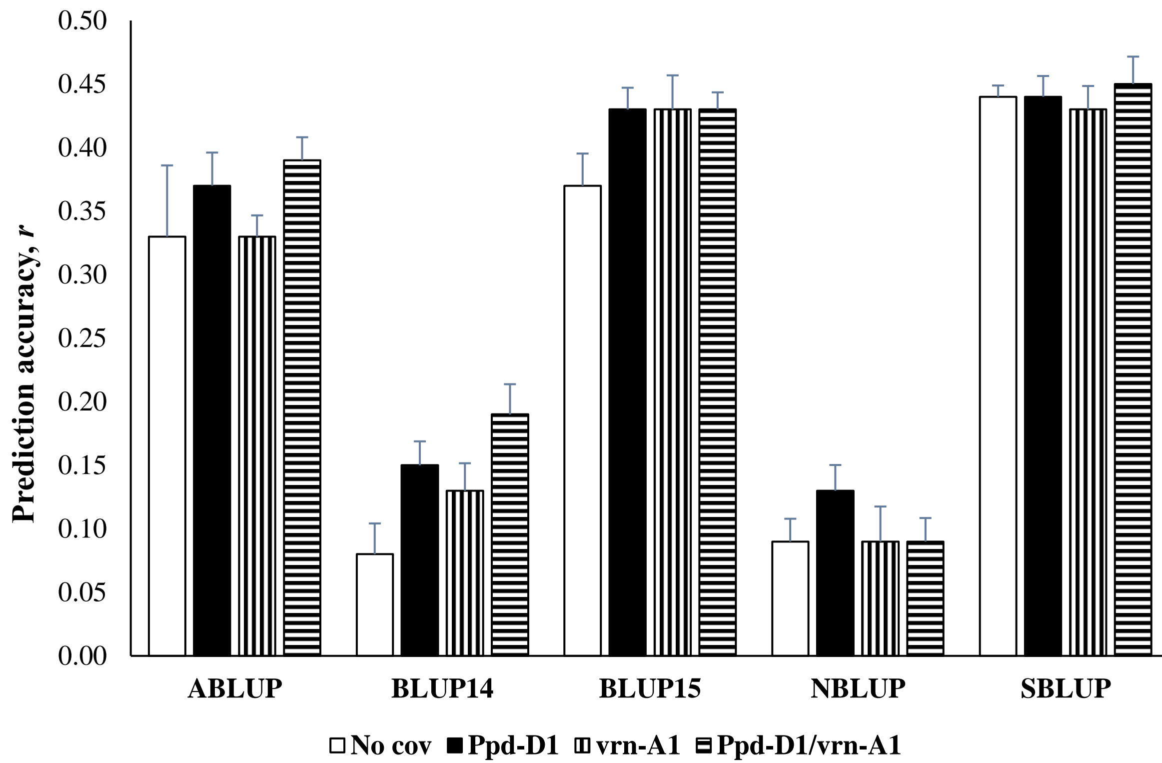


Figure 4. Effect of adding covariates to the predictability of GY across different datasets. Predictions were done using a standard single population CV; TP size= 144. *ABLUP*- BLUP across all environments; *BLUP14*- BLUP across 2014 environments; *BLUP15*- BLUP across all 2015 environments; *NBLUP*- BLUP across Northern environments; *SBLUP*- BLUP across southern environments. Bars indicate standard error

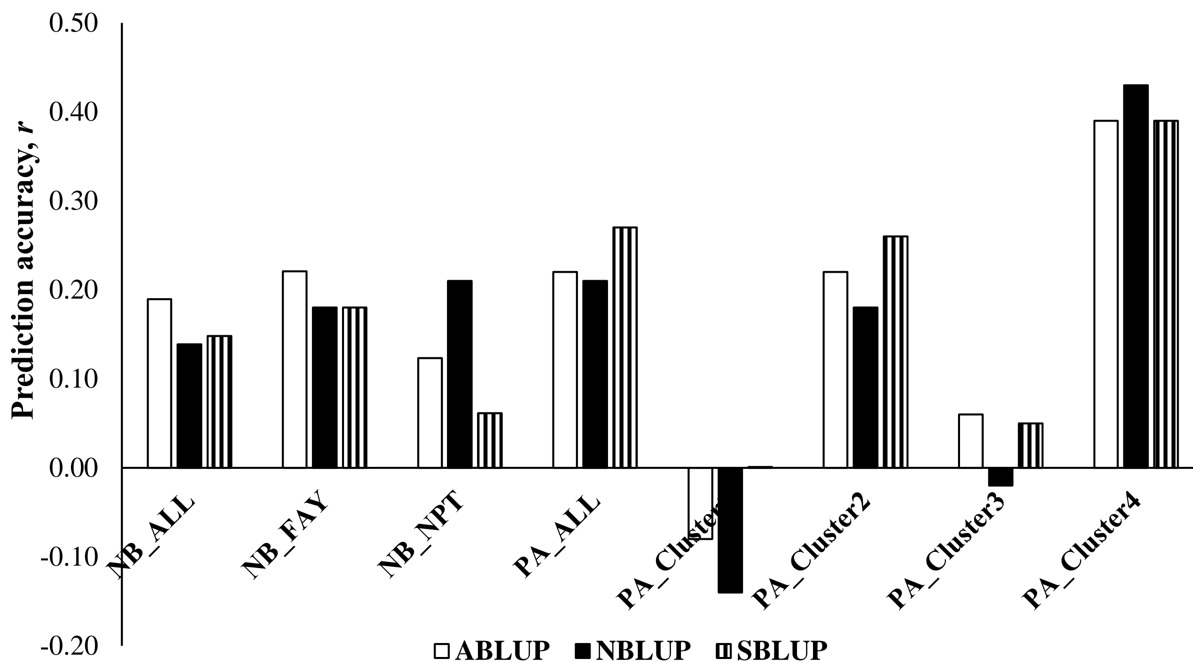


Figure 5. Accuracy for genomic selection using TP (*N*=239; ABLUP, NBLUP, and SBLUP datasets) to predict GY in the NB (*N*=100) and PA-RIL (*N*=156) across different site years and clusters. Predictions for the NB and PA-RIL were conducted with 1,089 and 1,632 SNP markers, respectively. *NB_ALL*- BLUP across all site-years for the NB; *NB_FAY*- BLUP across Fayetteville site-years (FAY15, FAY16, FAY17); *NB_NPT*- BLUP across Newport site-years (NPT16, NPT17); *PA_ALL* represents 12 site-years for the PA-RIL; *PA_Cluster1* includes site-years FAY12, STU12, and FAY14; *PA_Cluster2* includes FAY13 and MAR14; *PA_Cluster3* includes GA12 and GA13; *PA_Cluster4* includes TX12, TX13, MAR13, and STU13

CONCLUSIONS

Over-all conclusions

A genome-wide association study and genomic selection for grain yield and agronomic traits in soft winter wheat was conducted. Loci associated with multiple yield-related traits were identified in different genomic regions and showed potential to be used for marker-assisted breeding towards grain yield improvement in winter wheat. Validation of these yield-QTL using a spring wheat panel from CIMMYT, Mexico confirmed their effects and showed that different classes of wheat share common QTL which could also be exploited for marker-assisted selection. Candidate gene functions on the validated loci in chromosomes 3A, 4B, and 6B demonstrated the genetic complexity of grain yield. Effects of training population size, number of markers, relatedness, and covariates in the genomic prediction model on genomic selection accuracy were shown. Forward validation of selection model using winter wheat cultivars and breeding lines to predict grain yield in related biparental populations demonstrated the feasibility of this genome-wide selection approach to predict traits with complex genetic architecture. Combining phenotypic and genomic selection resulted to the highest response for grain yield and showed the ability of complementing these strategies to increase genetic gains and accelerate improvement in wheat breeding programs. Results of this study provide additional insights in the genetic complexity of grain yield and component traits and can be used to accelerate genetic improvement across different classes of wheat.

APPENDICES

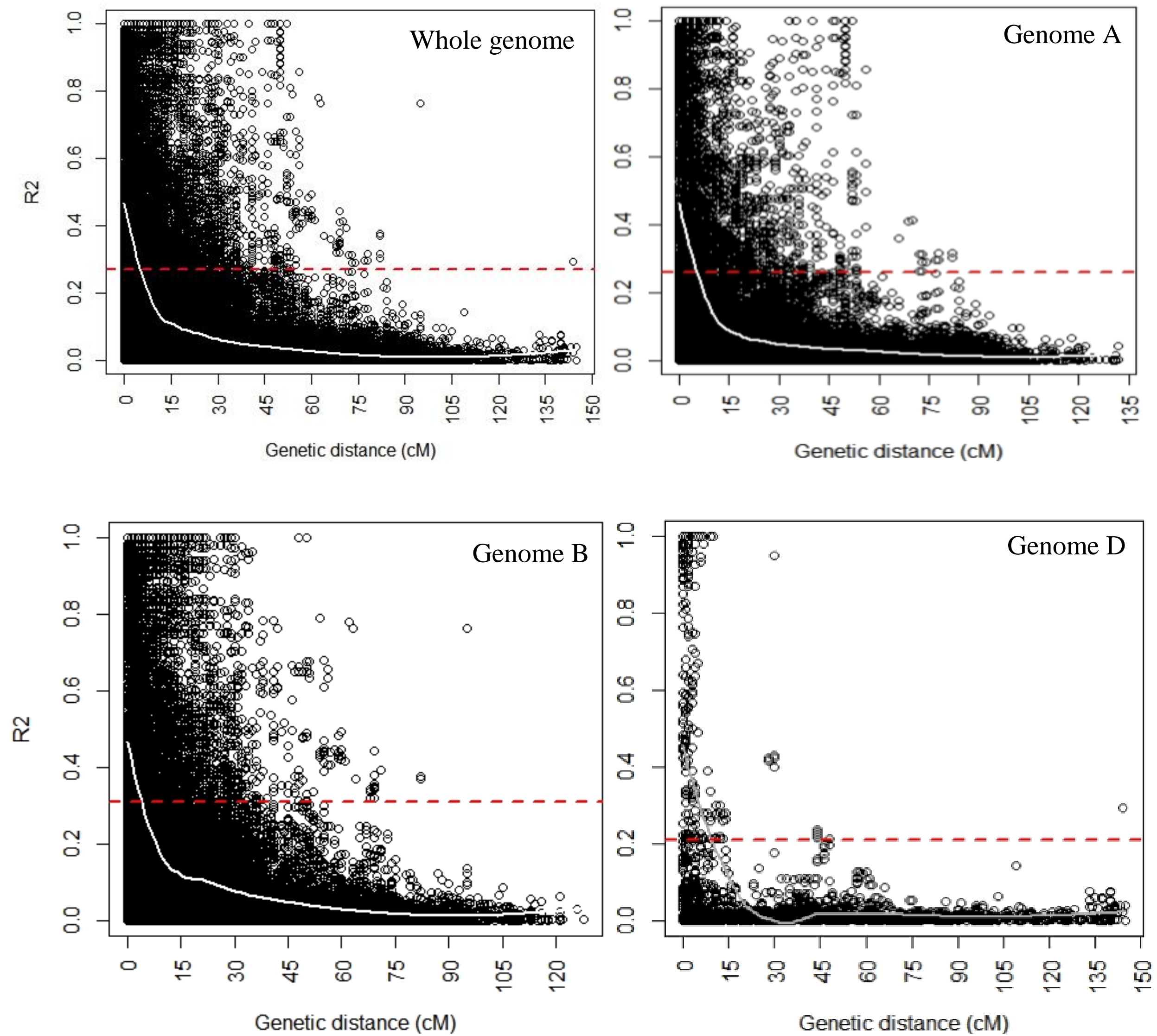
Appendix 1. Summary of linkage disequilibrium analyses for intrachromosomal marker pairs for the soft red winter wheat association mapping panel

Genome	**Total pairs**	**Mean r^2 for all pairs**	**No. of sig pairs[a]**	**Sig pairs (%)**	**Ave dist. of pairs in sig. LD (cM)**	**No. of linked pairs[b]**	**Linked pairs (%)**	**Mean r^2 for linked pairs**	**No. of pairs in comp. LD**	**Ave dist. of pairs in comp. LD (cM)**
Whole genome	153,600	0.16	74, 822	48.71	14.40	71, 800	46.74	0.33	6, 087	1.71
Genome A	73, 475	0.15	31, 979	435.52	14.70	61, 259	83.07	0.17	2, 485	1.34
Genome B	76, 125	0.17	41, 606	54.65	14.10	68, 063	89.41	0.19	3, 410	1.90
Genome D	4,000	0.20	1, 237	30.93	16.60	2, 643	66.08	0.02	192	3.14

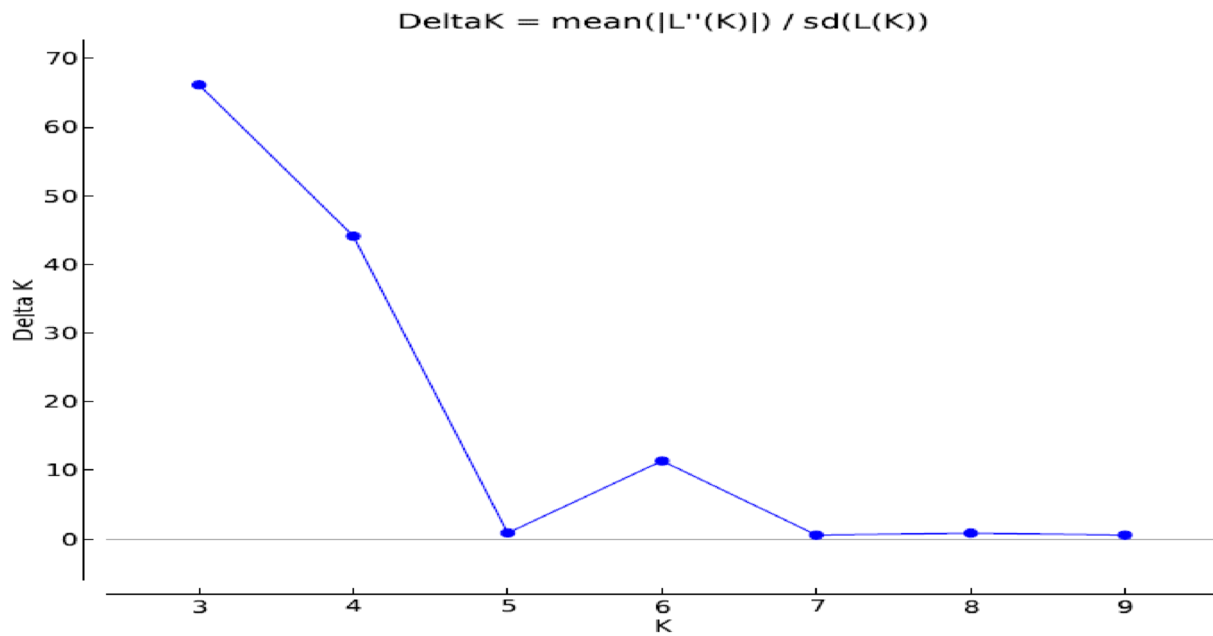
[a] Significant marker pairs, $p < 0.005$

[b] Physically linked pairs are those with genetic distance of ≤ 50 cM

[c] Marker pairs with r^2 value equal to 1.0 were regarded to be in complete LD



Appendix 2. Plot of intrachromosomal linkage disequilibrium (LD) represented by the square of correlation between alleles, r^2 against genetic distance (cM) showing LD decay with increasing distances among pairs of marker loci for the whole and individual sub-genomes. Curve shows the fitted second degree LOESS while dashed line represents the critical value beyond which LD is likely caused by physical linkage (equivalent to r^2= 0.269 (whole genome), 0.258 (Genome A), 0.309 (Genome B), and 0.210 (Genome D)), taken as 95th percentile of r^2 values for unlinked marker pairs (distance > 50 cM)



Appendix 3. Inference for the true number of subpopulations, *K* using the Evanno method showing K=3 as the optimum number of subgroup in the association mapping panel

Appendix 4. Inferred subgroup designation of the soft winter wheat lines based on Q values from STRUCTURE software

Entry	Variety name	*Q1*	*Q2*	*Q3*	Inferred group
AMP001	001169-7E15	0.56	0.246	0.194	Group 1
AMP002	01063-1-3-6-2-G2	0.194	0.726	0.079	Group 2
AMP003	011124-1-42-13	0.427	0.43	0.143	Group 2
AMP004	011388-8-4-5	0.246	0.15	0.604	Group 3
AMP005	031086-44-4-2	0.736	0.051	0.212	Group 1
AMP006	051336-B-B-1	0.671	0.003	0.326	Group 1
AMP007	071628-G3-G1-G4-G1	0	0.999	0	Group 2
AMP008	071694-G5-G5-G1pub	0.034	0.903	0.063	Group 2
AMP009	081515-G1-G2	0.17	0.788	0.042	Group 2
AMP010	09283-G1-G1	0.07	0.699	0.23	Group 2
AMP011	222-22-5	0.09	0.001	0.909	Group 3
AMP012	991227-6A33	0.104	0.713	0.183	Group 2
AMP013	991371-6E12	0.013	0.986	0	Group 2
AMP014	AG_2020	0.251	0.076	0.673	Group 3
AMP015	AGS_2000_JJ	0.003	0.997	0	Group 2
AMP016	AGS_2010	0.547	0.086	0.367	Group 1
AMP017	AGS_2020	0.124	0.793	0.083	Group 2
AMP018	AGS_2026	0.886	0.001	0.113	Group 1
AMP019	AGS_2031	0.411	0.118	0.47	Group 3
AMP020	AGS_2035	0	1	0	Group 2
AMP021	AGS_2060	0.16	0.276	0.564	Group 3
AMP022	AGS_2485	0.586	0.413	0.001	Group 1
AMP023	AGS_CL7	0.002	0.85	0.149	Group 2
AMP024	AR00255-16-1	0.181	0.443	0.377	Group 2
AMP025	AR00343-5-1	0.262	0.224	0.515	Group 3
AMP026	AR01039-4-1	0.084	0.532	0.384	Group 2
AMP027	AR01040-4-1	0.048	0.374	0.578	Group 3
AMP028	AR01044-1-1	0.043	0.452	0.505	Group 3
AMP029	AR01156-2-1	0.27	0.322	0.407	Group 3
AMP030	AR01163-3-1	0.253	0.147	0.6	Group 3
AMP031	AR01167-3-1	0.162	0.341	0.497	Group 3
AMP032	AR01179-4-1	0.069	0.424	0.507	Group 3
AMP033	AR01209-2-1	0.257	0.341	0.402	Group 3
AMP034	AR02061-1-1	0.247	0.326	0.427	Group 3
AMP035	AR910	0.065	0.175	0.759	Group 3
AMP036	AR97124-4-3	0.065	0.243	0.692	Group 3
AMP037	ARS05-0074	0.332	0.027	0.641	Group 3
AMP038	ARS05-0241	0.154	0.143	0.703	Group 3

Appendix 4. (Cont.)

Entry	Variety name	*Q1*	*Q2*	*Q3*	Inferred group
AMP039	ARS05-0401	0.006	0.297	0.697	Group 3
AMP040	ARS07-0203	0.353	0.07	0.578	Group 3
AMP041	ARS07-0404	0.534	0.012	0.454	Group 1
AMP042	ARS07-0558	0.001	0.001	0.998	Group 3
AMP043	ARS07-0815	0.006	0.003	0.99	Group 3
AMP044	ARS07-0912	0.089	0.184	0.727	Group 3
AMP045	ARS07-1208	0.014	0.156	0.83	Group 3
AMP046	ARS08-0111	0.101	0.502	0.397	Group 2
AMP047	ARS09-776	0.25	0.073	0.677	Group 3
AMP048	Arthur_CG	0.001	0.001	0.998	Group 3
AMP049	Baldwin	0	1	0	Group 2
AMP050	Blueboy_JJ	0.099	0.102	0.799	Group 3
AMP051	Boone	0.04	0.886	0.074	Group 2
AMP052	Branson	0.117	0.196	0.687	Group 3
AMP053	Caldwell	0.001	0.121	0.878	Group 3
AMP054	Chancellor	0.288	0.163	0.549	Group 3
AMP055	Chesapeake	0.497	0.166	0.336	Group 1
AMP056	Clark	0.19	0.21	0.6	Group 3
AMP057	Clemson_201	0.445	0.21	0.345	Group 1
AMP058	Coker_65-20	0.757	0.11	0.133	Group 1
AMP059	Coker_68-15_PM	0.948	0.052	0	Group 1
AMP060	Coker_747_CG	0.26	0.001	0.74	Group 3
AMP061	Coker_762	0.488	0.075	0.437	Group 1
AMP062	Coker_797_JJ	0.919	0.034	0.046	Group 1
AMP063	Coker_9134_CG	1	0	0	Group 1
AMP064	Coker_9134_Syn	1	0	0	Group 1
AMP065	Coker_9152	0.301	0.514	0.185	Group 2
AMP066	Coker_916_JJ	0.122	0.091	0.787	Group 3
AMP067	Coker_9375	0.679	0.008	0.313	Group 1
AMP068	Coker_9553	0.303	0.291	0.405	Group 3
AMP069	Coker_9663_Syn	0.115	0.3	0.585	Group 3
AMP070	Coker_9766	0.333	0.111	0.556	Group 3
AMP071	Coker_9803_CG	0.556	0	0.444	Group 1
AMP072	Coker_9835_PM	0.346	0.344	0.31	Group 1
AMP073	Delta_King_GR9108	0.085	0.171	0.744	Group 3
AMP074	Dominion	0.427	0.001	0.573	Group 3
AMP075	Doublecrop	0	0	0.999	Group 3
AMP076	Elkhart	0.162	0.397	0.441	Group 3
AMP077	Ernie_CS	0.085	0.067	0.848	Group 3
AMP078	FG95195	0.395	0.387	0.217	Group 1

Appendix 4 (Cont.)

Entry	Variety name	*Q1*	*Q2*	*Q3*	Inferred group
AMP079	FL_302_JJ	0.288	0.643	0.069	Group 2
AMP080	Flint	0.345	0.138	0.517	Group 3
AMP081	GA_1123	0.35	0.131	0.519	Group 3
AMP082	GA00067-8E35	0.53	0.359	0.111	Group 1
AMP083	GA001138-8E36	0.267	0.451	0.282	Group 2
AMP084	GA001142-9E23	0.204	0.503	0.293	Group 2
AMP085	GA001170-7E26	0.222	0.778	0	Group 2
AMP086	GA011493-8E18	0.003	0.996	0	Group 2
AMP087	GA021245-9E16	0.392	0.496	0.111	Group 2
AMP088	GA021338-9E15	0.299	0.701	0.001	Group 2
AMP089	GA031238-7E34	0.548	0.004	0.448	Group 1
AMP090	GA971127#1	0.163	0.193	0.644	Group 3
AMP091	Gore_JJ	0.968	0.031	0.001	Group 1
AMP092	Hazen	0.55	0.416	0.034	Group 1
AMP093	Holley	0.112	0.247	0.641	Group 3
AMP094	Hunter	0.635	0.001	0.364	Group 1
AMP095	IL00-8633	0.148	0.002	0.85	Group 3
AMP096	IL00-8641	0.001	0.083	0.916	Group 3
AMP097	IL05-4236	0.055	0.352	0.594	Group 3
AMP098	IL06-13721	0.134	0.003	0.864	Group 3
AMP099	IL06-23571	0.217	0.276	0.507	Group 3
AMP100	IL08-24578	0.173	0.08	0.748	Group 3
AMP101	IL96-6472	0.001	0.123	0.877	Group 3
AMP102	INW0304	0.001	0.001	0.998	Group 3
AMP103	Jackson_CG	0.831	0.079	0.09	Group 1
AMP104	Jamestown_PM	0.535	0.181	0.284	Group 1
AMP105	Jaypee_CS	0.225	0.182	0.593	Group 3
AMP106	Keiser	0.127	0.124	0.749	Group 3
AMP107	Knox_62	0.13	0.15	0.72	Group 3
AMP108	Kristy	0.127	0.151	0.723	Group 3
AMP109	KY02C-1043-04	0.298	0.272	0.43	Group 3
AMP110	KY02C-1058-03	0.386	0.452	0.162	Group 2
AMP111	KY02C-1076-07	0.21	0.658	0.132	Group 2
AMP112	KY02C-1121-11	0.081	0.443	0.477	Group 3
AMP113	KY02C-2215-02	0.137	0.23	0.633	Group 3
AMP114	KY03C-1002-02	0.262	0.347	0.391	Group 3
AMP115	KY03C-1237-39	0.238	0.365	0.396	Group 3
AMP116	LA01069D-23-4-4	0.001	0.65	0.35	Group 2
AMP117	LA0110D-150	0.453	0.457	0.09	Group 2
AMP118	LA01139D-56-1	0.068	0.387	0.545	Group 3

Appendix 4 (Cont.)

Entry	Variety name	*Q1*	*Q2*	*Q3*	Inferred group
AMP119	LA01164D-94-2-B	0.146	0.179	0.675	Group 3
AMP120	LA02015E201	0.369	0.207	0.424	Group 3
AMP121	LA02015E42	0.503	0.179	0.317	Group 1
AMP122	LA02015E58	0.373	0.209	0.418	Group 3
AMP123	LA02024E12	0.435	0.302	0.263	Group 1
AMP124	LA02024E7	0.459	0.311	0.231	Group 1
AMP125	LA03012E-27	0.004	0.699	0.297	Group 2
AMP126	LA03118E117	0.139	0.431	0.429	Group 2
AMP127	LA03136E71	0.125	0.273	0.602	Group 3
AMP128	LA03148E12	0.148	0.052	0.801	Group 3
AMP129	LA03155D-P13	0.171	0.333	0.496	Group 3
AMP130	LA03161D-P1	0.126	0.35	0.524	Group 3
AMP131	LA03217D-P2	0.001	0.536	0.463	Group 2
AMP132	LA03217E2	0.001	0.608	0.391	Group 2
AMP133	LA04013D-142	0.196	0.513	0.291	Group 2
AMP134	LA04041D-10	0.048	0.95	0.002	Group 2
AMP135	LA821	0.443	0.508	0.049	Group 2
AMP136	LA841	0.584	0.273	0.142	Group 1
AMP137	LA95135	0.324	0.526	0.151	Group 2
AMP138	LA97113UC-124	0	0.999	0	Group 2
AMP139	Madison_CS	0.043	0.086	0.87	Group 3
AMP140	MAGNOLIA	0.234	0.32	0.445	Group 3
AMP141	Mallard	0.005	0.279	0.716	Group 3
AMP142	Massey_CG	0.071	0.001	0.928	Group 3
AMP143	McCormick	0.119	0.001	0.88	Group 3
AMP144	McNair_1813	0.009	0.048	0.943	Group 3
AMP145	McNair_701	0.326	0.119	0.555	Group 3
AMP146	MD00W16-07-3	0.25	0.31	0.439	Group 3
AMP147	MD01W28-08-11	0.287	0.507	0.206	Group 2
AMP148	MD99W64-05-11	0.201	0.445	0.355	Group 2
AMP149	Merl	0.249	0.284	0.467	Group 3
AMP150	MO_011126	0.186	0.665	0.15	Group 2
AMP151	MO_080104	0.228	0.146	0.626	Group 3
AMP152	MO_081652	0.235	0.146	0.619	Group 3
AMP153	MO_980525	0.026	0.223	0.751	Group 3
AMP154	MPV_57_CG	0.237	0.236	0.527	Group 3
AMP155	NC06-19896	0.062	0.002	0.935	Group 3
AMP156	NC06-20401	0.417	0.038	0.545	Group 3
AMP157	NC06BGTAG12	0.619	0.001	0.38	Group 1
AMP158	NC07-22432	0.216	0.001	0.783	Group 3

Appendix 4 (Cont.)

Entry	Variety name	*Q1*	*Q2*	*Q3*	Inferred group
AMP159	NC07-23880	0.546	0.001	0.454	Group 1
AMP160	NC07-24445	0.368	0.001	0.631	Group 3
AMP161	NC07-25169	0.107	0.001	0.892	Group 3
AMP162	NC08-21273	0.141	0	0.858	Group 3
AMP163	NC08-23089	0.271	0.239	0.49	Group 3
AMP164	NC08-23090	0.239	0.264	0.497	Group 3
AMP165	NC08-23323	0.22	0.001	0.779	Group 3
AMP166	NC08-23324	0.225	0.001	0.775	Group 3
AMP167	NC08-23383	0.175	0.009	0.816	Group 3
AMP168	NC08-23925	0.926	0.001	0.073	Group 1
AMP169	NC09BGTS16	0.802	0.001	0.197	Group 1
AMP170	NC09BGTUM15	0.606	0.057	0.337	Group 1
AMP171	NC96BGTA4	0.949	0.001	0.05	Group 1
AMP172	NC96BGTA5	0.304	0.104	0.592	Group 3
AMP173	NC96BGTA6	1	0	0	Group 1
AMP174	NC96BGTD1	1	0	0	Group 1
AMP175	NC96BGTD2	1	0	0	Group 1
AMP176	NC96BGTD3	0.616	0.003	0.381	Group 1
AMP177	NC97BGTAB10	0.883	0	0.117	Group 1
AMP178	NC97BGTAB9	0.159	0.166	0.675	Group 3
AMP179	NC97BGTD7	0.693	0.117	0.191	Group 1
AMP180	NC97BGTD8	1	0	0	Group 1
AMP181	NC99BGTAG11	0.735	0.093	0.171	Group 1
AMP182	NC-Cape_Fear	0.478	0.152	0.37	Group 1
AMP183	NC-Neuse_PM	0.14	0	0.86	Group 3
AMP184	NC-Yadkin	0.137	0.001	0.862	Group 3
AMP185	Nelson	0.081	0.001	0.918	Group 3
AMP186	Oakes	0.294	0.044	0.662	Group 3
AMP187	Oasis	0.126	0.101	0.772	Group 3
AMP188	Oglethorpe	0.847	0.001	0.153	Group 1
AMP189	P03528A1-10	0.006	0.196	0.798	Group 3
AMP190	P0570A1-2	0.026	0.092	0.882	Group 3
AMP191	P07290A1-12	0.006	0.255	0.739	Group 3
AMP192	P99840C4-8	0.022	0.317	0.661	Group 3
AMP193	Panola	0.002	0.595	0.403	Group 2
AMP194	Pat	0.378	0.413	0.208	Group 2
AMP195	Pioneer_2548_CG	0.1	0.597	0.303	Group 2
AMP196	Pioneer_2555_CG	0.032	0.893	0.075	Group 2
AMP197	Pioneer_2568	0.144	0.627	0.229	Group 2
AMP198	Pioneer_2580_CG	0.002	0.588	0.41	Group 2

Appendix 4 (Cont.)

Entry	Variety name	*Q1*	*Q2*	*Q3*	Inferred group
AMP199	Pioneer_25W60	0.008	0.64	0.352	Group 2
AMP200	Pioneer_2643_CG	0.221	0.437	0.341	Group 2
AMP201	Pioneer_2684_CG	0.42	0.082	0.498	Group 3
AMP202	Pioneer_26R15	0.218	0.553	0.23	Group 2
AMP203	Pioneer_26R24_CG	0.992	0.008	0	Group 1
AMP204	Pioneer_26R31_CG	0.418	0.125	0.458	Group 3
AMP205	Pioneer_26R46_CG	0.371	0.629	0	Group 2
AMP206	Pioneer_26R61_JJ	0	0.999	0	Group 2
AMP207	Potomac_CG	0.182	0.235	0.584	Group 3
AMP208	Roane_CG	0.495	0.091	0.414	Group 1
AMP209	Rosen	0	0	0.999	Group 3
AMP210	Roy	0.214	0.114	0.672	Group 3
AMP211	Sabbe	0.117	0.224	0.659	Group 3
AMP212	Saluda_PM	1	0	0	Group 1
AMP213	Severn	0.38	0.001	0.619	Group 3
AMP214	Shirley_CG	0.394	0.136	0.47	Group 3
AMP215	Sisson_CG	0.245	0.063	0.692	Group 3
AMP216	SS_520	0.664	0.183	0.153	Group 1
AMP217	SS_5205	0.424	0.001	0.575	Group 3
AMP218	SS8641_JJ	0.629	0.004	0.367	Group 1
AMP219	Stacey	0.102	0.328	0.57	Group 3
AMP220	Tribute_CG	0.129	0.062	0.809	Group 3
AMP221	Tribute_PM	0.122	0.074	0.803	Group 3
AMP222	USG_3120	0.067	0.933	0	Group 2
AMP223	USG_3209_PM	0.372	0.006	0.621	Group 3
AMP224	USG_3295	0.431	0.113	0.457	Group 3
AMP225	USG_3555_JJ	0.401	0.019	0.58	Group 3
AMP226	USG_3592	0.87	0.068	0.062	Group 1
AMP227	VA_259	0.41	0.003	0.587	Group 3
AMP228	VA_90	0.484	0.15	0.367	Group 1
AMP229	VA_96W-247	0.266	0.06	0.674	Group 3
AMP230	VA00W-38	0.271	0.46	0.27	Group 2
AMP231	VA01W-21	0.508	0.164	0.327	Group 1
AMP232	VA01W713	0.421	0.18	0.398	Group 1
AMP233	VA03W-211	0.538	0.001	0.461	Group 1
AMP234	VA03W-235	0.523	0.258	0.219	Group 1

Appendix 4 (Cont.)

Entry	Variety name	*Q1*	*Q2*	*Q3*	Inferred group
AMP235	VA05W-139	0.62	0.21	0.17	Group 1
AMP236	VA05W-151	0.53	0.001	0.469	Group 1
AMP237	Wakefield_CG	0.047	0.14	0.813	Group 3
AMP238	Wakeland_CG	0.243	0.239	0.518	Group 3
AMP239	Wheeler_CG	0.191	0.001	0.808	Group 3

Appendix 5. Analysis of Molecular Variance (AMOVA) for the soft winter wheat association mapping panel under a ploidy independent allele model (ρ) tested using 999 permutations

Source of variation	d.f.	Mean squares	Variance components	% Variance
Within population	236	892.781	892.781	89.1
Among population	2	8855.489	109.222	10.9
Total	239			

Appendix 6. Subpopulation pairwise *Gst* and *Fst* values

Subgroup [+]	***Q1***	***Q2***	***Q3***
1	-		
2	0.17	-	
3	0.13	0.12	-
Fst	0.69	0.43	0.23

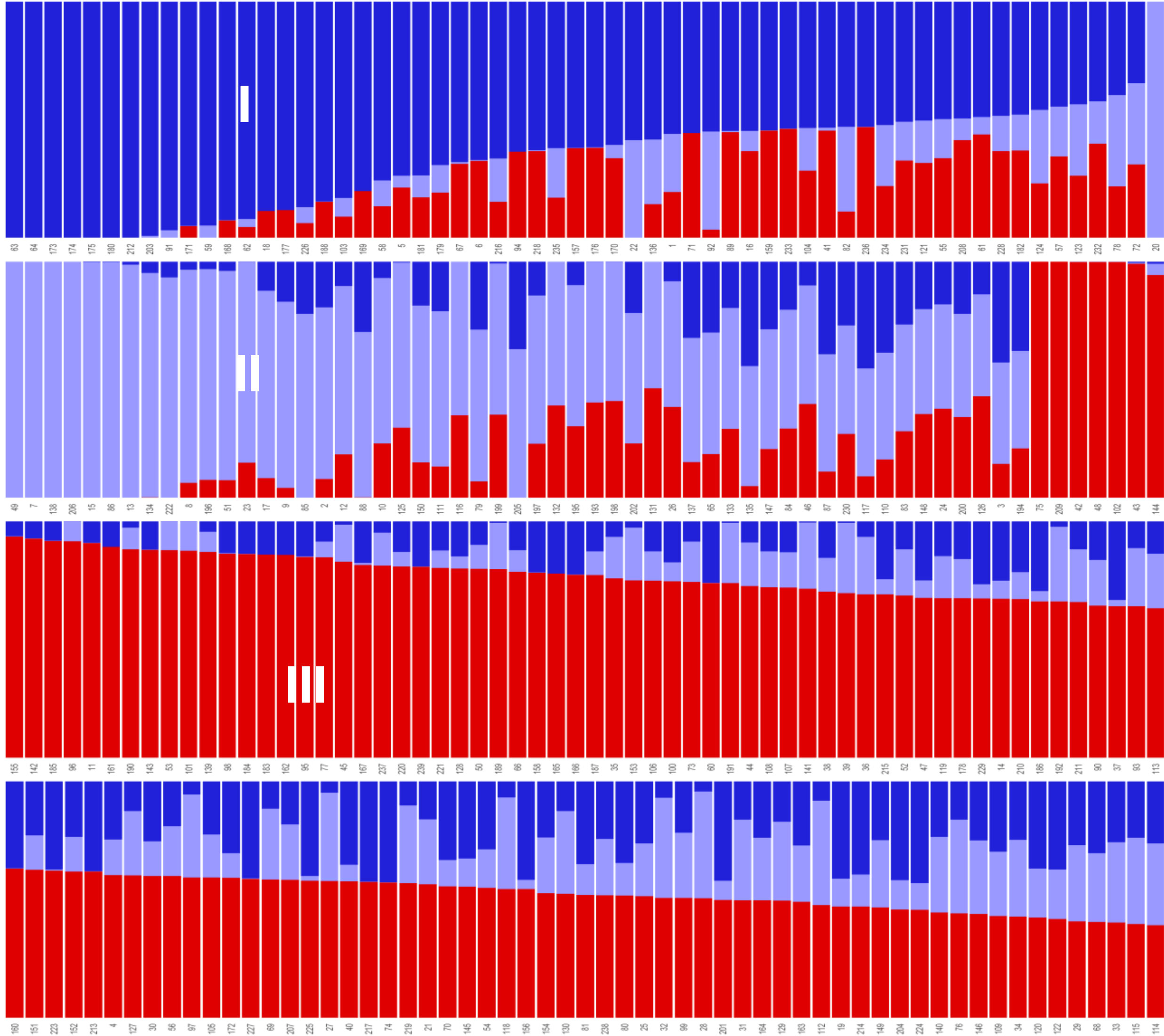
[+] Subgroups based on the values of coefficient of membership, *Q* inferred from STRUCTURE where individuals were assigned to a subpopulation based on the highest value of a corresponding *Q*

Appendix 7. Number of trait-specific and multi-trait markers associated with the measured phenotypes for the soft red winter wheat germplasm panel

Trait	No. of significant markers [a]	Chromosomes	R^2 [b]
Grain yield (GY)	15	1A, 2A, 2B, 3B, 4A, 4B, 5A, 7A	0.08-0.28
Heading date (HD)	12	1A, 1B, 2D, 6B	0.06-0.13
Kernel number per spike (KNS)	19	1A, 2B, 2D, 3A, 3B, 4B, 5A, 5B, 6A, 7A, 7B	0.06-0.16
Kernel weight (KW)	9	1A, 2B, 3A, 6A	0.10-0.29
Kernel weight per spike (KWS)	19	3A, 4A, 4B, 4D, 5A, 6B,7D	0.08-0.26
Plant height (PH)	24	1A, 2A, 2B, 2D, 3B, 4A, 4D, 5A, 6B, 7B	0.15-0.34
Peduncle length (PL)	11	1A,2A,2D, 3A, 3B, 7A	0.08-0.15
Spike length (SL)	8	1A, 1B, 7B, 7D	0.06-0.16
HD, KNS	1	1A	0.07-0.09
KNS, PH	1	3B	0.06-0.18
GY, KNS	1	4B	0.08-0.26
GY, KWS	1	4B	0.21-0.28
HD, PH	1	2D	0.08-0.28

[a] Significant markers p value < 0.0005

[b] R^2 values reported as a range; reflect the r^2 of the model with SNP calculated in GAPIT across all BLUP datasets and compressed mixed linear model (CMLM) used for GWAS



Appendix 8. Population structure of the soft red winter wheat germplasm panel showing K=3 different clusters inferred using STRUCTURE**.** Horizontal coordinate represents the specific designation for the entries comprising the association mapping panel while vertical axis is the coefficient of membership, Q for each of the individuals in the population. Each entry was assigned to one of the three subpopulations based on the largest value of a corresponding Q

Appendix 9. Chi square table for the allele frequency of the top 100 highest yielding lines

	Observed	**Expected**	**Observed (O) - Expected (E)**	**$(O-E)^2$**	**$(O-E)^2/E$**
Ppd-D1a	29	50	-21	441	8.82
Ppd-D1b	64	50	14	256	5.12
Total	95	100			13.94

H_0: There is no significant difference between observed and expected allele frequency

H_1: There is significant difference between observed and expected allele frequency

Degrees of freedom = N- 1 = 2-1 =1

At degrees of freedom equal to 1 and df/area at 0.05, we have a critical value of 3.84

Since $\chi^2_c = 13.94 > 3.8$, we reject H_0 and conclude that there is significant difference between allele frequency at the Ppd-D1 locus

Appendix 10. Heritability of the measured traits on each environment

Environment	Country	Location	Trait		
			Grain yield	**Grain number**	**Thousand grain weight**
BJ10	Bangladesh	Joydebpur	- [a]	-	0.71
BJ11	Bangladesh	Joydebpur	0.68	0.69	0.85
EE10	Egypt	El mat	0.78	0.76	0.69
ES10	Egypt	Souhag	-	-	0.82
ID10	India	Delhi	0.66	0.66	-
ID11	India	Delhi	0.66	0.66	-
IH10	India	Dharwad	-	-	0.72
IH11	India	Dharwad	0.91	0.86	0.95
II10	India	Indore	0.64	-	0.80
II11	India	Indore	-	-	0.82
IK10	India	Karnal	-	-	0.78
IK11	India	Karnal	-	-	0.74
IL10	India	Ludhiana	0.57	0.54	0.91
IL11	India	Ludhiana	0.79	0.72	0.52
IV10	India	Varanasi	0.67	0.63	-
IV11	India	Varanasi	-	-	-
MD10 [b]	Mexico	Obregon	0.71	0.77	0.93
MH10 [c]	Mexico	Obregon	0.78	0.86	0.91
MHD10 [d]	Mexico	Obregon	0.68	0.79	0.93
MI10 [e]	Mexico	Obregon	0.74	0.83	0.96
NB10	Nepal	Bhairahwa	-	0.55	0.83
NB11	Nepal	Bhairahwa	0.55	0.53	0.78
PI10	Pakistan	Islamabad	-	-	0.71
PI11	Pakistan	Islamabad	-	-	-
RA10	Iran	Ahwaz	-	-	-
RA11	Iran	Ahwaz	-	-	-
SD10	Sudan	Dongola	0.57	0.59	-
SH10	Sudan	Hudeiba	-	0.52	0.86
SW10	Sudan	Wad Madani	-	0.60	0.74

[a] Indicates that calculated heritability of the trait for that environment was $h^2 < 0.50$

[b] Mexico drought

[c] Mexico heat

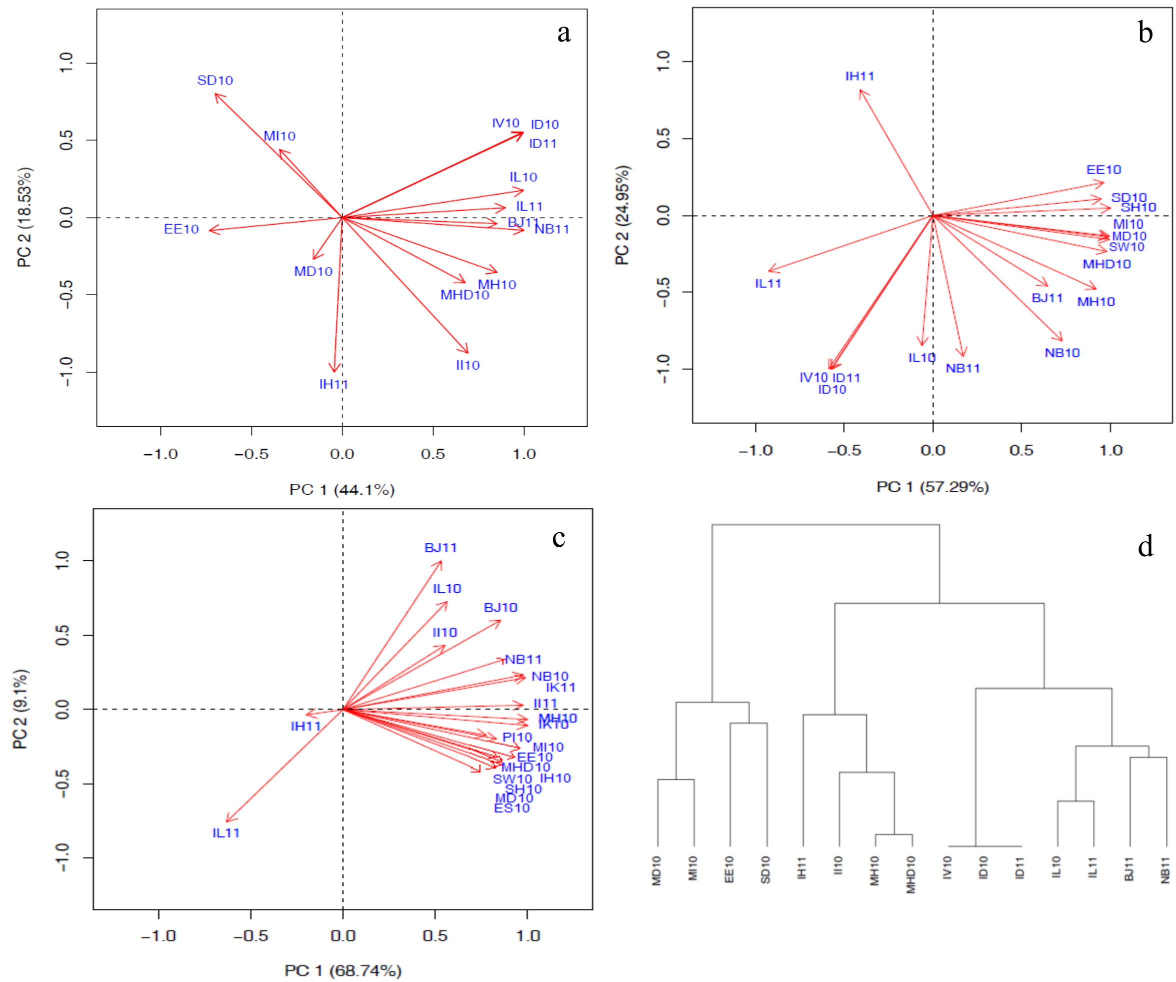
[d] Mexico heat drought

[e] Mexico irrigated

Appendix 11. Phenotypic correlations, *r* for grain yield (GY), grain number (GNO), and thousand grain weight (TGW) for the WAMI panel, CIMMYT, Mexico

Trait	**GY**	**GNO**	**TGW**
GY	-		
GNO	0.54***	-	
TGW	0.27***	-0.66***	-

***- significant at $p < 0.001$



Appendix 12. PCA biplot of the different environments for (a) GY, (b) GNO, and (c) TGW. Environments with $h^2 < 0.50$ were not included for analysis. Dendogram showing relationship among environments for GY (d) using the Ward method. See Appendix 10 for abbreviations

Appendix 13. Description of the different phenotypic datasets used for GWAS

Dataset	**Description**
ABLUP	BLUP across all environments
ABLUP10	BLUP across all environments, 2010
ABLUP11	BLUP across all environments, 2011
ABLUPH	BLUP across environments with $H^2 > 0.50$
ABLUPH10	BLUP across all environments with $H^2 > 0.50$, 2010
ABLUPH11	BLUP across all environments with $H^2 > 0.50$, 2011
AFRBLUP	BLUP across all African environments
AFRBLUP10	BLUP across all African environments, 2010
AFRBLUP11	BLUP across all African environments, 2011
AFRBLUPH	BLUP across all African environments with $H^2 > 0.50$
ASIABLUP	BLUP across all Asian environments
ASIABLUP10	BLUP across all Asian environments, 2010
ASIABLUP11	BLUP across all Asian environments, 2011
ASIABLUPH	BLUP across all Asian environments with $H^2 > 0.50$
BJ10BLUE	BLUE across Bangladesh (Joydebpur) environment, 2010
BJ11BLUE	BLUE across Bangladesh (Joydebpur) environment, 2011
BLUPBANG	BLUP across Bangladesh environments
BLUPBANGH	BLUP across Bangladesh environments, $H^2 > 0.50$
BLUPEGYPT	BLUP across Egypt environments
BLUPEGYPTH	BLUP across Egypt environments, $H^2 > 0.50$
BLUPIND	BLUP across India environments
BLUPINDH	BLUP across India environments, $H^2 > 0.50$
BLUPMEX	BLUP across Mexico environments
BLUPMEXH	BLUP across Mexico environments, $H^2 > 0.50$
BLUPNEPAL	BLUP across Nepal environments
BLUPNEPALH	BLUP across Nepal environments, $H^2 > 0.50$
BLUPPAKISTAN	BLUP across Pakistan environments
BLUPPAKISTANH	BLUP across Pakistan environments, $H^2 > 0.50$
BLUPSUD	BLUP across Sudan environments
BLUPSUDH	BLUP across Sudan environments, $H^2 > 0.50$
EE10BLUE	BLUE across Egypt (El mat) environment, 2010
ES10BLUE	BLUE across Egypt (Souhag) environment, 2010
ID10BLUE	BLUE across India (Delhi) environment, 2010
ID11BLUE	BLUE across India (Delhi) environment, 2011
IH10BLUE	BLUE across India (Dharwad) environment, 2010
IH11BLUE	BLUE across India (Dharwad) environment, 2011
II10BLUE	BLUE across India (Indore) environment, 2010
II11BLUE	BLUE across India (Indore) environment, 2011

Appendix 13 (Cont.)

Dataset	Description
IK10BLUE	BLUE across India (Karnal) environment, 2010
IL10BLUE	BLUE across India (Ludhiana) environment, 2010
IL11BLUE	BLUE across India (Ludhiana) environment, 2011
IV10BLUE	BLUE across India (Varanasi) environment, 2010
MD10BLUE	BLUE across Mexico (Obregon) environment, 2010, Drought
MEXBLUP	BLUP across all Mexico environments; also equivalent to BLUP across all North American locations
MEXBLUP10	BLUP across all Mexico environments; also equivalent to BLUP across all North American locations, 2010
MEXBLUP11	BLUP across all Mexico environments; also equivalent to BLUP across all North American locations, 2011
MEXBLUPH	BLUP across all Mexico environments with $H^2 > 0.50$; also equivalent to BLUP across all North American locations
MEXBLUPH10	BLUP across all Mexico environments; also equivalent to BLUP across all North American locations, 2010, $H^2 > 0.50$
MEXBLUPH11	BLUP across all Mexico environments; also equivalent to BLUP across all North American locations, 2011, $H^2 > 0.50$
MH10BLUE	BLUE across Mexico (Obregon) environment, 2010, Heat
MHD10BLUE	BLUE across Mexico (Obregon) environment, 2010, Heat Drought
MI10BLUE	BLUE across Mexico (Obregon) environment, 2010, Irrigated
NB10BLUE	BLUE across Nepal (Bhairahwa) environment, 2010
NB11BLUE	BLUE across Nepal (Bhairahwa) environment, 2011
PI10BLUE	BLUE across Pakistan (Islamabad) environment, 2010
SD10BLUE	BLUE across Sudan (Dongola) environment, 2010
SH10BLUE	BLUE across Sudan (Hudeiba) environment, 2011
SW10BLUE	BLUE across Sudan (Wad Madani) environment, 2010

Appendix 14. Candidate genes and sequences identified for the validated QTL in the WAMI spring wheat panel

SNP (Chr)	Organism	Genes/Sequences	Loc	E-value	%ID	Putative function
wsnp_Ex_c361_708712 (3A)	*O. sativa indica*	BGIOSGA005238	1	0.23	89.7	Protein coding
wsnp_Ex_c361_708712 (3A)	*H. vulgare*	HORVU3Hr16112690	3H	1.55E-77	93.4	Uncharacterized
wsnp_Ex_c361_708712 (3A)	*B. distachyon*	BRAD12G61670	2	3.9E-11	82.6	Protein coding
wsnp_Ex_c361_708712 (3A)	*B. distachyon*	BRAD12G39600	2	9.7E-9	84.0	Protein coding
wsnp_Ex_c361_708712 (3A)	*T. durum/T. aestivum*	IWGSC_CSS_3AL_scaff_3069047	3A	4.00E-94		Protein coding
wsnp_Ex_c361_708712 (3A)	*T. aestivum*	TRIAE_CS42_3AL_TGACv1_196872_AA0663440	3A	7.60E-103	98.5	Protein coding
wsnp_Ex_c361_708712 (3A)	*T. aestivum*	TRIAE_CS42_3DL_TGACv1_250847_AA0874060	3D	1.50E-52	93.1	Protein coding
wsnp_Ex_c361_708712 (3A)	*T. aestivum*	TRIAE_CS42_3B_TGACv1_220619_AA0711760	3B	1.40E-43	96.2	Protein coding
wsnp_Ex_c361_708712 (3A)	*T. aestivum*	TRIAE_CS42_1BS_TGACv1_049558_AA0156790	1B	1.20E-04	82.7	Protein coding
wsnp_Ex_c361_708712 (3A)	*T. aestivum*	TRIAE_CS42_1DS_TGACv1_080727_AA0252850	1D	0.029	86.7	Protein coding
wsnp_Ex_c361_708712 (3A)	*T. aestivum*	TRIAE_CS42_5BL_TGACv1_406471_AA1345910	5B	7.2	100	Protein coding
wsnp_Ex_c361_708712 (3A)	*T. aestivum*	TRIAE_CS42_1AS_TGACv1_020044_AA0074220	1A	7.2	84.4	Protein coding
wsnp_Ex_c361_708712 (3A)	*T. aestivum*	TRIAE_CS42_3DL_TGACv1_249377_AA0847020	3D	7.2	100	Protein coding
wsnp_Ex_c361_708712 (3A)	*T.urartu*	TRIUR3_21942		2.10E-103	98.5	Putative E3 ubiquitin-protein ligase HERC1
wsnp_Ex_c361_708712 (3A)	*T.urartu*	TRIUR3_04359		0.0082	86.7	Putative E3 ubiquitin-protein ligase HERC2
wsnp_Ex_c361_708712 (3A)	*T.urartu*	TRIUR3_22549		8	95.5	Expansin-B9
wsnp_Ex_c361_708712 (3A)	*T.urartu*	TRIUR3_31959		8	100	Protein coding
wsnp_Ex_c361_708712 (3A)	*T.urartu*	TRIUR3_10995		8	100	Putative WRKY transcription factor 23
wsnp_Ex_c361_708712 (3A)	*A.tauschii*	F775_19709		9.10E-51	92.4	Putative E3 ubiquitin-protein ligase HERC1
wsnp_Ex_c361_708712 (3A)	*A.tauschii*	F775_21039		0.0072	86.7	Putative E3 ubiquitin-protein ligase HERC2

Appendix 14 (Cont.)

SNP (Chr)	Organism	Genes/Sequences	Loc	E-value	%ID	Putative function
wsnp_Ex_c361_708712 (3A)	*A.tauschii*	F775_32805		1.8	100	GDSL esterase/lipase
wsnp_Ex_c361_708712 (3A)	*A.tauschii*	F775_25619		7.1	95.5	Protein coding
wsnp_Ex_c361_708712 (3A)	*A.tauschii*	F775_08716		7.1	100	Putative WRKY transcription factor 23
wsnp_Ex_c361_708712 (3A)	*A.tauschii*	F775_04691		7.1	95.5	GDSL esterase/lipase
wsnp_Ex_c13849_21698240 (4B)	*O. sativa japonica*	OS04G0662900	4	0.22	100.0	Repressor of RNA Pol III
wsnp_Ex_c13849_21698240 (4B)	*O. sativa indica*	BG1OSGA014164	4	0.25	100.0	Uncharacterized
wsnp_Ex_c13849_21698240 (4B)	*A. thaliana*	AT1G51720	1	0.078	95.7	Amino acid dehydrogenase family protein
wsnp_Ex_c13849_21698240 (4B)	*A. thaliana*	AT5G57150	5	0.28	100.0	Transcription factor
wsnp_Ex_c13849_21698240 (4B)	*H. vulgare*	HORVU4Hr1G073630	4H	5.8E-20	96.7	Protein coding
wsnp_Ex_c13849_21698240 (4B)	*T. durum/T. aestivum*	IWGSC_CSS_4BL_scaff_7026506	4B	3.00E-95		Protein coding
wsnp_Ex_c13849_21698240 (4B)	*T.aestivum*	TRIAE_CS42_4BL_TGACv1_321575_AA1062410	4B	7.40E-105	99.5	1,2-dihydroxy-3-keto-5-methylthiopentene dioxygenase
wsnp_Ex_c13849_21698240 (4B)	*T.aestivum*	TRIAE_CS42_4DL_TGACv1_342984_AA1127000	4D	5.90E-50	95.9	1,2-dihydroxy-3-keto-5-methylthiopentene dioxygenase
wsnp_Ex_c13849_21698240 (4B)	*T.aestivum*	TRIAE_CS42_4AS_TGACv1_307193_AA1018250	4A	1.70E-25	95.9	1,2-dihydroxy-3-keto-5-methylthiopentene dioxygenase
wsnp_Ex_c13849_21698240 (4B)	*T.aestivum*	TBP1		7.9	100	TATA-box-binding protein 1
wsnp_Ex_c13849_21698240 (4B)	*T.urartu*	TRIUR3_27052		4.80E-26	95.9	Heat stress transcription factor A-2d
wsnp_Ex_c13849_21698240 (4B)	*T.urartu*	TRIUR3_32346		8.7	100	TATA-box-binding protein 1
wsnp_Ex_c13849_21698240 (4B)	*A.tauschii*	F775_30508		1.50E-50	95.9	1,2-dihydroxy-3-keto-5-methylthiopentene dioxygenase

Appendix 14 (Cont.)

SNP (Chr)	Organism	Genes/Sequences	Loc	E-value	%ID	Putative function
wsnp_CAP11_c3599_1741800 (6B)	*A. thaliana*	ERF113	5	0.018	100.0	Transcription factor
wsnp_CAP11_c3599_1741800 (6B)	*A. thaliana*	T2R2	4	0.28	100.0	Positive regulator of ubiquitin protein ligase activity
wsnp_CAP11_c3599_1741800 (6B)	*A. thaliana*	AT3G20460	3	0.28	100.0	Sugar transporter
wsnp_CAP11_c3599_1741800 (6B)	*H. vulgare*	HORVU6Hr1G060720	6H	3.3E-49	96.6	Protein coding
wsnp_CAP11_c3599_1741800 (6B)	*B. distachyon*	BRAD13G50010	3	8.0E-19	87.5	Protein coding
wsnp_CAP11_c3599_1741800 (6B)	*T. durum/T. aestivum*	IWGSC_CSS_6BL_scaff_4224574	6B	1.00E-93		Protein coding
wsnp_CAP11_c3599_1741800 (6B)	*T.aestivum*	TRIAE_CS42_6BL_TGACv1_500374_AA1603810	6B	6.30E-56	99.1	Protein coding
wsnp_CAP11_c3599_1741800 (6B)	*T.aestivum*	TRIAE_CS42_4DL_TGACv1_343189_AA1131240	4D	2	100	Protein coding
wsnp_CAP11_c3599_1741800 (6B)	*T.aestivum*	TRIAE_CS42_U_TGACv1_643081_AA2127000		7.9	95.7	Protein coding
wsnp_CAP11_c3599_1741800 (6B)	*T.aestivum*	TRIAE_CS42_6DL_TGACv1_527878_AA1709470	6D	1.30E-41	94	Protein coding
wsnp_CAP11_c3599_1741800 (6B)	*T.aestivum*	TRIAE_CS42_6AL_TGACv1_471085_AA1502490	6A	1.10E-29	96.3	Protein coding
wsnp_CAP11_c3599_1741800 (6B)	*T.urartu*	TRIUR3_02969		2.30E-40	93.9	Cytochrome b561
wsnp_CAP11_c3599_1741800 (6B)	*T.urartu*	TRIUR3_18345		8.7	100	Protein coding
wsnp_CAP11_c3599_1741800 (6B)	*A.tauschii*	F775_30054		8.00E-40	93.2	Cytochrome b561
wsnp_CAP11_c3599_1741800 (6B)	*A.tauschii*	F775_02015		0.49	100	Zinc finger CCCH domain-containing protein 44
wsnp_CAP11_c3599_1741800 (6B)	*A.tauschii*	F775_52312		2	100	Protein coding
wsnp_CAP11_c3599_1741800 (6B)	*A.tauschii*	F775_43432		2	100	Protein coding
wsnp_CAP11_c3599_1741800 (6B)	*A.tauschii*	F775_12532		7.7	95.5	Auxin response factor

Appendix 15. Corresponding single nucleotide variants (SNVs) for the BLAST hit region of the for the validated loci in the WAMI.

Locus	Line	Sequence ID	pos	chr	Wild type base	Mut base	Effect gene	Effect consequence
wsnp Ex c361 708712	Cadenza 0230	IWGSC CSS 3AL scaff 3069047	795	3A	C	T	Traes 3AL 10A1A8D E3.2	synonymous variant
wsnp Ex c361 708712	Cadenza 0381	IWGSC CSS 3AL scaff 3069047	864	3A	C	T	Traes 3AL 10A1A8D E3.2	synonymous variant
wsnp Ex c361 708712	Cadenza 0401	IWGSC CSS 3AL scaff 3069047	831	3A	C	T	Traes 3AL 10A1A8D E3.2	synonymous variant
wsnp Ex c361 708712	Cadenza 1158	IWGSC CSS 3AL scaff 3069047	825	3A	C	T	Traes 3AL 10A1A8D E3.2	synonymous variant
wsnp Ex c361 708712	Cadenza 1521	IWGSC CSS 3AL scaff 3069047	873	3A	G	A	Traes 3AL 10A1A8D E3.2	synonymous variant
wsnp Ex c361 708712	Kronos2 208	IWGSC CSS 3AL scaff 3069047	855	3A	C	T	Traes 3AL 10A1A8D E3.2	synonymous variant
wsnp Ex c361 708712	Kronos3 622	IWGSC CSS 3AL scaff 3069047	861	3A	G	A	Traes 3AL 10A1A8D E3.2	synonymous variant
wsnp Ex c361 708712	Kronos9 10	IWGSC CSS 3AL scaff 3069047	863	3A	C	T	Traes 3AL 10A1A8D E3.2	missense variant
wsnp Ex c13849 21698240	Cadenza 1231	IWGSC CSS 4BL scaff 7026506	8196	4B	G	A	Traes 4BL F435C85B F.1	3 prime UTR variant
wsnp Ex c13849 21698240	Cadenza 1265	IWGSC CSS 4BL scaff 7026506	8201	4B	G	A	Traes 4BL 2E125A70 2.1	upstream gene variant
wsnp Ex c13849 21698240	Cadenza 0759	IWGSC CSS 4BL scaff 7026506	8270	4B	C	T	Traes 4BL 2E125A70 2.1	upstream gene variant
wsnp Ex c13849 21698240	Cadenza 1800	IWGSC CSS 4BL scaff 7026506	8276	4B	C	T	Traes 4BL 2E125A70 2.1	upstream gene variant
wsnp Ex c13849 21698240	Cadenza 1429	IWGSC CSS 4BL scaff 7026506	8276	4B	C	T	Traes 4BL F435C85B F.1	3 prime UTR variant
wsnp Ex c13849 21698240	Kronos2 345	IWGSC CSS 4BL scaff 7026506	8277	4B	C	T	Traes 4BL F435C85B F.1	3 prime UTR variant
wsnp Ex c13849 21698240	Kronos3 339	IWGSC CSS 4BL scaff 7026506	8301	4B	C	T	Traes 4BL F435C85B F.1	3 prime UTR variant
wsnp Ex c13849 21698240	Cadenza 1174	IWGSC CSS 4BL scaff 7026506	8334	4B	C	T	Traes 4BL F435C85B F.1	3 prime UTR variant

Appendix 15 (Cont.)

Locus	Line	Sequence ID	pos	chr	Wild type base	Mutated base	Effect gene	Effect consequence
wsnp Ex c13849 21698240	Kronos2025	IWGSC CSS 4BL scaff 7026506	8346	4B	C	T	Traes 4BL 2E125A702.1	upstream gene variant
wsnp Ex c13849 21698240	Cadenza0554	IWGSC CSS 4BL scaff 7026506	8350	4B	G	A	Traes 4BL 2E125A702.1	upstream gene variant
wsnp CAP11 c3599 1741800	Kronos925	IWGSC CSS 6BL scaff 4224574	8578	6B	C	T	Traes 6BL 65F472130.1	missense variant
wsnp CAP11 c3599 1741800	Kronos2933	IWGSC CSS 6BL scaff 4224574	8589	6B	G	A	Traes 6BL 65F472130.1	missense variant
wsnp CAP11 c3599 1741800	Kronos2042	IWGSC CSS 6BL scaff 4224574	8601	6B	G	A	Traes 6BL 65F472130.1	missense variant
wsnp CAP11 c3599 1741800	Cadenza0956	IWGSC CSS 6BL scaff 4224574	8575	6B	G	A	Traes 6BL 65F472130.1	missense variant
wsnp CAP11 c3599 1741800	Cadenza0044	IWGSC CSS 6BL scaff 4224574	8585	6B	G	A	Traes 6BL 65F472130.1	synonymous variant
wsnp CAP11 c3599 1741800	Cadenza0773	IWGSC CSS 6BL scaff 4224574	8590	6B	C	T	Traes 6BL 65F472130.1	missense variant
wsnp CAP11 c3599 1741800	Cadenza1315	IWGSC CSS 6BL scaff 4224574	8592	6B	G	A	Traes 6BL 65F472130.1	missense variant
wsnp CAP11 c3599 1741800	Kronos4607	IWGSC CSS 6BL scaff 4224574	8575	6B	G	A	Traes 6BL 65F472130.1	missense variant

Appendix 16. Accuracy of genomic selection for measured traits across different TP sizes at a constant VP size (N=60)

	Grain yield	**Plant height**	**Heading date**	**Thousand kernel weight**	**Kernel no. spike^{-1}**	**Kernel weight spike^{-1}**
TP25	0.18	0.27	0.19	0.28	0.12	0.19
TP50	0.3	0.46	0.27	0.41	0.23	0.32
TP75	0.38	0.5	0.3	0.43	0.24	0.36
TP100	0.4	0.55	0.35	0.43	0.35	0.43
TP125	0.44	0.65	0.39	0.53	0.38	0.53
TP150	0.46	0.73	0.47	0.58	0.4	0.59

Appendix 17. Accuracy of genomic selection across different marker subsets (SS) from association analyses using BLUP across all environments (ABLUP) dataset

Trait	$SS_{0.15}$	$SS_{0.10}$	$SS_{0.05}$	Whole genotype
Grain yield	0.54	0.56	0.54	0.33
Heading date	-0.011	-0.013	0.003	0.17
Plant height	0.31	0.31	0.25	0.31
Kernel no. $spike^{-1}$	0.19	0.19	0.21	0.31
Kernel weight $spike^{-1}$	0.26	0.26	0.28	0.44
Thousand kernel weight	0.47	0.45	0.42	0.49

Appendix 18. Accuracy of genomic selection for grain yield using inferred subgroups *Q* from STRUCTURE analyses; TP=50; VP= 30

TP/VP	Grain yield	Kernel no. spike^{-1}	Kernel weight spike^{-1}	Thousand kernel weight
Q1/Q2	0.09	0.07	0.04	0.08
Q2/Q1	0.22	0.08	0.08	0.26
Q1/Q3	0.10	0.14	0.08	0.16
Q3/Q1	0.09	0.09	0.21	0.21
Q2/Q3	0.30	0.25	0.16	0.28
Q3/Q2	0.26	0.10	0.20	0.37

Appendix 19. Accuracy using covariates (*Ppd-D1* and *vrn-A1*) in genomic selection for different grain yield datasets

	ABLUP	**BLUP14**	**BLUP15**	**NBLUP**	**SBLUP**
No covariates	0.33	0.08	0.37	0.09	0.44
Ppd-D1	0.37	0.15	0.43	0.13	0.44
Vrn-A1	0.33	0.13	0.43	0.09	0.43
Ppd-D1/Vrn-A1	0.39	0.19	0.43	0.09	0.45

[a]*ABLUP*- BLUP across all environments; *BLUP14*- BLUP across 2014 site-years; *BLUP15*- BLUP across 2015 site-years; *NBLUP*- BLUP across northern environments; *SBLUP*- BLUP across southern environments

Appendix 20. Accuracy of genomic selection for grain yield using NB and PA-RIL as VP

TP	VP							
	NB_ALL	**NB_FAY**	**NB_NPT**	**PA_ALL**	**PA_Cluster1**	**PA_Cluster2**	**PA_Cluster3**	**PA_Cluster4**
ABLUP	0.19	0.22	0.12	0.22	-0.08	0.22	0.06	0.39
NBLUP	0.14	0.18	0.21	0.21	-0.14	0.18	-0.02	0.43
SBLUP	0.15	0.18	0.06	0.27	0.001	0.26	0.05	0.39

[a] Training population- *ABLUP*- BLUP across all environments; *NBLUP*- BLUP across northern environments; *SBLUP*- BLUP across southern environments