Exploring Novel Germplasm and Traits in the National Small Grains Collection Using Genome-Wide Association Mapping Studies

A Dissertation
Presented in Partial Fulfillment of the Requirements for the
Degree of Doctor of Philosophy
with a
Major in Plant Science
in the
College of Graduate Studies
University of Idaho
by
Tyler Gordon

Major Professor: Jianli Chen, Ph.D
Committee Members: J. Michael Bonman, Ph.D.; David Hole, Ph.D.; Daolin Fu, Ph.D.
Department Administrator: Robert Tripepi, Ph.D.

December 2019

Authorization to Submit Dissertation

This dissertation of Tyler Gordon submitted for the degree of Doctor of Philosophy with a Major in Plant Sciences and titled "Exploring Novel Germplasm and Traits in the National Small Grains Collection Using Genome-Wide Association Mapping Studies," has been reviewed in final form. Permission, as indicated by the signatures and dates below, is now granted to submit final copies to the College of Graduate Studies for approval.

Major Professor: _____ Date: _____
Jianli Chen, Ph.D.

Committee Members: _____ Date: _____
J. Michael Bonman, Ph.D.

_____ Date: _____
David Hole, Ph.D.

_____ Date: _____
Daolin Fu, Ph.D.

Department Administrator: _____ Date: _____
Robert Tripepi, Ph.D.

ABSTRACT

Wheat (*Triticum aestivum* L.), and barley (*Hordeum vulgare* L.) are two important small grains crops that fulfill important human caloric and nutritional requirements globally. More efficient small grain production systems are required to meet increasing grain demand while minimizing adverse environmental impacts associated with production. The purpose of this dissertation was to identify novel germplasm, traits, and associated quantitative trait loci (QTL) in the National Small Grains Collection (NSGC) through genome-wide association studies (GWAS). The specific objectives were to: 1) identify novel accessions and QTLs associated with dwarf bunt resistance; and 2) characterize barley accessions and identify QTLs associated with drought tolerance and agronomic traits. In objective 1, 292 winter bread wheat accessions were tested for dwarf bunt (*Tilletia controversa* J.G. Kühn) resistance over a three year trial period in Logan, UT. Ninety-eight accessions were identified as resistant with disease incidence ≤10%, of which 28 accessions were highly resistant with an incidence ≤1% across all three trials. Two genetic loci on chromosome 6D were found to be associated with resistance across three trials. In objective 2, 480 two-row spring barley accessions were assessed for terminal drought tolerance in Aberdeen, ID over two years. Agronomic traits including yield, test weight and protein were measured when the accessions were grown and harvested under regular irrigation and terminal drought treatments. Twenty accessions were highlighted for use in the malting industry and ten accessions were highlighted for use in the food industry. Fifteen genetic loci were associated with at least one agronomic trait across treatments. Accessions and genetic loci identified in these studies can be used to enhance locally adapted wheat and barley cultivars.

ACKNOWLEDGEMENTS

I would like to extend a sincere thank you to my major advisor, Dr. Jianli Chen, and committee members Dr. Mike Bonman, Dr. David Hole and Dr. Daolin Fu, for their advice, input and assistance throughout the dissertation process. Dr. Rui Wang at the University of Idaho assisted greatly with his statistical insights and comradery. Dr. Brian Bowman collected preliminary data for the barley study. Dr. Harold Bockelman provided necessary germplasm, curated GRIN data, and added NSGC expertise. Shannon Ristrem has been a dedicated and resourceful problem solver in the lab and field. Adam Morris provided expert JMP Genomics technical support, and has been an important resource for resolving statistical issues and technical problems within the JMP software suite. I also would like to thank Dr. Kathy Esvelt Klos at the USDA-ARS in Aberdeen, ID for her willingness to serve on my committee for a brief period, and for her assistance with association mapping design and interpretation. This barley study was supported in part by the Triticeae Coordinated Agricultural Project (TCAP), grant no. 2011-68002-30029, funded by the United States Department of Agriculture - National Institute of Food and Agriculture (USDA-NIFA).

DEDICATION

To Teresa and Emma

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

Common bunt (CB)

Dwarf bunt (DB)

False discovery rate (FDR)

Germplasm Resources Information Network (GRIN)

Genome-wide association study (GWAS)

National Plant Germplasm System (NPGS)

Single nucleotide polymorphism (SNP)

USDA National Small Grains Collection (NSGC)

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1:
LITERATURE REVIEW

WHEAT AND BARLEY PRODUCTION

Wheat (*Triticum aestivum* L.) and barley (*Hordeum vulgare* L.) are important cereal staples and wheat alone fulfills about 20% of global caloric and protein requirements (Braun, et al., 2010; Shiferaw, et al., 2013). In 2017 an estimated 771.7 and 147.4 of wheat and barley were harvested, respectively, out of 215 and 54.4 million ha planted (FAOSTAT, 2019). Forecasts for future agricultural demands estimate that a 100-110% increase in global crop production will be required by 2050 to accommodate a projected increase of 2.3 billion people while maintaining current consumption trends (Tilman, et al., 2011). However, annual cereal production gains of only 1-2% have been realized over the past several decades, and annual gains of 3-4% would be required to meet the 2050 consumption demand estimate (Alston, et al., 2009; Fischer and Edmeades, 2010). Additionally, existing agriculture practices contribute to greenhouse gas emissions, loss of biodiversity, water degradation, and deforestation (West, et al., 2014). Better cereal production systems are required to attain increases in yield and quality while minimizing environmental impacts (Alston, et al., 2009; Tester and Langridge, 2010).

GERMPLASM COLLECTIONS: A RESOURCE FOR ADDRESSING EVOLVING PRODUCTION CONSTRAINTS

Addressing evolving constraints in cereal crop production requires a multi-pronged breeding approach that encompasses the assistance of farmers, extension consultants, breeding

programs, end-users and international research centers. One of the underlying foundations of this approach is access and assessment of germplasm with important traits, such as yield, early maturity, and tolerance to biotic and abiotic stresses. These traits can be introgressed into adapted germplasm for cultivation and generation of new cultivars with improved characteristics. Germplasm pools consist of any living tissue from which new plants can be grown, including seeds, leaves, stems, and pollen. Preservation of these resources can be *in situ*, *ex situ* or both with the primary goal of preserving the genetic variability within a crop species, genus or family. *Ex situ* conservation, which includes storage of germplasm in gene banks, allows for systematic storage, data acquisition, and release of accessions for evaluation and introgression.

History of germplasm conservation

Several individuals were instrumental in introducing the theory of germplasm conservation in the early 1900s. The founder of many of these principals was Nikolai Vavilov (1887-1943). Vavilov was a Russian biologist and geographer who dedicated his life to understanding crop diversity and genetics. From 1921 to 1940, Vavilov worked as the Director of the All-Union Institute of Plant Industry, an organization that would eventually be known as N.I. Vavilov Research Institute of Plant Industry (Vavilov, 1992). Early in his tenure, drought along the Volga River led to five million Russian deaths due to starvation. This tragedy focused Vavilov's interest in genetics and crop diversity. Drawing from the theories of De Candolle, Darwin, and Mendel, Vavilov formulated ideas about crop origins that would help alleviate the cycles of hunger and famine in Russia and the world. Impressed with Vavilov's agricultural prowess, and struggling after years of war and

famine, the Bolshevik government directed Vavilov to lead the Lenin All-Union Academy of Agricultural Sciences. In this capacity Vavilov and his teams traveled to 40 countries on five continents and systematically collected 250,000 plant specimens for the newly established germplasm bank in St. Petersburg (Vavilov, 1992). These genetic resources provided Vavilov with abundant material from which to frame and test his hypotheses of homologous variation, plant immunity, and phyto-geography. They also greatly increased the agricultural potential in Russia: barley and potatoes could be established along the Arctic Circle, cotton was grown in the droughty conditions east of the Caspian Sea, and tea, citrus, and quinine plantations were expanded along the Black Sea. Sadly, Vavilov was sent to a gulag as a political prisoner, but his legacy continues to provide genetic resources to counter evolving crop threats through germplasm collections.

Other important contributors to germplasm conservation as an interdisciplinary field of scientific inquiry were Harry Harlan (1882-1944) and his son Jack Harlan (1917-1998). Harry Harlan was a barley breeder who used Vavilov's theories to target the collection of thousands of barley accessions from their centers of origin for the USDA Bureau of Plant Industry. He realized the importance and urgency of collecting genetic diversity and storing it for the foundation of future crop breeding efforts (Harlan, 1957). Jack Harlan, like his father, collected plant accessions extensively and refined Vavilov's theory of crop origins with the concept of primary, secondary, and tertiary gene pools from which to prioritize breeding and collection efforts (Harlan, 1971). As an example of the importance of conserving plant germplasm he described one of his collected accessions from Hakkari Province, Turkey in Practical Problems of Exploration: Seed Crops (1975):

"The potential value of a collection cannot be assessed in the field. Perhaps this statement could best be illustrated by PI 178383, a wheat I collected in a remote part of Eastern Turkey in 1948. It is a miserable looking wheat, tall, thin-stemmed, lodges badly, is susceptible to leaf rust, lacks winter hardiness yet is difficult to vernalize, and has poor baking qualities. Understandably, no one paid any attention to it for some 15 years. Suddenly, stripe rust became serious in the northwestern states and PI 178383 turned out to be resistant to four races of stripe rust, 35 races of common bunt, ten races of dwarf bunt and to have good tolerance to flag smut and snow mold. The improved cultivars based on PI 178383 are reducing losses by a matter of some millions of dollars per year."

Organizations, policies, and locations for germplasm conservation

The Food and Agriculture Organization of the United Nations (FAO) reports there are 1,750 gene bank facilities that store a total of 7.4 million accessions, two million of which are considered unique (FAO, 2010). Most countries, including the United States, have a national gene bank or germplasm collection. In the United States, the U.S. National Plant Germplasm System (NPGS) supports agricultural production by acquiring, conserving, and evaluating, documenting, and distributing crop germplasm. Managed by the Agricultural Research Service (ARS), the NPGS aims to safeguard genetic diversity of important crops (Williams, 2005). In addition to national gene banks, a global system of eleven Consultative Group on International Agricultural Research (CGIAR) gene banks conserves and enables plant researcher access to unique sources of crop diversity. CGIAR centers are recognized by the International Treaty on Plant Genetic Resources for Food and Agriculture (ITPGRFA) and funded by the Crop Diversity Endowment Fund. The eleven CGIAR centers include

AfricaRice, Bioversity International, International Center for Tropical Agriculture (CIAT), International Maize and Wheat Improvement Center (CIMMYT), International Potato Center (CIP), International Center for Agricultural Research in the Dry Areas (ICARDA), World Agroforestry (ICRAF), International Crop Research Institute for the Semi-Arid Tropics (ICRISAT), International Institute for Tropical Agriculture (IITA), International Livestock Research Institute (ILRI), and International Rice Research Institute (IRRI).

Backup samples of worldwide gene bank accessions exist at the Svalbard Global Seed Vault as a type of germplasm conservation insurance policy. The Svalbard Global Seed Vault was built halfway between Norway and the North Pole as a storage facility meant to safeguard important plant genetic material from around the world. The Seed Vault stores germplasm collections from the CGIAR gene banks as well as national gene banks. Because back-up samples of plant germplasm are stored at Svalbard, valuable genetic diversity will not be lost to natural disasters or accidents at the contributing germplasm banks. Funding for the Svalbard Global Seed Vault comes from the Global Crop Diversity Trust (FAO, 2001), which provides financial support for work associated with the International Treaty on Plant Genetic Resources for Food and Agriculture (IT PGRFA).

The IT PGRFA was signed by the Food and Agriculture Organization of the United Nations in 2001 and became effective in 2004. The treaty has several goals, including those of recognizing farmers' contributions to crop diversity, creating a global system for involved parties to have access to plant genetic materials, and sharing any derived benefits from genetic materials with their countries of origin (FAO, 2001). In 1983, leading up to the adoption of the IT PGRFA, the Commission of Genetic Resources for Food and Agriculture was established as a long-standing, cross-governmental organization to target biological

diversity in relation to food and agriculture. Every ten years the Commission assesses the current state of the global system for access to and use of plant genetic resources as well as gaps and challenges in the system.

The most recent global assessment, entitled the Second Report on the State of the World's Plant Genetic Resources for Food and Agriculture (SoWPGR-2), was completed in 2010. The SoWPGR-2 reports the status of conservation and utilization of plant diversity in terms of global, regional, and national achievements, important scientific and technical progress, and pertinent issues (FAO, 2010). According to the SoWPGR-2 findings, there has been progress in securing germplasm diversity *ex situ*. Similarly, more minor crop and crop wild relative accessions have been conserved yet they continue to be underrepresented in germplasm collections (FAO, 2010). The SoWPGR-2 further details information on crop accession numbers stored at various institutions. For instance, 14 percent of wheat accessions are stored at CIMMYT, seven percent are stored at the USDA-ARS National Small Grains Collection (NSGC), and five percent are stored at the Institute for Commercial Forestry Research (ICFR-CAAS) in South Africa. Remaining institutions store between 1 to 4 percent of wheat accessions. For barley, 9% of accessions are stored at the Plant Gene Resources of Canada 6% at NSGC, 6% at Embrapa Genetic Resources & Biotechnology and 6% at ICARDA. Remaining institutions store between 1 to 5 percent of barley accessions (FAO, 2010).

Use of germplasm resources

Germplasm resources have been used extensively to enhance crop production and respond to evolving climate or disease pressures. The value of these genetic resources support

continued conservation efforts. Some striking examples of this utility can be found in research to identify new sources of resistance to plant diseases. Cereal rusts encompass an important group of plant diseases that affect wheat, barley and oats, and have plagued small grain production since early cultivation of these crops in the Bronze Age (Kislev, 1982). Rust diseases spread rapidly, severely reduce grain yield and quality, and the rust pathogens can evolve quickly to overcome known sources of genetic resistance (Roelfs, et al., 1992; McIntosh, et al., 1995). New races of the stem rust pathogen (*Puccinia graminis* f.sp. *tritici* Eriks. & E. Henn.) emerged in Eastern Africa within the past two decades (Pretorius, et al., 2000) and have overcome widely adopted resistance genes in wheat including *Sr24*, *Sr31*, *Sr36*, and *SrTmp* (Pretorius, et al., 2010; Newcomb, et al., 2016). Screening for novel sources of resistance to these emerging races has become an international priority (Singh, et al., 2008), and gene banks have provided a source of wheat landraces (Rouse, et al., 2011a; Newcomb, et al., 2013; Babiker, et al., 2017) and wild relatives (Xu, et al., 2009; Rouse, et al., 2011b; Edae, et al., 2016) to test for resistance to the new races.

Fusarium head blight (*Fusarium graminearum* Schwabe) is another serious disease of wheat and barley that can reduce grain yields and quality, and also produces deoxynivalennol (DON), a mycotoxin that adversely affects human and animal health (McMullen, et al., 1997). Researchers have screened thousands of bread wheat (Yu, et al., 2008; He, et al., 2014), tetraploid wheat (Oliver, et al., 2008) and barley (Mamo and Steffenson, 2015) accessions from gene banks for resistance to fusarium head blight infection and DON accumulation. Resistance is rare, but is present in bread wheat landrace accessions originating from China (Zhang, et al., 2012), tetraploid wheat accessions

originating from Tunisia, Georgia, and Ethiopia (Elias, et al., 2005; Oliver, et al., 2008) and barley landrace accessions originating from Ethiopia (Mamo and Steffenson, 2015).

Rising global temperatures exacerbate small grain yield and quality losses associated with drought and heat stress (Lesk, et al., 2016; Zhao, et al., 2017; Xie, et al., 2018). Discovering drought and heat tolerant accessions and dissecting the genetic basis of their tolerance to these stressors will help plant breeders address challenges associated with rising global temperatures (Sachs, 2012). Drought and heat tolerance are complex traits that can be measured in a variety of ways either by direct measurements of stomatal conductance (Fischer, et al., 1998) or indirect methods such as near-infrared reflectance (Mwadzingeni, et al., 2016a), integrated water-use efficiency (Araus, et al., 2002), biochemical markers (Yang, et al., 2010), thermometers (Araus and Cairns, 2014), and drought selection indices (Fischer and Maurer, 1978). The combination of these inferential methods with genotyping arrays has been proposed as a way to identify genetic loci associated with stress tolerance in targeted environments (Fleury, et al., 2010). Some of these methods have been used to identify accessions from germplasm collections with drought and heat tolerance. Several examples include wheat accessions from the NSGC (Bowman, et al., 2015; Liu, et al., 2017b) and CIMMYT drought nurseries (Mwadzingeni, et al., 2016b), barley landraces accessions from ICARDA (Acevedo, et al., 1991; Zhao, et al., 2010), and oat accessions from Spain (Sánchez-Martín, et al., 2012; Sánchez-Martín, et al., 2017).

GENOMIC TOOLS FOR GERMPLASM EXPLORATION

Recent advances in cereal genotyping arrays (Wang, et al., 2014; Chaffin, et al., 2016; Bayer, et al., 2017) and genomic sequences (Mascher, et al., 2017; IWGSC, 2018) have increased the efficiency and power of plant crop trait detection using previously established statistical models. Genetic mapping refers generally to the identification of genomic regions associated with specific traits of interest. Mapping was first reported in 1923 (Sax, 1923) to detect linkage between pigmentation and seed size in common bean (*Phaseolus vulgaris* L.). Since then, the techniques have evolved to include dense molecular marker arrays and sophisticated statistical procedures, but the goals of mapping have been consistent: identify and locate important traits that can be introgressed into modern cultivars. Two popular mapping procedures used today are linkage mapping and association mapping. Linkage mapping draws inferences between markers and traits through recent recombination in bi-parental populations, while association mapping utilizes natural populations and historical recombination events to find associations between markers and traits of interest (Kruglyak, 2008).

Molecular markers

Molecular markers can be thought of as genome-wide road markers: the chromosome location of trait variations can be inferred by how closely the trait is associated with these small genomic fragments. This is analogous to inferring the location of Pocatello, Idaho by reading the mile marker on Interstate 84. There are many types of molecular markers, and they broadly fall into three categories including: hybridization, polymerase chain reaction

(PCR) and next generation sequencing (NGS) (Amom and Nongdam, 2017). Restriction fragment length polymorphisms (RFLPs) use restriction enzymes to cleave certain genomic sites and the resulting fragments of interest are hybridized with chemically labeled probes (Lander and Botstein, 1989). Variations in fragment sizes between genotypes can be differentiated through Southern blots (Amom and Nongdam, 2017). Single nucleotide polymorphisms (SNPs) are the most common type of genomic variation, and the allelic diversity of SNPs can be detected using PCR or NGS strategies (Rafalski, 2002). In the PCR-based approach, SNPs are identified by amplifying genomic regions of interest, then sequencing the products in order to find variations among genotypes (Rafalski, 2002). Genotyping-by-sequencing (GBS) utilizes a different method of SNP discovery where methylation-sensitive restriction enzymes with sample-specific barcodes are used to excise specific genomic regions in a set of genotypes which then undergo a round of target-specific amplification and sequencing (Elshire, et al., 2011). Genomic polymorphisms identified through any of these methods allow for linkage and trait mapping, but the SNP-based approaches typically have the most dense marker profiles which can provide better trait location estimates.

Linkage mapping

There are several types of linkage mapping including single marker analysis, interval mapping, composite interval mapping, and multiple interval mapping. The single marker analysis approach uses a t-test or ANOVA to infer a QTL (Thoday, 1961; Soller, et al., 1976), Interval mapping (IM) uses a pair of markers and the likelihood ratio test statistic to infer QTL relationships through mixed models (Lander and Botstein, 1989; Jansen, 1993).

While IM can estimate QTL positions and effects more accurately than the single marker analysis, requires fewer individuals, and can differentiate recombination from the effect of the QTL, it has several notable limitations. For instance, the IM procedure can bias results when there are more than one QTL on a chromosome, and only two markers can be tested simultaneously, which limits the efficiency of the model (Zeng, 1994). Composite Interval Mapping (CIM) was first proposed by Zeng (1993; 1994) as a statistical procedure to separate multiple, linked quantitative trait loci (QTL). CIM added a partial regression equation to the interval mapping model, which simultaneously confines the model to one dimension, while allowing for stepwise introgression of multiple marker information. This increases sensitivity and precision of the QTL result. CIM has been broadly used to map many cereal crops traits (Babiker, et al., 2015; Babiker, et al., 2017; Huang, et al., 2018; Solis, et al., 2018; Wang, et al., 2019).

Practically, the CIM procedure uses molecular marker and trait information derived from a cross between two parents. Polymorphic markers are ordered and assigned positions on chromosomes or linkage groups. First, a single marker analysis procedure is performed, and the significant QTL are discovered and used to calculate a multiple regression equation. Another round of single marker analysis is performed after removing the first round QTL from the data. If additional QTL are discovered, they are included in the multiple regression equation as cofactors. Interval mapping is performed to determine whether the QTL exists either in the interval or near the cofactors.

The underlying assumptions for the phenotypic data are that the traits are quantitative, normally distributed, heritable, and lack epistatic effects (Zeng, 1994; Doerge, 2002). Even with the advantages that CIM affords over IM, the model has difficulty locating

two QTL in adjacent marker intervals. Four properties of CIM established by Zeng (1994) include: 1) partial regression coefficient of the trait depends only on the QTLs located on the interval bracketed by the neighboring markers; 2) unlinked markers can be used in the regression and reduce sampling variance thereby increasing the power of the test; 3) linked markers used in the regression will reduce interference from multiple linked QTL, but may decrease the power of the test by increasing sampling variance; and 4) test statistics calculated for two different intervals should not be correlated.

When phenotypic data is not normally distributed (Shapiro and Wilk, 1965), as in the case for many Mendelian traits, the test statistic used to determine significance can also be affected. Skewness is a measure of distribution symmetry where a normal curve would give the value 0, negative values would indicate skewness to the left, and positive values would indicate skewness to the right. Kurtosis is a measure of the distribution tails where the kurtosis value of a normal curve is three, excessive kurtosis would have a value beyond three. Normality assumptions can be visually inspected using a quantile-quantile (Q-Q) plot, or formally measured with a Shapiro-Wilk test of normality (Razali and Wah, 2011). Several methods have been recommended for testing non-normal phenotypic distributions including: simulation of significance thresholds (Lander and Botstein, 1989), permutations (Doerge and Churchill, 1996), transformations (Bajgain, et al., 2016; Braun, et al., 2017; Jia, et al., 2018), and nonparametric statistical tests (Kruglyak and Lander, 1995).

Output for CIM includes a likelihood profile expressed as the logarithm of the odds (LOD), phenotypic variation explained by the QTL (R^2), allelic effect, and QTL position. LOD expresses the log of the likelihood ratio of there being a QTL at a marker. The phenotypic variation (R^2) calculation is based on the partial correlation of the marker with

the phenotype and indicates how well a marker explains the phenotype. The allelic effect is calculated by taking the average difference in trait scores between those individuals in the population that are homozygous AA and homozygous BB.

An appropriate LOD significance threshold for a genome-wide analysis will depend on the size of the genome and number of markers. After performing extensive simulations on the IM model, Lander and Botstein (1989) determined that an LOD threshold between 2 and 3 would ensure an overall false positive rate of 0.05. Piepho (2001) proposed another method for approximating QTL statistical thresholds. This method uses experimentally calculated CIM LOD values and can be used when accepted thresholds are not available, and phenotypic distributions are not normal. A later review (Ott, et al., 2015) noted that an LOD of 3.3 would correspond to a genome-wide significance level of $\alpha = 0.05$.

Some linkage mapping refinements include QTL by environment interactions (QEI) proposed by Li et al. (2015) and multiple interval mapping (MIM) which uses multiple regression to detect main and epistatic QTL effects in the population (Kao, et al., 1999). Both techniques have been used extensively in plant research to refine the search for the genetic basis of complex traits (Mauricio, 2001; Maccaferri, et al., 2008; Jia, et al., 2018; Kolmer, et al., 2018).

Genome-wide association studies

Genome-wide association studies (GWAS) have been used to locate genetic variants associated with traits of interest using existing populations coupled with expansive molecular marker arrays (Kraakman, et al., 2004; Hirschhorn and Daly, 2005; Breseghello and Sorrells, 2006; Zhao, et al., 2011; Bush and Moore, 2012). GWAS have several

advantages over linkage mapping studies including the potential to localize the trait of interest to a smaller genetic region, and the ability to use available populations instead of having to develop them. One potential problem with GWAS is not having enough markers to adequately cover the genome, as markers need to be highly correlated with the trait causal allele to detect the association. Hirschhorn and Daly (2005) indicate that even 100,000 markers in the human genome would place 1 marker at every 30 kb which would provide adequate information for only 50% of the common variants.

Both quantitative and qualitative phenotypes have been used in GWAS (Zhao, et al., 2011; Bush and Moore, 2012). Analysis of quantitative traits is usually conducted with an ANOVA with these assumptions: 1) the trait is normally distributed, 2) each group has a similar level of variance, and 3) the groups are independent (Bush and Moore, 2012).

Output for an association analysis is like the output for CIM. It includes the marker position and a *P*-value often reported as the negative logarithm of the *P*-value where a larger number indicates a more significant response. Additionally, the phenotypic variation explained by the QTL (R^2) is reported. As in CIM, the phenotypic variation (R^2) calculation is based on the partial correlation of the marker with the phenotype and indicates how well a marker explains the phenotype.

There can be significant effects of cryptic population stratification and kinship relatedness in GWAS (Yu, et al., 2006). Well-matched case-control reduces the effects of population stratification (Hirschhorn and Daly, 2005). Population stratification can also be introduced into the GWAS model as a cofactor by correcting ancestor correlations through the use of principal component analysis or a Q matrix (Price, et al., 2006; Yu, et al., 2006). Kinship or relatedness measures (K) can be introduced into a mixed model as a random

effect calculated through identity by state (IBS) or identity by descent (IBD) matrices, which calculate an allele sharing matrix between individuals (Yu, et al., 2006). Compensating for population structure and relatedness in a mixed association model adequately controls Type I errors, but can introduce Type II errors (i.e. false negatives) (Zhao, et al., 2011). A QK mixed model for structured association was proposed by Yu et al. (2006):

$Y = X\beta + Q\upsilon + Zu + e$

Y is the phenotypic vector,

X is the molecular marker matrix,

β is the unknown vector of allele effects to be estimated,

Q is the posterior probabilities matrix of belonging to each population obtained from a population structure (PCA) analysis

υ is the vector of population effects (parameters),

Z is a matrix that relates each measurement to the individual from which it was obtained, a relationship matrix

u is the vector of random background polygenic effects

e is the residual error.

Only fixed effect factors are included as parameters for general linear models, while both fixed and random factors are included in general mixed linear models (Breslow and Clayton, 1993). A fixed factor is the specific factor of interest (i.e. treatments or genotypes) and is expected to be correlated with the independent variable. Interpretation of the fixed effect comes from the differences in the mean treatment responses, and the residual variance is the error term in the resulting *F*-test (Moore and Dixon, 2015). Fixed-effects linear models assume the variance components sum to zero across the levels of the experiment.

Random effect factors, on the other hand, are intended to be representative of the population in general terms and are not expected to be correlated with the independent variable. Interpretation of the random effect comes from the factor variance, and the random treatment interaction is used as the error term in the *F*-test (Moore and Dixon, 2015). In mixed model association analyses covariance matrices included for population structure (Q) and relatedness (K) are typically included as fixed and random factors, respectively.

In addition to controlling for cryptic population structure and relationships, a multiple testing correction should also be included as a *post-hoc* transformation on the resulting GWAS *P*-values. Lander and Botstein (1989) proposed the use of a Bonferroni multiple testing correction where each individual test is given a comparison-wise significance level of α/M (M is the number of tests performed.) Although the Bonferroni multiple-testing correction calculation is straightforward, several groups have shown it to be overly conservative (Hirschhorn and Daly, 2005; Bush and Moore, 2012), giving rise to Type II errors, because the independence assumption between tests is violated.

Bush and Moore (2012) used the false discovery rate (FDR) multiple testing correction reported by Benjamini and Hochberg (1995) for multiple allele testing correction in GWAS. The FDR multiple correction sorts the *P*-values in ascending order, then each *P*-value ranked percentile is multiplied by the α level. These values can be converted to FDR-adjusted *P*-values, called *q*-values where tests below the critical value are rejected (Storey and Tibshirani, 2003). The FDR correction procedure is easy to implement using either Excel or other software programs and is less conservative than the Bonferroni correction.

Instead of a *P*-value adjustment, some propose a permutation testing method to empirically determine an appropriate *P*-value threshold (Hirschhorn and Daly, 2005; Pe'er,

et al., 2008). In the permutation testing method, the phenotypic values are shuffled with respect to their genotypes and run through the association mapping algorithm to find a threshold that does not reliably detect significance in these randomly generated phenotype-genotype pairs. While the permutation method is empirically derived, it does not preserve population structure dynamics within the genotype set (Hayes, 2013), and often thousands of permutations are required to find a reliable significance threshold (Pe'er, et al., 2008; Backes, et al., 2014).

Quantile-quantile (Q-Q) plots of the observed and expected *P*-values provide a visual inspection of the association analysis output. If the observed values fall on the expected significance line there is little evidence of association or population structure, whereas an indication of population structure or association is observed when values deviate from the expected line (McCarthy, et al., 2008).

Akaike information criterion (AIC) or Bayesian information criterion (BIC) values (Akaike, 1974; Schwarz, 1978), and/or the mean squared deviation (MSD) can be used to formally compare various association models with one another. The AIC and BIC use the number of parameters in the model combined with the maximum likelihood for each model to identify the model with the lowest value. MSD, however, calculates the residual sum of squares of the observed values compared to those expected from the model, the lower the MSD value the better the model fits the expectation (Wallach and Goffinet, 1989; Mamidi, et al., 2011).

True marker-trait association identification is the primary objective of most GWAS. However, the power and reproducibility of the results is contingent on several factors including variation explained by the marker nearest the mutations of interest, mutation effect

size, and mutation frequency within the GWAS population. Rare variant alleles (those with less penetrance) or those alleles with relatively small effects are difficult to detect especially in GWAS with small population size (Gibson, 2012; Korte and Farlow, 2013). Nevertheless, GWAS offers a complementary approach to linkage mapping, and has been used in numerous small grain studies to identify common variants linked with traits that are important for global crop production (Korte and Farlow, 2013; Bajgain, et al., 2015; Pantalião, et al., 2016; Shi, et al., 2017).

Specifically, the NSGC has been a resource that provides germplasm for GWAS studies with the aim of identifying resistance to diseases including stripe rust (Bulli, et al., 2016; Liu, et al., 2017a), stem rust (Bajgain, et al., 2015), FHB (Mamo and Steffenson, 2015) and insect pests including Russian wheat aphids (Valdez, et al., 2012) and stem sawfly (Varella, et al., 2017). GWAS has also been used to identify genetic regions associated with important small grain agronomic characteristics within the NSGC including yield, test weight and drought tolerance (Bowman, et al., 2015; Winkler, et al., 2016; Liu, et al., 2017b).

REFERENCES

Acevedo, E., P. Craufurd, R. Austin and P. Perez-Marco. 1991. Traits associated with high yield in barley in low-rainfall environments. J Agric Sci 116: 23-36.

Akaike, H. 1974. A new look at the statistical model identification. Selected Papers of Hirotugu Akaike. Springer. p. 215-222.

Alston, J.M., J.M. Beddow and P.G. Pardey. 2009. Agricultural research, productivity, and food prices in the long run. Science 325: 1209-1210.

Amom, T. and P. Nongdam. 2017. The Use of Molecular Marker Methods in Plants: A Review. Int J Curr Res 9: 1-7.

Araus, J., G. Slafer, M. Reynolds and C. Royo. 2002. Plant breeding and drought in C3 cereals: what should we breed for? Ann Bot 89: 925-940.

Araus, J.L. and J.E. Cairns. 2014. Field high-throughput phenotyping: the new crop breeding frontier. Trends Plant Sci 19: 52-61.

Babiker, E.M., T.C. Gordon, S. Chao, M.N. Rouse, Y. Jin, S. Bhavani, et al. 2017. Genetic loci conditioning adult plant resistance to the Ug99 race group and seedling resistance to races TRTTF and TTTTF of the stem rust pathogen in wheat landrace CItr 15026. Plant Dis 101: 496-501.

Babiker, E.M., T.C. Gordon, E.J. Jackson, S. Chao, S.A. Harrison, M.L. Carson, et al. 2015. Quantitative trait loci from two genotypes of oat (*Avena sativa* L.) conditioning resistance to *Puccinia coronata*. Phytopathology 105: 239-245.

Backes, C., F. Rühle, M. Stoll, J. Haas, K. Frese, A. Franke, et al. 2014. Systematic permutation testing in GWAS pathway analyses: identification of genetic networks in dilated cardiomyopathy and ulcerative colitis. BMC Genomics 15: 622.

Bajgain, P., M.N. Rouse, P. Bulli, S. Bhavani, T. Gordon, R. Wanyera, et al. 2015. Association mapping of North American spring wheat breeding germplasm reveals loci conferring resistance to Ug99 and other African stem rust races. BMC Plant Biol 15: 249-268.

Bajgain, P., M.N. Rouse, T.J. Tsilo, G.K. Macharia, S. Bhavani, Y. Jin, et al. 2016. Nested association mapping of stem rust resistance in wheat using genotyping by sequencing. PLoS One 11: e0155760.

Bayer, M.M., P. Rapazote-Flores, M. Ganal, P.E. Hedley, M. Macaulay, J. Plieske, et al. 2017. Development and evaluation of a barley 50k iSelect SNP array. Front Plant Sci 8: 1792.

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

Bowman, B., J. Chen, J. Zhang, J. Wheeler, Y. Wang, W. Zhao, et al. 2015. Evaluating grain yield in spring wheat with canopy spectral reflectance. Crop Sci 55: 1881-1890.

Braun, H.J., G. Atlin, T. Payne. 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, CAB International 115-138.

Braun, S.R., J.B. Endelman, K.G. Haynes and S.H. Jansky. 2017. Quantitative trait loci for resistance to common scab and cold-induced sweetening in diploid potato. Plant Genome 10.

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

Bulli, P., J. Zhang, S. Chao, X. Chen and M. Pumphrey. 2016. Genetic architecture of resistance to stripe rust in a global winter wheat germplasm collection. G3-Genes Genom Genet 6: 2237-2253.

Bush, W.S. and J.H. Moore. 2012. Genome-wide association studies. PLoS Comput Biol 8: e1002822.

Chaffin, A.S., Y.-F. Huang, S. Smith, W.A. Bekele, E. Babiker, B.N. Gnanesh, et al. 2016. A consensus map in cultivated hexaploid oat reveals conserved grass synteny with substantial subgenome rearrangement. Plant Genome 9.

Doerge, R.W. 2002. Multifactorial genetics: Mapping and analysis of quantitative trait loci in experimental populations. Nat Rev Genet 3: 43.

Doerge, R.W. and G.A. Churchill. 1996. Permutation tests for multiple loci affecting a quantitative character. Genetics 142: 285-294.

Edae, E.A., P.D. Olivera, Y. Jin, J.A. Poland and M.N. Rouse. 2016. Genotype-by-sequencing facilitates genetic mapping of a stem rust resistance locus in Aegilops umbellulata, a wild relative of cultivated wheat. BMC Genomics 17: 1039.

C. S. M. e. al. 2005. Breeding efforts to develop Fusarium head blight resistant durum wheat in North Dakota. Proc Natl Fusarium Head Blight Forum, Milwaukee. Michigan State Univ.

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

FAO. 2001. International Treaty on Plant Genetic Resources for Food and Agriculture. Food & Agriculture Organization of the United Nations.

FAO. 2010. The second report on the state of the world's plant genetic resources for food and agricultureFood & Agriculture Organization of the United Nations.

FAOSTAT. 2019. Production and yield quantities of cereal grains. Food and agriculture organization of the United Nations.

Fischer, R. and G.O. Edmeades. 2010. Breeding and cereal yield progress. Crop Sci 50: 85-98.

Fischer, R. and R. Maurer. 1978. Drought resistance in spring wheat cultivars. I. Grain yield responses. Aust J Agric Res 29: 897-912.

Fischer, R., D. Rees, K. Sayre, Z.-M. Lu, A. Condon and A.L. Saavedra. 1998. Wheat yield progress associated with higher stomatal conductance and photosynthetic rate, and cooler canopies. Crop Sci 38: 1467-1475.

Fleury, D., S. Jefferies, H. Kuchel and P. Langridge. 2010. Genetic and genomic tools to improve drought tolerance in wheat. J Exp Bot 61: 3211-3222.

Gibson, G. 2012. Rare and common variants: twenty arguments. Nat Rev Genet 13: 135.

Harlan, J. 1975. Practical problems in exploration: Seed crops. Crop genetic resources for today and tomorrow. Cambridge Univ. Press, Cambridge, UK: 111-115.

Harlan, J.R. 1971. Agricultural origins: Centers and noncenters. Science 174: 469-474.

Hayes, B. 2013. Overview of statistical methods for genome-wide association studies (GWAS). Genome-Wide Association Studies and Genomic Prediction. Springer. p. 149-169.

He, X., P.K. Singh, N. Schlang, E. Duveiller, S. Dreisigacker, T. Payne, et al. 2014. Characterization of Chinese wheat germplasm for resistance to Fusarium head blight at CIMMYT, Mexico. Euphytica 195: 383-395.

Hirschhorn, J.N. and M.J. Daly. 2005. Genome-wide association studies for common diseases and complex traits. Nat Rev Genet 6: 95.

Huang, Y., M. Haas, S. Heinen, B. Steffenson, K. Smith and G. Muehlbauer. 2018. QTL mapping of fusarium head blight and correlated agromorphological traits in an elite barley cultivar Rasmusson. Front Plant Sci 9: 1260.

IWGSC. 2018. Shifting the limits in wheat research and breeding using a fully annotated reference genome. Science 361.

Jansen, R.C. 1993. Interval mapping of multiple quantitative trait loci. Genetics 135: 205-211.

Jia, A., Y. Ren, F. Gao, G. Yin, J. Liu, L. Guo, et al. 2018. Mapping and validation of a new QTL for adult-plant resistance to powdery mildew in Chinese elite bread wheat line Zhou8425B. Theor Appl Genet 131: 1063-1071.

Kao, C.-H., Z.-B. Zeng and R.D. Teasdale. 1999. Multiple interval mapping for quantitative trait loci. Genetics 152: 1203-1216.

Kislev, M.E. 1982. Stem rust of wheat 3300 years old found in Israel. Science 216: 993-994.

Kolmer, J., S. Chao, G. Brown-Guedira, U. Bansal and H. Bariana. 2018. Adult plant leaf rust resistance derived from the soft red winter wheat cultivar ‘Caldwell' maps to chromosome 3BS. Crop Sci 58: 152-158.

Korte, A. and A. Farlow. 2013. The advantages and limitations of trait analysis with GWAS: a review. Plant Methods 9: 29.

Kraakman, A.T., R.E. Niks, P.M. Van den Berg, P. Stam and F.A. Van Eeuwijk. 2004. Linkage disequilibrium mapping of yield and yield stability in modern spring barley cultivars. Genetics 168: 435-446.

Kruglyak, L. 2008. The road to genome-wide association studies. Nat Rev Genet 9: 314.

Kruglyak, L. and E.S. Lander. 1995. A nonparametric approach for mapping quantitative trait loci. Genetics 139: 1421-1428.

Lander, E.S. and D. Botstein. 1989. Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121: 185-199.

Lesk, C., P. Rowhani and N. Ramankutty. 2016. Influence of extreme weather disasters on global crop production. Nature 529: 84.

Li, S., J. Wang and L. Zhang. 2015. Inclusive composite interval mapping of QTL by environment interactions in biparental populations. PLoS One 10: e0132414.

Liu, W., M. Maccaferri, P. Bulli, S. Rynearson, R. Tuberosa, X. Chen, et al. 2017a. Genome-wide association mapping for seedling and field resistance to *Puccinia striiformis* f. sp. *tritici* in elite durum wheat. Theor Appl Genet 130: 649-667.

Liu, Y., B. Bowman, Y.-G. Hu, X. Liang, W. Zhao, J. Wheeler, et al. 2017b. Evaluation of agronomic traits and drought tolerance of winter wheat accessions from the USDA-ARS national small grains collection. Agronomy 7: 51.

Maccaferri, M., M.C. Sanguineti, S. Corneti, J.L.A. Ortega, M.B. Salem, J. Bort, et al. 2008. Quantitative trait loci for grain yield and adaptation of durum wheat (Triticum durum Desf.) across a wide range of water availability. Genetics 178: 489-511.

Mamidi, S., S. Chikara, R.J. Goos, D.L. Hyten, D. Annam, S.M. Moghaddam, et al. 2011. Genome-wide association analysis identifies candidate genes associated with iron deficiency chlorosis in soybean. Plant Genome 4: 154-164.

Mamo, B.E. and B.J. Steffenson. 2015. Genome-wide association mapping of Fusarium head blight resistance and agromorphological traits in barley landraces from Ethiopia and Eritrea. Crop Sci 55: 1494-1512.

Mascher, M., H. Gundlach, A. Himmelbach, S. Beier, S.O. Twardziok, T. Wicker, et al. 2017. A chromosome conformation capture ordered sequence of the barley genome. Nature 544: 427-433.

Mauricio, R. 2001. Mapping quantitative trait loci in plants: uses and caveats for evolutionary biology. Nat Rev Genet 2: 370-381.

McCarthy, M.I., G.R. Abecasis, L.R. Cardon, D.B. Goldstein, J. Little, J.P. Ioannidis, et al. 2008. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nat Rev Genet 9: 356.

McIntosh, R.A., C.R. Wellings and R.F. Park. 1995. Wheat Rusts an atlas of resistance genesCSIRO, East Melbourne.

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

Mwadzingeni, L., H. Shimelis, E. Dube, M.D. Laing and T.J. Tsilo. 2016a. Breeding wheat for drought tolerance: Progress and technologies. J Integr Agr 15: 935-943.

Mwadzingeni, L., H. Shimelis, S. Tesfay and T.J. Tsilo. 2016b. Screening of bread wheat genotypes for drought tolerance using phenotypic and proline analyses. Front Plant Sci 7: 1276.

Newcomb, M., M. Acevedo, H.E. Bockelman, G. Brown-Guedira, B.J. Goates, Y. Jin, et al. 2013. Field resistance to the Ug99 race group of the stem rust pathogen in spring wheat landraces. Plant Dis 97: 882-890.

Newcomb, M., P.D. Olivera, M.N. Rouse, L.J. Szabo, J. Johnson, S. Gale, et al. 2016. Kenyan isolates of Puccinia graminis f. sp. tritici from 2008 to 2014: Virulence to SrTmp in the Ug99 race group and implications for breeding programs. Phytopathology 106: 729-736.

Oliver, R., X. Cai, T. Friesen, S. Halley, R. Stack and S. Xu. 2008. Evaluation of Fusarium head blight resistance in tetraploid wheat (Triticum turgidum L.). Crop Sci 48: 213-222.

Ott, J., J. Wang and S.M. Leal. 2015. Genetic linkage analysis in the age of whole-genome sequencing. Nat Rev Genet 16: 275.

Pantalião, G.F., M. Narciso, C. Guimarães, A. Castro, J.M. Colombari, F. Breseghello, et al. 2016. Genome wide association study (GWAS) for grain yield in rice cultivated under water deficit. Genetica 144: 651-664.

Pe'er, I., R. Yelensky, D. Altshuler and M.J. Daly. 2008. Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. Genetic Epidemiology 32: 381-385.

Piepho, H.-P. 2001. A quick method for computing approximate thresholds for quantitative trait loci detection. Genetics 157: 425-432.

Pretorius, Z.A., C.M. Bender, B. Visser and T. Terefe. 2010. First report of a *Puccinia graminis* f. sp. *tritici* race virulent to the *Sr24* and *Sr31* wheat stem rust resistance genes in South Africa. Plant Dis 94: 784-784.

Pretorius, Z.A., R.P. Singh, W.W. Wagoire and T.S. Payne. 2000. Detection of virulence to wheat stem rust resistance gene *Sr31* in *Puccinia graminis f.sp. tritici* in Uganda. Plant Dis 84: 203-203.

Price, A.L., R.M. Patterson, R.M. Plenge, M.E. Weinblatt, N.A. Shadick and D. Reich. 2006. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 38: 904-909.

Rafalski, A. 2002. Applications of single nucleotide polymorphisms in crop genetics. Curr Opin Plant Biol 5: 94-100.

Razali, N.M. and Y.B. Wah. 2011. Power comparisons of shapiro-wilk, kolmogorov-smirnov, lilliefors and anderson-darling tests. J Stat Mod Analyt 2: 21-33.

Roelfs, A.P., R.P. Singh and E.E. Saari. 1992. Diseases of wheat: concepts and methods of disease management. CIMMYT, Mexico City.

Rouse, M., R. Wanyera, P. Njau and Y. Jin. 2011a. Sources of resistance to stem rust race Ug99 in spring wheat germplasm. Plant Dis 95: 762-766.

Rouse, M.N., E.L. Olson, B.S. Gill, M.O. Pumphrey and Y. Jin. 2011b. Stem rust resistance in Aegilops tauschii germplasm. Crop Sci 51: 2074-2078.

Sachs, J.D. 2012. From millennium development goals to sustainable development goals. The Lancet 379: 2206-2211.

Sánchez-Martín, J., L.A. Mur, D. Rubiales and E. Prats. 2012. Targeting sources of drought tolerance within an Avena spp. collection through multivariate approaches. Planta 236: 1529-1545.

Sánchez-Martín, J., N. Rispail, F. Flores, A.A. Emeran, J.C. Sillero, D. Rubiales, et al. 2017. Higher rust resistance and similar yield of oat landraces versus cultivars under high temperature and drought. Agron Sustain Dev 37: 3.

Sax, K. 1923. The association of size differences with seed-coat pattern and pigmentation in Phaseolus vulgaris. Genetics 8: 552.

Schwarz, G. 1978. Estimating the dimension of a model. Ann Stat 6: 461-464.

Shi, W., C. Hao, Y. Zhang, J. Cheng, Z. Zhang, J. Liu, et al. 2017. A combined association mapping and linkage analysis of kernel number per spike in common wheat (Triticum aestivum L.). Front Plant Sci 8: 1412.

Shiferaw, B., M. Smale, H.-J. Braun, E. Duveiller, M. Reynolds and G. Muricho. 2013. Crops that feed the world 10. Past successes and future challenges to the role played by wheat in global food security. J Food Secur 5: 291-317.

Singh, R.P., D.P. Hodson, J. Huerta-Espino, Y. Jin, S.A. Herrera-Foessel, R. Wanyera, et al. 2008. Will stem rust destroy the world's wheat crop? Advances in Agronomy 98: 271-309.

Solis, J., A. Gutierrez, V. Mangu, E. Sanchez, R. Bedre, S. Linscombe, et al. 2018. Genetic mapping of quantitative trait loci for grain yield under drought in rice under controlled greenhouse conditions. Front Chem 5: 129.

Soller, M., T. Brody and A. Genizi. 1976. On the power of experimental designs for the detection of linkage between marker loci and quantitative loci in crosses between inbred lines. Theor Appl Genet 47: 35-39.

Storey, J.D. and R. Tibshirani. 2003. Statistical significance for genomewide studies. Proc Nat Acad Sci 100: 9440-9445.

Tester, M. and P. Langridge. 2010. Breeding technologies to increase crop production in a changing world. Science 327: 818-822.

Thoday, J. 1961. Location of polygenes. Nature 191: 368-370.

Tilman, D., C. Balzer, J. Hill and B.L. Befort. 2011. Global food demand and the sustainable intensification of agriculture. Proc Nat Acad Sci 108: 20260-20264.

Valdez, V.A., P.F. Byrne, N.L. Lapitan, F.B. Peairs, A. Bernardo, G. Bai, et al. 2012. Inheritance and genetic mapping of Russian wheat aphid resistance in Iranian wheat landrace accession PI 626580. Crop Sci 52: 676-682.

Varella, A.C., D.K. Weaver, J.P. Cook, N.K. Blake, M.L. Hofland, P.F. Lamb, et al. 2017. Characterization of resistance to the wheat stem sawfly in spring wheat landrace accessions from targeted geographic regions of the world. Euphytica 213: 153.

Vavilov, N.I. 1992. Origin and geography of cultivated plantsCambridge University Press, Cambridge.

Wallach, D. and B. Goffinet. 1989. Mean squared error of prediction as a criterion for evaluating and comparing system models. Ecol Model 44: 299-306.

Wang, R., T. Gordon, D. Hole, W. Zhao, K. Isham, J.M. Bonman, et al. 2019. Identification and assessment of two major QTL for dwarf bunt resistance in winter wheat line ‘IDO835’. Theor and Appl Genet: 1-12..

Wang, S.D., D. Wong, K. Forrest, A. Allen, S. Chao, E. Huang, 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.

West, P.C., J.S. Gerber, P.M. Engstrom, N.D. Mueller, K.A. Brauman, K.M. Carlson, et al. 2014. Leverage points for improving global food security and the environment. Science 345: 325-328.

Williams, K.A. 2005. An overview of the US National Plant Germplasm System's exploration program. HortScience 40: 297-301.

Winkler, L.R., J. Michael Bonman, S. Chao, B. Admassu Yimer, H. Bockelman and K. Esvelt Klos. 2016. Population structure and genotype–phenotype associations in a collection of oat landraces and historic cultivars. Front Plant Sci 7: 1077.

Xie, W., W. Xiong, J. Pan, T. Ali, Q. Cui, D. Guan, et al. 2018. Decreases in global beer supply due to extreme drought and heat. Nat Plants 4: 964-973.

Xu, S., Y. Jin, D. Klindworth, R.-C. Wang and X. Cai. 2009. Evaluation and characterization of seedling resistances to stem rust Ug99 races in wheat–alien species derivatives. Crop Sci 49: 2167-2175.

Yang, S., B. Vanderbeld, J. Wan and Y. Huang. 2010. Narrowing down the targets: towards successful genetic engineering of drought-tolerant crops. Mol Plant 3: 469-490.

Yu, J.-B., G.-H. Bai, S.-B. Cai, Y.-H. Dong and T. Ban. 2008. New Fusarium head blight-resistant sources from Asian wheat germplasm. Crop Sci 48: 1090-1097.

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

Zeng, Z.-B. 1993. Theoretical basis for separation of multiple linked gene effects in mapping quantitative trait loci. Proc Nat Acad Sci USA 90: 10972-10976.

Zeng, Z.B. 1994. Precision mapping of quantitative trait loci. Genetics 136: 1457-1468.

Zhang, X., H. Pan and G. Bai. 2012. Quantitative trait loci responsible for Fusarium head blight resistance in Chinese landrace Baishanyuehuang. Theor Appl Genet 125: 495-502.

Zhao, C., B. Liu, S. Piao, X. Wang, D.B. Lobell, Y. Huang, et al. 2017. Temperature increase reduces global yields of major crops in four independent estimates. Proc Nat Acad Sci USA 114: 9326-9331.

Zhao, J., H. Sun, H. Dai, G. Zhang and F. Wu. 2010. Difference in response to drought stress among Tibet wild barley genotypes. Euphytica 172: 395-403.

Zhao, K., C.-W. Tung, G.C. Eizenga, M.H. Wright, M.L. Ali, A.H. Price, et al. 2011. Genome-wide association mapping reveals a rich genetic architecture of complex traits in Oryza sativa. Nat Commun 2: 467.

CHAPTER 2:

GENETIC CHARACTERIZATION AND GENOME-WIDE ASSOCIATION MAPPING FOR DWARF BUNT RESISTANCE IN BREAD WHEAT ACCESSIONS FROM THE USDA NATIONAL SMALL GRAINS COLLECTION

ABSTRACT

Dwarf bunt (DB), caused by *Tilletia controversa* J.G. Kühn, can significantly reduce grain yield and quality on autumn-sown wheat in regions with prolonged snow cover. DB can be managed with the use of resistant cultivars. The objectives of the present study were to characterize DB resistance in a large set of bread wheat accessions from the National Small Grains Collection (NSGC) and use a genome-wide association study (GWAS) approach to identify genetic loci associated with DB resistance. A total of 292 accessions were selected using historical DB resistance data recorded across many trials and years in the Germplasm Resources Information Network (GRIN) and re-tested for DB resistance in replicated field nurseries in Logan, UT, in 2017, 2018 and 2019. Ninety-eight accessions were resistant with DB normalized incidence ≤10%, and twenty-eight of these were highly resistant with DB normalized incidence ≤1% in both GRIN and the field nurseries. Based on the presence of marker haplotypes of the four published dwarf bunt QTL on 6DS, 6DL, 7AL, and 7DS, highly resistant accessions identified in this study may provide novel resistance and should be further evaluated. This study validated one previously identified QTL on 6DS and identified an additional locus on 6DS. These loci explained 9-15% of the observed phenotypic variation. The resistant accessions and molecular markers identified in the

present study may provide valuable resources for characterization and deployment of DB resistance in bread wheat.

INTRODUCTION

Bread wheat (*T. aestivum* L.) is an important food staple and 772 million t were harvested in 2017 globally (FAOSTAT 2019). Dwarf bunt (DB), caused by the basidiomycete *Tilletia controversa* J.G. Kühn [as '*contraversa*'] in L. Rabenhorst (Kühn 1874), and common bunt (CB), caused by two closely related fungi *Tilletia caries* (DC.) Tul. & C. Tul. [syn. *T. tritici* (Bjerk.) G. Wint.] and *Tilletia laevis* J. G. Kühn [syn. *T. foetida* (Wallr.) Liro], are destructive diseases of bread wheat and durum wheat (*T. turgidum* subsp. *durum* Desf.) (Goates 1996). While these three pathogens vary slightly in their spore morphology and etiology, they are closely related with similar modes of infection and means of control. DB and CB differ slightly in that *T. controversa* infects autumn-sown wheat and requires several months of snow cover for teliospore germination on the soil surface, whereas *T. caries* and *T. laevis* primarily infect spring planted wheat from spores in the soil.

Initiation of DB and CB begins when dikaryotic infection hyphae penetrate emerging seedlings thereby infecting the developing apical meristem (Kollmorgen and Ballinger 1987). The resulting systemic infection is often cryptic until flowering, when the fungal hyphae invades and replaces developing ovaries with darkly pigmented teliospores that comprise a fungal sorus or bunt ball (Goates 1996; Castlebury et al. 2005). Yield losses due to DB and CB can exceed 80% and trimethylamine emitted by the teliospores causes a fetid, rotting fish odor which reduces flour quality (Goates 1996; Castlebury et al. 2005).

Difenoconazole, a seed treatment fungicide, effectively controls both diseases without causing yield reductions or phytotoxicity (Keener et al. 1995; Goates 1996) though genetic resistance offers a cost effective compliment to seed treatments particularly in organic production systems. Most of the wheat landraces in the USDA National Small Grains Collection (NSGC) were screened for bunt resistance over the past 30 years and resistance was found primarily in germplasm originating from regions in Iran, Macedonia, Montenegro, Serbia, and Turkey (Bonman et al. 2006). However, resistance was relatively rare. Among 10,759 landrace accessions tested for CB resistance only 597 (5.5%) were resistant, and of 8,167 landrace accessions tested for DB resistance only 104 (1.3%) were resistant (Bonman et al. 2006).

DB and CB resistance is putatively controlled by gene-for-gene interactions and it is assumed that the same genes confer resistance to both diseases (Hoffman and Metzger 1976; Goates 2012). An expanded set of bunt differential wheat accessions representing 16 *Bt* genes was developed to elucidate host-pathogen interactions (Goates 2012). Using these *Bt* differentials, Goates (2012) found 19 pathogenic races of *T. controversa*, 36 races of *T. caries*, and 15 races of *T. laevis,* and determined that *Bt8* (PI 554120), *Bt11* (PI 554119), and *Bt12* (PI 119333) were broadly effective against most races of DB and CB.

Genomic tools in wheat including dense molecular marker arrays with annotations (Wang et al. 2014), genotyping-by-sequencing, and reference genome sequences (IWGSC 2018) have enabled the identification of genetic loci underpinning DB and CB resistance (Table A2.1). Linkage mapping (Chen et al. 2016; Singh et al. 2016; Steffan et al. 2017) and association mapping techniques (Bhatta et al. 2018; Mourad et al. 2018) have located bunt

resistance loci on 19 wheat chromosomes. Identifying markers tightly linked to resistance will enable the discovery of additional resistance genes and introgression of multiple resistance genes into adapted cultivars.

The NSGC is a worldwide collection of the small grains and contains 42,544 bread wheat accessions. Of these, 19,378 accessions have been systematically characterized for DB resistance since the early 1980s, and only 129 (0.7%) are classified as resistant based on a DB incidence threshold of ≤10% proposed by Goates (2012). The purpose of this study was to: 1) verify the DB resistance in the NSGC bread wheat accessions with replicated field trials, and 2) identify genetic loci associated with DB resistance using a genome-wide association study (GWAS) approach.

MATERIALS AND METHODS

Plant materials

DB resistant and susceptible accessions were selected for this panel based on data from the U.S. National Germplasm System online database: Germplasm Resources Information Network (GRIN), accessed at https://npgsweb.ars-grin.gov/gringlobal/search.aspx. Using a resistance threshold of ≤10% disease incidence relative to the susceptible check (Goates 2012), only 129 GRIN accessions were classified as DB resistant. An additional seven accessions with DB incidence below 13% were also included in the panel for a total of 136 bread wheat accessions classified as resistant for the GWAS. In an attempt to mitigate the effects of population structure on the GWAS, one susceptible accession from the same

geographic region as each resistant accession was selected. For example, PI 470452 was classified as resistant and originated in Agri Province, Turkey; therefore, a susceptible accession from Agri Province, Turkey, PI 470470, was also selected. Additionally, the bunt differentials (Goates 2012), including *Bt0* through *Bt15*, *Btp*, and PI 173438 (unknown *Bt*), and two known susceptible winter cultivars 'Wanser' (CItr 13844) and 'Cheyenne' (CItr 8885), were also included in the GWAS panel. Supplementary File A2.1 lists the accession number, name, taxon, geographic origin, improvement status, pedigree, and DB incidence for each of the 292 accessions.

Field trials

Since the 1980's, GRIN DB normalized incidence (NI) relative to the susceptible check cultivar 'Cheyenne' in each trial was collected from NSGC accessions grown at the Green Canyon USDA-ARS disease screening nursery in Logan, UT (approximately 3 km east of Logan: 41°46'21.05"N, 111°46'52.68"W. elevation 1450 m). DB field trials conducted in 2017, 2018, and 2019 were evaluated near the Green Canyon site at the Utah State University (USU) Research Farm in Logan, UT (41°45'46.46"N, 111°48'54.98"W, elevation 1400 m). USU field trials were sown with a head row planter on October 10, 2016, September 27, 2017 and September 18, 2018 with one accession per 1-m row and two replications in 2017 and 2018 and one replication in 2019. Each row was inoculated after seedling emergence on November 4, 2016, October 24, 2017, and November 6, 2018 with approximately 100 ml of a concentrated DB teliospore suspension ($2x10^6$ spores ml^{-1} water). A composite of teliospores from infected spikes previously collected in the USU DB nursery

were used for the inoculations. Disease incidence was assessed on fully mature adult plants, Zadoks stage 94 (Zadoks et al. 1974), on August 3, 2017, July 26, 2018, and August 6, 2019 by counting the number of spikes where at least one floret was infected, and dividing by the total number of spikes in the row. DB incidence per replicate was normalized to the average of the six plots of the susceptible cultivar, Wanser. Accession DB incidence and NI are reported in Supplementary File A2.1, and the mean DB NI for each field trial can be accessed through GRIN.

Molecular marker assessment

A modified DNA CTAB protocol was used to extract genomic DNA from seedlings at the 2-3 leaf stage (Babiker et al. 2015). A 2 cm segment of leaf tissue was placed into 96 well Corning® Costar® tubes (Corning, NY, U.S.) and macerated in CTAB extraction buffer with a bead grinder. The aqueous layer was separated in chloroform, extracted, and the precipitate was washed with isopropanol and then ethanol. Resulting DNA pellets were suspended in Tris (10 mM) and sent to the USDA-ARS Small Grains Genotyping Laboratory in Fargo, ND where samples were genotyped using the 90K iSelect SNP assay as described by the manufacturer (Illumina, San Diego, CA). Allele clustering was completed using Genome Studio v.2.0.2 (Illumina) and the resulting set of 41,511 polymorphic SNPs were exported to JMP Genomics v.9.0 (SAS Institute Inc., Cary, NC, USA) for filtering. Markers were excluded if minor allele frequency (MAF) was <4%, or missing data was >10%. Heterozygous calls were also removed. Accessions were classified as duplicates and removed if they were ≥ 99.7% identical across all polymorphic SNPs. A final group of 246

bread wheat accessions was selected and 19,281 SNP markers were aligned with the physical wheat annotation (IWGSC 2018) and used for subsequent marker-trait associations.

Statistical analyses

Unless stated otherwise, all statistical analyses were conducted using JMP® Genomics v. 9.0. By design, the DB NI had a bimodal distribution, and a Shipiro-Wilk normality test (Shapiro and Wilk 1965) of trial residuals indicated a significant ($P < 0.0001$) shift from normality. Similarly, square root and $\log_{10}$ transformations of the trials indicated significant ($P < 0.0001$) deviations from normality, and the untransformed DB NI data was used in all further analyses. A mixed model with genotype set as a fixed effect and trial as a random effect was used to calculate best linear unbiased estimates (BLUEs) for DB NI across trials and replications (Henderson 1975). Broad sense heritability (H^2) was calculated using the formula: $H^2 = \sigma^2_G/[\sigma^2_G + \sigma^2_{ExG}/r + \sigma^2_e/r]$ where σ^2_G is the genotypic variance, σ^2_{ExG} is the interaction variance between trial and genotype, σ^2_e is the residual variance, and r is the number of data sets (Hanson et al. 1956). Correlation coefficient estimates between trials were calculated using a Spearman's Rho nonparametric rank-sum correlation procedure.

Genome-wide linkage disequilibrium (LD) was calculated as r^2 values between each marker within chromosome groups (Figure A2.1). An IBS familial relationship matrix (k matrix) and heat map were generated using the Ward hierarchical clustering method (Ward Jr and Hook 1963) to explore potential subpopulations within the panel. STRUCTURE v.2.3.4 (Pritchard et al. 2000) and STRUCTURE HARVESTER (Earl and vonHoldt 2012) software packages were used to optimize the number of subpopulations (k). In

STRUCTURE, the burn-in iterations and Markov-chain Monte Carlo replications were set to 10,000, the admixture correlated model was selected, and five replicate iterations were performed. Proposed subpopulations with k between 1 and 10 were evaluated in STRUCTURE HARVESTER using the Evanno method (Evanno et al. 2005), and the number of subpopulations that corresponded with the highest Δk value was selected as the optimal model.

A principal component analysis with 10 principal components (PCs) was generated (Q matrix) to explore population stratification , and the resulting scree plot was used to estimate the optimal number of PCs that would explain the most variation in the models (Price et al. 2006). Bayesian information content (BIC) assessments (Burnham and Anderson 2004) were used to formally test the various association analysis models. Tested models included a general linear model (GLM) without corrections for K or Q (the naïve model), a GLM that corrected for population stratification with 2, 3 or 5 PCs, and mixed linear models (MLMs) that controlled for both familial relationships, as a random effect, and population stratification with 2, 3 or 5 PCs as a fixed effect (Yu et al. 2006). All models correcting for familial relationships performed better than the naïve model. A MLM with a kinship covariate matrix and 2 PCs had the lowest BIC value and was therefore chosen for further marker-trait association analysis.

Marker-trait associations between DB NI and SNP markers were conducted on trial means, and BLUEs. Resulting *P*-values were adjusted using an FDR multiple testing procedure (Benjamini and Hochberg 1995) and a significance threshold of $P \leq 0.05$ on FDR-adjusted *P*-values was used to identify SNP-trait associations for further analysis.

SNPs significantly associated with DB NI in any trial or BLUE were aligned with the Chinese Spring reference genome sequence v1.0 (IWGSC 2018) using IWGSC BLAST (Alaux et al. 2018) with the highest coverage and identity location available. To assess potentially linked SNPs, the most significant marker in each putative marker-trait association group were included one at a time in the MLM as covariates. Markers in high LD with the covariate marker were no longer significantly associated with DB resistance and were grouped with the covariate SNP group.

RESULTS

Field trials

Two susceptible check cultivars, Cheyenne and Wanser, showed a high incidence of DB in both USU field trials. The mean DB incidence for Wanser was 63.9% in 2017, 82.8% in 2018, and 67.3% in 2019 while Cheyenne had a mean DB incidence of 79.2% in 2017, 84.6% in 2018 and 88.6% in 2019. Across the three trials, all differentials showed consistent responses except for the *Bt9* differential, which was classified as resistant in 2017 but susceptible in 2018 and 2019, and the *Bt5* differential which was classified as susceptible in 2017 and 2019 but resistant in 2018 (Table 2.1).

A mixed model ANOVA (Table A2.2) found no significant trial effect, but there was a significant genotype and genotype-by-trial effect ($P < 0.0001$). Broad-sense heritability (H^2) for DB NI was estimated at 0.93. Best linear unbiased estimates derived from the mixed model of DB NI across trials are listed in Supplementary File A2.1. By design, the field trials were composed of approximately 50% resistant and 50% susceptible accessions as

classified based on GRIN data (Fig. 2.1). USU field trials produced a similar response, with 50.3% showing resistance in 2017, 50.7% showing resistance in 2018 and 45.2% with resistance in 2019 (Table 2.2). Accessions classified as susceptible based on GRIN showed a wide array of disease incidence in the field trials (Fig. 2.1). Most of the accessions classified as resistant based on GRIN data, yet susceptible in 2017, 2018 or 2019, were breeding lines from the U.S. Another group of accessions, about half being landraces from Turkey, showed the opposite reaction; they were susceptible based on GRIN data, but resistant in the USU field trials (Supplementary File A2.1).

A Spearman's rank-sum nonparametric correlation was used to measure the degree of similarity between and amongst the GRIN, 2017 2018 and 2019 USU field trial means and BLUEs. Correlation coefficients (r^2) between data sets ranged from 0.70 and 0.93, and all estimates were significant at $P < 0.0001$. GRIN was correlated with the 2017, 2018 and 2019 trials, and BLUEs with correlation coefficients of 0.76, 0.77, 0.70 and 0.85, respectively. The correlation coefficient between the 2017, 2018 and 2019 USU trials was 0.88, 0.76 and 0.78, respectively. In the 2017 USU field trials, the r^2 between replications was 0.86 and in 2018 the r^2 was 0.87 between the two replications.

There were 98 accessions that were resistant with a DB NI $\leq$10% across GRIN and 2017, 2018 and 2019 USU trial means (Supplementary File A2.1). Of these, 28 were highly resistant with a DB NI $\leq$1% across all trials (Table 4). These highly resistant accessions included eight Turkish landraces, and 14 U.S. lines with Turkish landraces in their pedigree. The remaining six highly resistant accessions were landraces from Serbia (1), Montenegro (1), Iran (1), and three breeding lines from the U.S (Table 2.3).

Population structure

There were 44 accessions that were ≥99.7% identical across the 19,281 SNPs. These duplicate and near-duplicate accessions originated from similar geographic regions and had similar DB NI across data sets (Supplementary File A2.1), and were removed for further analyses. Genetic similarity among the 246 non-duplicated accessions ranged from 53% to 99% with a mean similarity of 67%.

Based on the STRUCTURE HARVESTER Δk values there were six distinct subpopulations (k = 6) in the panel, and these groupings were supported by visual assessment of the Ward hierarchical clustering heat map and principal component analysis (Fig. 2.2A and 2.2B). The first three principal components explained 22.2% of the total variation, and 12.7%, 5.8% and 3.7% of the variation was explained by PC1, PC2 and PC3, respectively (Fig. 2.2B). Subpopulations based on the marker data corresponded well with geographic origin (Table 2.4). Subpopulation 1 and 4 consisted primarily of breeding lines and cultivars from the U.S. Subpopulation 2 consisted of accessions from Turkey and breeding lines from the U.S. Serbian landraces predominated in subpopulation 3, while landraces from Iran were primarily located in subpopulation 5. Landraces from Hakkari province, Turkey and breeding lines from the U.S. that had Turkish landraces in their pedigree, were grouped into subpopulation 6 (Table 2.4). The bunt differentials were distributed across all the subpopulations (Table 2.1).

The BLUE DB NI estimate for the entire panel was 47.3% (Fig 2C), and BLUE values for each subpopulation differed significantly at $P < 0.0001$. Subpopulation 6 had the

lowest mean BLUE DB NI of 8.5%, and subpopulation 1 had the highest mean BLUE DB NI of 74.0%. Of the 98 accessions that were resistant across trials (Table 2.3), 7% were in subpopulation 6, 13% were in subpopulation 2, with the remainder in subpopulations 1, 3, 4, and 5. Of the 28 highly resistant accessions, 75% were in subpopulation 6, 7% were in subpopulation 2 and 3, and 4% were in each of the subpopulations 1, 4 and 5.

Linkage disequilibrium

Genome-wide marker-pair r^2 correlations between 19,281 SNPs were plotted as a function of intrachromosomal inter-marker genetic distance (Fig. A2.1). A median r^2 of 1 was found between SNP markers that were completely linked with an inter-marker physical distance of 0 Mbp. LD median r^2 decreased to 0.1 at an inter-marker distance of 0.1 to 1 Mbp indicating an LD decay rate of 90% over the 1 Mbp interval.

A smoothing spline curve with lambda equal to 10,000 was fit to the LD scatter plot to determine a genome-wide QTL confidence interval (Supplementary File A2.1). Others (Maccaferri et al. 2015; Liu et al. 2017), have used an LD of $r^2 = 0.3$ as a threshold for genome-wide QTL confidence intervals in wheat. In the present study, the largest spline curve r^2 value was 0.45. When the smoothing spline curve was set to $r^2 = 0.3$ the physical distance was 0.67 Mbp, and when the curve was set to $r^2 = 0.1$ the distance was 6.80 Mbp.

Marker-trait associations

After controlling for kinship and population stratification, GWAS revealed 4 SNPs significantly (FDR-adjusted $P < 0.05$) associated with DB incidence in at least one trial or

BLUE (Table 2.5, Figs. 2.3 and A2.3, and Supplementary File A2.3). FDR-adjusted negative $\log_{10}$ *P*-values for BLUE DB NI from these six marker-trait association groups ranged from 1.7 to 5.1, phenotypic variance (R^2) ranged from 0.09 to 0.15, and average DB NI BLUE values for accessions carrying resistance alleles ranged from 16.1 to 40.8 (Table 2.5). One marker-trait association group represented by 2 SNPs on chromosome 6DS, was significant in three of the data sets (Table 2.5). Marker-trait association groups aligned with the 246 bread wheat accessions used for the GWAS (Supplementary File A2.4) show a corresponding decrease in DB NI as the number of resistant allele haplotypes increases (Table A2.3).

DISCUSSION

Uniform DB infection requires specific environmental conditions that include several weeks of stable cool soil temperatures, a moist environment at the soil surface, and low light levels. These conditions are most reliably provided by continuous snow cover and are critical for teliospore germination (Chen et al. 2016). The two susceptible check cultivars, Wanser and Cheyenne, showed high DB incidence in all field trials indicating that the environmental conditions in these years favored infection by the DB pathogen.

There were 28 highly resistant accessions with a DB NI $\leq$1% across all data sets. Twenty-one of these highly resistant accessions either originated in Turkey or have Turkish landraces in their pedigrees. Similarly the four bunt differentials that were highly resistant across trials, PI 554119 (*Bt11*), PI 119333 (*Bt12*), PI 173437 (*Btp*), and PI 173438

(unknown Bt), all either originated in Turkey or had a Turkish landrace in their pedigree. PI 119333 and PI 173437 had similar haplotype profiles (Table A2.6); and they shared SNP marker haplotypes with some of the other highly resistant accessions (Supplementary File A2.4). For instance, PI 119333 (*Bt12*) shares a similar haplotype profile to six other highly resistant accessions and PI 173438 (with unknown *Bt*) shares a similar profile with two other highly resistant accessions including PI 476212 (Table 2.6).

Based on the pedigree analysis (Supplementary File A2.1), many highly resistant breeding lines are derived from resistant Turkish landraces. PI 178383 and PI 476212 are in the pedigrees of several DB resistant cultivars, such as 'Weston', 'DW', 'Golden Spike', and 'UI Silver' (Hole et al. 2002). However, some highly resistant landraces PI 345106 from Serbia, PI 345428 from Montenegro, and PI 627677 from Gilan province, Iran have unique haplotypes and geographic origins (Table 2.6). Therefore, mapping the DB resistance within these unexploited resistance sources is an important step towards future molecular breeding for DB resistance.

In the present study, accessions were selected based on a DB NI resistance threshold of ≤10%. Other accessions with intermediate levels of resistance are of interest to geneticists and plant breeders as they may contain a complex of minor or partial resistance genes. Specifically, PI 362710 from Montenegro, PI 345480 from Serbia, and PI 636153 a breeding line from Idaho, U.S., had intermediate levels of DB resistance across data sets. Additionally, in the GRIN database, there are 976 bread wheat accessions that have a DB incidence recorded between 11 and 30%. Environmental conditions can make bunt disease incidence variable from one year to the next. Thus, to confirm the partial resistance that may

exist in these accessions, more research is warranted. Single-seed derived lines of each accession could be tested for multiple years in the field. Alternatively, molecular marker-assisted evaluation could be undertaken to identify accessions that do not carry known resistance QTL haplotypes. A quantitative PCR assay, like those developed for rust diseases (Admassu-Yimer et al. 2019), that reliably measures the degree of tissue colonization by the bunt pathogen could also provide a means for assessing partial resistance to the disease under greenhouse conditions.

Six subpopulations were selected in this panel of 246 bread wheat accessions based on Δk value optimization using STRUCTURE and STRUCTURE HARVESTER. These six subpopulations roughly corresponded to the geographic origins listed in GRIN (Table 2.5). We attempted to control for population relatedness by selecting both resistant and susceptible accessions from the same geographic area. Unfortunately, the subpopulations differed significantly in their levels of DB incidence (Fig. 2.2C) which could affect marker-trait associations. Specifically, those accessions in subpopulation 6 which corresponded with a Hakkari province, Turkey origin, had significantly lower DB NI values than the other five subpopulations. Investigators may need to limit the origin of accessions to one region or locality to better balance population structure when designing future bunt GWAS. For instance, it might be of interest to examine all landrace accessions from Turkey as one study, and all landrace accessions from Iran as a separate study.

Broad-sense heritability, 0.93, was high for DB NI in this panel. Others have also reported high broad-sense heritability estimates for bread wheat resistance to dwarf bunt, 0.88 to 0.93, (Chen et al. 2016) and common bunt, 0.58 to 0.78 (Bhatta et al. 2018; Mourad

et al. 2018). Although the broad-sense heritability estimate and correlations between replications and years were high in this study, there were no significant SNPs that were consistent between data sets and met the FDR-adjusted *P*-value threshold of 0.05 (Supplementary File A2.3). Less stringent significance thresholds have been used in other bread wheat GWAS panels with small population sizes (Zegeye et al. 2014; Gao et al. 2016). A less stringent threshold would allow identification of additional marker-trait associations in this panel (Supplementary File A2.2), but would increase the likelihood of false positive associations.

Of the two marker-trait associations that were significant in the present study (Table 2.5), only one corresponds with a previously reported QTL for DB or CB resistance (Table A2.1). Menzies et al. (2006) and Singh et al. (2016) found a QTL on 6DS with a peak marker at 6.17 Mbp, which is likely the same QTL identified as *DB-6D2* in this study. *DB-6D2* is composed of two SNP markers and is most significantly associated with resistance identified in the present study (Table 2.5). Accessions containing the resistance alleles had a mean DB NI BLUE value of 16.3 (Table 2.5). The *Bt10* differential, PI 178383 and another 30 accessions in this GWAS panel have this resistance-associated haplotype (Supplementary File A2.4). Based on the physical position, this QTL spanned a relatively narrow section of the chromosome from 6.97 to 7.29 Mbp, which is within the flanking position of the *Bt10* gene (Menzies et al. (2006). Markers in this region can be developed and used in marker assisted selection, but must first be validated in bi-parental populations.

Additionally, Menzies et al. (2006) hypothesized that the *Bt10* QTL contributed by the bread wheat cultivar ‘AC Cadillac’ was closely linked with effective Ug99 stem rust

resistance genes on 6DS, *SrTmp* or *SrCad* (Hiebert et al. 2016; Kassa et al. 2016). To determine whether *Bt10* confers a stem rust resistance phenotype like *SrCad* or *SrTmp*, PI 554118 (*Bt10*) and PI 178383 were screened with several Ug99 stem rust races. These two accessions were resistant to many of the same stem rust races as were lines containing *SrTmp* and *SrCad* (unpublished data). Further studies are needed to determine if the 6DS region contains one or more genes that confer resistance to DB, CB and wheat stem rust.

One of the markers in the *DB-6D2* locus group was shown to be on 6A according to the Chinese Spring Reference Sequence (Table 2.5, Fig. 2.3). The marker on 6A, IWB64081, and the marker on 6D at 1.77 Mbp, IWB21614, were in high LD with one another and the marker on 6A was previously shown to be on 6D (Wang et al, 2014). We therefore assumed both of these markers are from the same marker-trait association group and combined them into one QTL.

Two major QTL, *Q.DB.ui-7DS* (Chen et al. 2016) and *Q.DB-6DL* (Steffan et al. 2017; Wang et al. 2019) and *Q.DB.ui-7AL* (Wang et al. 2019), that were previously reported in bi-parental populations were not detected in this study. The QTL *Q.DB.ui-7DS* was reported in the 'Rio Blanco'/'IDO444' population on 7DS with a peak marker, *wPt-2565,* at 5.9 Mbp near the telomere (Chen et al. 2016). Based on pedigree information, the resistance in IDO444 was thought to be derived from PI 476212, the same parent contributing resistance in cv. 'Blizzard'. PI 476212 was initially selected for snow mold and DB resistance (Sunderman et al. 1986) and is in the pedigree of resistant cultivars 'DW' (PI 620629), 'Bonneville' (PI 557015), 'Golden Spike' (PI 614813), and 'UI Silver' (PI 658467). PI 476212 was highly resistant in the present study and was 99.99% similar to PI 173438

(unknown *Bt*) across the 19,281 SNPs, but the 7DS QTL reported by Chen et al. (2016) was not detected, possibly because too few accessions with this QTL were included in the present study. A haplotype analysis using SNPs in the 7DS region indicated that three of the highly resistant landraces in addition to the *Bt12* differential and PI 476212 may contain the 7DS QTL (Table 2.6, Supplementary File A2.4)

Similarly, *Bt9* has been mapped to 6DL between 172.8 to 175.9 Mbp in a population derived from the *Bt9* differential PI 554099 (Steffan et al. 2017). However, our GWAS did not detect any markers significantly associated with 6DL in any data sets. Using a bi-parental mapping population derived from a University of Idaho wheat breeding line 'IDO835', Wang et al. (2019) found two QTL for DB resistance, one on 6DL corresponding with the *Bt9* locus, and one on 7AL. We used the resistant haplotypes for both loci to find accessions that contain these QTL (Table 2.6, Supplementary File A2.4). The *Bt9* differential and PI 178383 contained the haplotype profile for the 6DL locus, but none of the highly resistant accessions contained the 6DL or 7AL haplotype (Table 2.6, Supplementary File A2.4).

Aside from the possible low frequency of certain known loci in our GWAS panel, SNP maker filtering could also have reduced detection of known loci. SNP markers were filtered at a MAF threshold of 4% and any marker with fewer than thirteen individuals in each allelic state would have been filtered before analysis. This filtering threshold could mask SNP-trait associations that were present at low frequencies. To find such QTL, bi-parental populations could be developed using resistant accessions from the panel that lack alleles for the previously identified QTL.

Several marker-trait groups were associated with specific subpopulations (Table A2.4). For instance, 48% of accessions with the *QDB-6D2* resistant haplotype are in subpopulation 4 (Table A2.4 and Supplementary File A2.4). All the highly resistant accessions and 12 of the bunt differentials contained the *DB-6D1* haplotype group. Conversely, the resistant haplotype for *DB-6D2* was strongly associated resistance (Table 2.5); however none of the highly resistant landrace accessions contained this haplotype (Table 2.6, Supplementary File A2.4).

CONCLUSIONS

The present study evaluated the DB responses recorded in the GRIN database for 292 wheat accessions rated in three field trials and identified 98 accessions that were resistant and 28 accessions that were highly resistant across all three years of USU field trials and in GRIN. Additionally, four SNP markers associated with DB resistance were identified, one marker-trait association group on 6D was consistent across several data sets, and one marker-trait association group on chromosome 6D was not previously reported. Of the highly resistant landrace accessions, six have novel resistance haplotype profiles. These resistant accessions and haplotype regions can be used to confirm resistance loci in bi-parental mapping populations for introgression into advanced wheat breeding lines.

REFERENCES

Admassu-Yimer B, Gordon T, Bonman JM, Klos KE (2019) Development and validation of a quantitative PCR assay method of assessing relative resistance of oat (*Avena sativa*) to crown rust (*Puccinia coronata* f.sp. *avenae*). Plant Pathol 68:669-677

Alaux M, Rogers J, Letellier T, Flores R, Alfama F, Pommier C, Mohellibi N, Durand S, Kimmel E, Michotey C, al. e (2018) Linking the International Wheat Genome Sequencing Consortium bread wheat reference genome sequence to wheat genetic and phenomic data. Genome Biol 19:111

Babiker EM, Gordon TC, Chao S, Newcomb M, Rouse MN, Jin Y, Waynera R, Acevedo M, Brown-Guedira G, Williamson S, Bonman JM (2015) Mapping resistance to the Ug99 race group of the stem rust pathogen in a spring wheat landrace. Theor Appl Genet 128:605-612

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

Bhatta M, Morgounov A, Belamkar V, Yorgancilar A, Baenziger PS (2018) Genome-wide association study reveals favorable alleles associated with common bunt resistance in synthetic hexaploid wheat. Euphytica 214:200-210

Bonman JM, Bockelman HE, Goates BJ, Obert DE, McGuire PE, Qualset CO, Hijmans RJ (2006) Geographic distribution of common and dwarf bunt resistance in landraces of *Triticum aestivum* subsp. *aestivum*. Crop Sci 46:1622-1629

Burnham KP, Anderson DR (2004) Multimodel inference: understanding AIC and BIC in model selection. Sociol Methods Res 33:261-304

Castlebury LA, Carris LM, Vánky K (2005) Phylogenetic analysis of *Tilletia* and allied genera in order Tilletiales (Ustilaginomycetes; Exobasidiomycetidae) based on large subunit nuclear rDNA sequences. Mycologia 94:888-900

Chen J, Guttieri MJ, Zhang J, Hole D, Souza E, Goates BJ (2016) A novel QTL associated with dwarf bunt resistance in Idaho 444 winter wheat. Theor Appl Genet 129:2313-2322

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

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

FAOSTAT (2019) Production and yield quantities of cereal grains. Food and agriculture organization of the United Nations

Gao L, Turner MK, Chao S, Kolmer J, Anderson JA (2016) Genome wide association study of seedling and adult plant leaf rust resistance in elite spring wheat breeding lines. PLoS One 11:e0148671

Goates BJ (1996) Common Bunt and Dwarf Bunt. In: Wilcoxson RD, Saari EE (eds) Bunt and Smut Diseases of Wheat: Concepts and Methods of Disease Management. CIMMYT, D.F., Mexico, pp 12-25

Goates BJ (2012) Identification of new pathogenic races of common bunt and dwarf bunt fungi, and evaluation of known races using an expanded set of differential wheat lines. Plant Dis 96:361-369

Hanson CH, Robinson HF, Cornstock RE (1956) Biometrical studies of yield in segregating populations of Korean Lespedeza. Agron J 48:268-272

Henderson CR (1975) Best linear unbiased estimation and prediction under a selection model. Biometrics 31:423-447

Hiebert CW, Kassa MT, McCartney CA, You FM, Rouse MN, Fobert P, Fetch TG (2016) Genetics and mapping of seedling resistance to Ug99 stem rust in winter wheat cultivar Triumph 64 and differentiation of SrTmp, SrCad, and Sr42. Theor Appl Genet 129:2171-2177

Hoffman JA, Metzger RJ (1976) Current status of virulence genes and pathogenic races of the wheat bunt fungi in the northwestern USA. Phytopathology 66:657-660

Hole DJ, Clawson SM, Young SA, Roche D (2002) Registration of 'Golden Spike' wheat. Crop Sci 42:1376

IWGSC (2018) Shifting the limits in wheat research and breeding using a fully annotated reference genome. Science 361

Kassa M, You FM, Fetch TG, Fobert P, Sharpe A, Pozniak CJ, Menzies JG, Jordan MC, Humphreys G, Zhu T, Lou M-C, McCartney CA, Hiebert CW (2016) Genetic mapping of SrCad and SNP marker development for marker-assisted selection of Ug99 stem rust resistance in wheat. Theor and Appl Genet 129:1373-1382

Keener TK, Stougaard RN, Mathre DE (1995) Effect of winter wheat cultivar and difenoconazole seed treatment on dwarf bunt. Plant Dis 79:601-604

Kollmorgen JF, Ballinger J (1987) Detection and morphology of hyphae of common bunt fungi (*Tilletia laevis* and *T. tritici*) in wheat seedlings. Trans Br Mycol Soc 88:555-558

Kühn JG (1874) Tilletia contraversa Hedwigia 13:188-189

Liu W, Maccaferri M, Bulli P, Rynearson S, Tuberosa R, Chen X, Pumphrey M (2017) Genome-wide association mapping for seedling and field resistance to *Puccinia striiformis* f. sp. *tritici* in elite durum wheat. Theor Appl Genet 130:649-667

Maccaferri M, Zhang J, Bulli P, Abate Z, Chao S, Cantu D, Bossolini E, Chen X, Pumphrey M, Dubcovsky J (2015) A genome-wide association study of resistance to stripe rust (*Puccinia striiformis* f. sp. *tritici*) in a worldwide collection of hexaploid spring wheat (*Triticum aestivum* L.). G3: Genes Genom Genet 114:014563

Menzies JG, Knox RE, Popovic Z, Procunier JD (2006) Common bunt resistance gene *Bt10* located on wheat chromosome 6D. Can J Plant Sci 86:1409-1412

Mourad AMI, Sallam A, Belamkar V, Mahdy E, Bakheit B, El-Wafaa AA, Baenziger PS (2018) Genetic architecture of common bunt resistance in winter wheat using genome-wide association study. BMC Plant Biol 18:280-294

Price AL, Patterson RM, Plenge RM, Weinblatt ME, Shadick NA, Reich D (2006) Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 38:904-909

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

Shapiro SS, Wilk MB (1965) An analysis of variance test for normality. Biometrika 52:591-611

Singh A, Knox RE, DePauw RM, Singh AK, Cuthbert RD, Kumar S, Campbell HL (2016) Genetic mapping of common bunt resistance and plant height QTL in wheat. Theor Appl Genet 129:243-256

Steffan PM, Torp AM, Borgen A, Backes G, Rasmussen SK (2017) Mapping of common bunt resistance gene *Bt9* in wheat. Theor Appl Genet 130:1031-1040

Sunderman DW, Hoffman JA, O'Connell BT (1986) Registration of four winter wheat germplasm lines with resistance to dwarf bunt. Crop Sci 26:651-652

Wang R, Gordon T, Hole D, Zhao W, Isham K, Bonman JM, Goates BJ, Chen J (2019) Identification and assessment of two major QTL for dwarf bunt resistance in winter wheat line 'IDO835'. Theor and Appl Genet:1-12

Wang SD, Wong D, Forrest K, Allen A, Chao S, Huang E, Maccaferri M, Salvi S, Milner S, Cattievelli L, Mastrangelo AM, Whan A, Stephen S, Barker G, Wieseke R, Plieske J, Consortium IWGS, Lillemo M, Mather D, Apples R, Dolferus R, Brown-Guedira G, Korol A, Akhunova AR, Feuillet C, Salse J, Morgante M, Pozniak C, Ming-Cheng L, Dvorak J, Morell M, Dubcovsky J, Ganal M, Tuberosa R, Lawley C, Mikoulitch I, Cavanagh C, Edwards KE, Hayden M, Akhunov E (2014) Characterization of polyploid wheat genomic diversity using a high-density 90 000 single nucleotide polymorphism array. Plant Biotechnol J 12:787-796

Ward Jr JH, Hook ME (1963) Application of an hierarchical grouping procedure to a problem of grouping profiles. Educational and Psychological Measurement 23:69-81

Yu J, Pressoir G, Briggs WH, Vroh Bi I, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB, Kresovich S, Buckler ES (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet 38:203-208

Zadoks JC, Chang TT, Konzak CF (1974) A decimal code for the growth stages of cereals. Weed Res 14:415-421

Zegeye H, Rasheed A, Makdis F, Badebo A, Ogbonnaya FC (2014) Genome-wide association mapping for seedling and adult plant resistance to stripe rust in synthetic hexaploid wheat. PLoS One 9:e105593

TABLES AND FIGURES

Table 2.1 Bunt differential lines and known susceptible and resistant sources showing subpopulations, and DB normalized incidence across five data sets[a].

Accession	Name	Bt gene	Subpopulation	GRIN	2017	2018	2019	BLUE
CItr 8885	Cheyenne	Susceptible	4	100	119.1	102.2	131.6	116.2
CItr 13844	Wanser	Susceptible	4	100	100	100	100	104.4
PI 209794	Heins VII	*Bt0*	1	97	135.3	85.2	130.2	111.2
PI 554101	Selection 2092	*Bt1*	1	.	103.3	92.1	101.7	104.0
PI 554097	Selection 1102	*Bt2*	1	.	121.6	98.1	131.4	119.2
CItr 6703	Ridit	*Bt3*	3	76	10.5	42.6	67.2	51.2
PI 11610	CI 1558	*Bt4*	3	100	150.2	98.9	130.4	120.4
CItr 11458	Hohenheimer	*Bt5*	1	.	69.4	3.4	26.7	32.1
CItr 10061	Rio	*Bt6*	3	.	132.0	54.5	144.1	67.4
PI 554100	Selection 50077	*Bt7*	1	100	150.0	93.2	103.7	112.7
PI 554120	M72-1250	*Bt8*	2	.	0	3.0	1.9	7.5
PI 554099	R63-6968	*Bt9*	2	.	0	44.0	111.1	55.3
PI 554118	R63-6982	*Bt10*	2	.	17.9	19.0	69.3	37.0
PI 554119	M82-2123	*Bt11*	2	1	0	1.2	2.2	4.5
PI 119333	1696	*Bt12*	6	0	0	0	0	3.4
PI 181463	Thule III	*Bt13*	5	15	2.7	9.6	0.9	11.0
CItr 13711	Doubbi	*Bt14*	.	.	.	0.0	2.8	3.7
CItr 12064	Carleton	*Bt15*	.	.	.	9.6	15.5	14.8
PI 173437	7838	*Btp*	6	0	.	0.7	0	0.1
PI 173438	7845	Unknown	6	0	.	0	0.9	0.1
PI 178383	6256	*Bt8, 9, 10*	6	0	0	2.1	0	4.5
PI 476212	SM Selection 4	Unknown	6	1	0	0	0	4.0

[a]Data sets including the germplasm resources information network (GRIN), 2017, 2018 and 2019 Logan, UT field trials, and best linear unbiased estimates (BLUEs)

Table 2.2 Number and percent of resistant and susceptible bread wheat accessions in five data sets and number of accessions that were consistent for resistance or susceptibility across all data sets.

Data set[a]	Resistant[b]	Susceptible[c]	Percent resistant
GRIN	128	162	44.1%
2017	146	144	50.3%
2018	147	143	50.7%
2019	131	159	45.2%
BLUE	116	174	40.0%
Consistent across all trials	98	116	45.8%

[a]Data sets including the germplasm resources information network (GRIN), mean 2017, 2018 and 2019 Logan, UT field trials, and best linear unbiased estimate (BLUE) from across trails
[b]Resistance based on a DB normalized incidence $\leq$10%
[c]Susceptibility based on a DB normalized incidence >10%

Table 2.3 Geographic origin and number of bread wheat accessions highly resistant, resistant, and susceptible to DB across all data sets[a].

Accession Origin	DB resistance category		
Country	HR[b]	R[c]	S[d]
Azerbaijan	0	0	3 (2)
Germany	0	0	1
Iran	1 (1)	8 (8)	10 (10)
Montenegro	1 (1)	1 (1)	2 (2)
Russia	0	1	1 (1)
Serbia	1 (1)	6 (6)	9 (9)
Spain	0	0	1 (1)
Turkey	8 (8)	26 (25)	17 (13)
United States	17	56	72
Total	28	98	116

[a]Data sets including the germplasm resources information network (GRIN), mean 2017, 2018 and 2019 Logan, UT field trials; the number of landraces within each group shown in parenthesis
[b]Highly resistant accessions with a DB NI ≤1%
[c]Resistant accessions with a DB NI ≤10%
[d]Susceptible accessions with a DB NI >10%

Table 2.4 Geographic origin and number of bread wheat accessions in each subpopulation.

Accession Origin	Subpopulation					
Country	1	2	3	4	5	6
Azerbaijan	0	0	1	0	2	0
Germany	1	0	0	0	0	0
Iran	0	5	0	0	19	3
Montenegro	1	0	2	1	0	0
Russia	0	0	1	0	1	0
Serbia	2	1	15	0	0	0
Spain	1	0	0	0	0	0
Sweden	0	0	0	0	1	0
Turkey	4	25	1	0	11	42
United States	24	16	31	43	0	36
Total	33	47	51	44	34	81

Table 2.5 Marker-trait association groups significantly (FDR-adjusted $P \leq 0.05$) associated with dwarf bunt (DB) resistance.

DB Marker-trait group	Chr.[a]	QTL Range (Mbp)[b]	SNP Index[c]	Markers[d]	SNP[e]	FDR-adjusted negative $\log_{10}(P)$	R^2	RAF[f]	R SNP[g]	S SNP[g]	Significant data sets
DB-6D1	6D	1.77	IWB21614	2	[T/**C**]	1.8	0.09	0.85	40.8	87.0	BLUE
DB-6D2	6D	6.97 to 7.29	IWB59793	2	[A/**G**]	5.0	0.15	0.14	16.1	52.2	GRIN, 2018, BLUE

[a] Physical chromosome locations of each marker trait association group with 99 or 100% identity based on the physical annotation of wheat (IWGSC, 2018)

[b]Physical regions in Megabase pairs (Mbp) based on the physical annotation of wheat (IWGSC, 2018)

[c]Single nucleotide polymorphism (SNP) 90K index according to Wang et al., 2014 in each marker-trait association group with the lowest *P*-value; additional associated SNPs are reported in Supplementary file 3

[d]Additional SNPs in the marker-trait association group in high LD (≥ 0.8) with the SNP index

[e]SNP with resistance allele in bold and underlined

[f]Resistance allele frequency (RAF) for each indicator SNP

[g]Average DB normalized incidence values associated with the resistant (R) and susceptible (S) SNP allele; a low value indicates a high level of resistance

Table 2.6 Highly resistant landrace accessions with DB normalized incidence ≤1% across all data sets, bunt differentials and several known resistant and susceptible accessions and marker-trait group haplotypes detected in this study and three QTL reported in previous studies.

Accession	Bt gene	Origin	BLUE	*DB-6D1*[a]	*DB-6D2*	*QDB.ui-6DL*[b]	*QDB.ui-7AL*[c]	*QDB.ui-7DS*[d]
PI 345106		Serbia	3.9	+	-	-	-	-
PI 345428		Montenegro	4	-	-	-	-	-
PI 476212		United States	4	+	-	-	-	+
PI 560601		Turkey	3.8	+	-	-	-	-
PI 560602		Turkey	3.8	+	-	-	-	-
PI 560842		Turkey	3.8	+	-	-	-	+
PI 560843		Turkey	3.8	+	-	-	-	-
PI 560848		Turkey	3.8	+	-	-	-	+
PI 627677		Iran	4.1	+	-	-	-	-
CItr 8885	Susceptible	United States	116.2	-	-	-	-	-
PI 209794	Susceptible	Germany	111.2	+	-	-	-	-
PI 554101	*Bt1*	United States	104	-	-	-	-	-
PI 554097	*Bt2*	United States	119.2	-	-	-	-	-
CItr 6703	*Bt3*	United States	51.2	+	-	-	-	-
PI 11610	*Bt4*	United States	120.4	+	-	-	-	-
CItr 11458	*Bt5*	United States	32.1	+	-	-	-	-
CItr 10061	*Bt6*	United States	67.4	+	+	-	-	-
PI 554100	*Bt7*	United States	112.7	-	-	-	-	-
PI 554120	*Bt8*	United States	7.5	+	-	-	-	-
PI 554099	*Bt9*	United States	55.3	-	-	+	-	-
PI 554118	*Bt10*	United States	37	+	+	-	-	-
PI 554119	*Bt11*	United States	4.5	+	-	-	-	-
PI 119333	*Bt12*	Turkey	3.4	+	-	-	-	+
PI 181463	*Bt13*	Sweden	11	+	-	-	-	-
PI 173437	*Btp*	Turkey	0.1	+	-	-	-	-
PI 173438	*Bt* (unknown)	Turkey	0.1	+	-	-	-	+
PI 178383	*Bt8,9,10*	Turkey	4.5	+	+	+	-	-

[a]Presence (+) or absence (-) of the resistant allele haplotypes from each marker-trait association group detected in this study

[b]6DL haplotype SNP markers are reported in Wang, et al. (2019)

[c]7AL haplotype SNP markers are reported in Wang, et al. (2019)

[d]7DS haplotype SSR markers are reported in Chen, et al. (2016) and SNP markers for this region were provided by Rui Wang (personal communication)

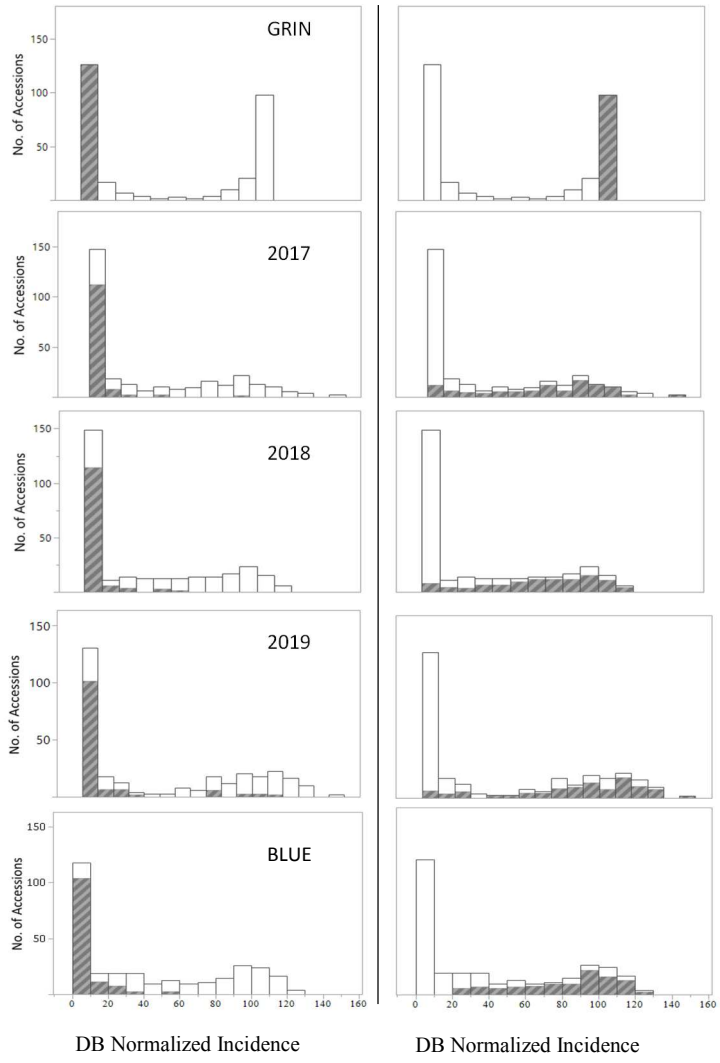

Fig. 2.1 DB normalized incidence distributions across 292 wheat accessions from four data sets including the germplasm resources information network (GRIN), mean 2017,2018 and 2019 Logan, UT field trials, and best linear unbiased estimates (BLUEs) from across trials; left pane: shaded accessions with DB normalized incidence ≤10% in GRIN, right pane: shaded accessions with DB normalized incidence ≥90% in GRIN.

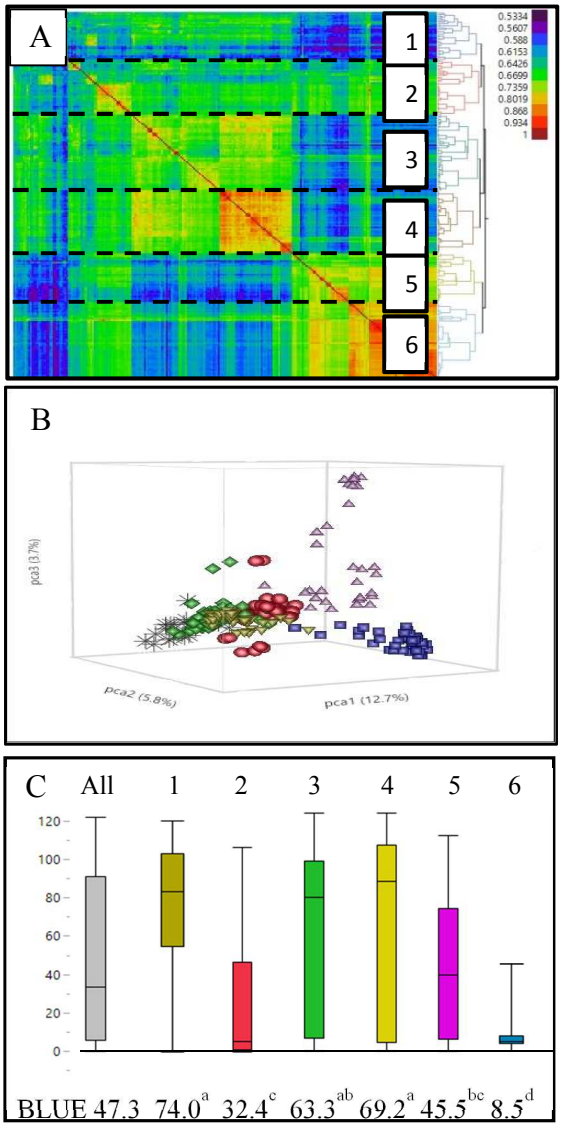

Fig. 2.2 **A** Genetic similarity heat map derived from an identity-by-state relationship matrix of 246 by 246 bread wheat accessions, regions of high (red) and low (purple) similarity between accessions; and a dendrogram showing six sub-populations (1 to 6) each separated by a dashed line. **B** Accessions plotted with three principal components showing sub-populations: 1 (brown stars), 2 (red circles), 3 (green diamonds), 4 (brown triangles), 5 (purple triangles), 6 (blue squares). **C** Best linear unbiased estimate (BLUE) DB normalized incidence quantile box plots, left to right: all 246 accessions (gray) and subpopulations: 1 (blue), 2 (red), 3 (green), 4 (yellow), 5 (purple), 6 (blue); mean BLUE values are listed for each sub-population below their respective box plots, means followed by a common letter are not significantly different by Tukey's HSD at $P \leq 0.05$.

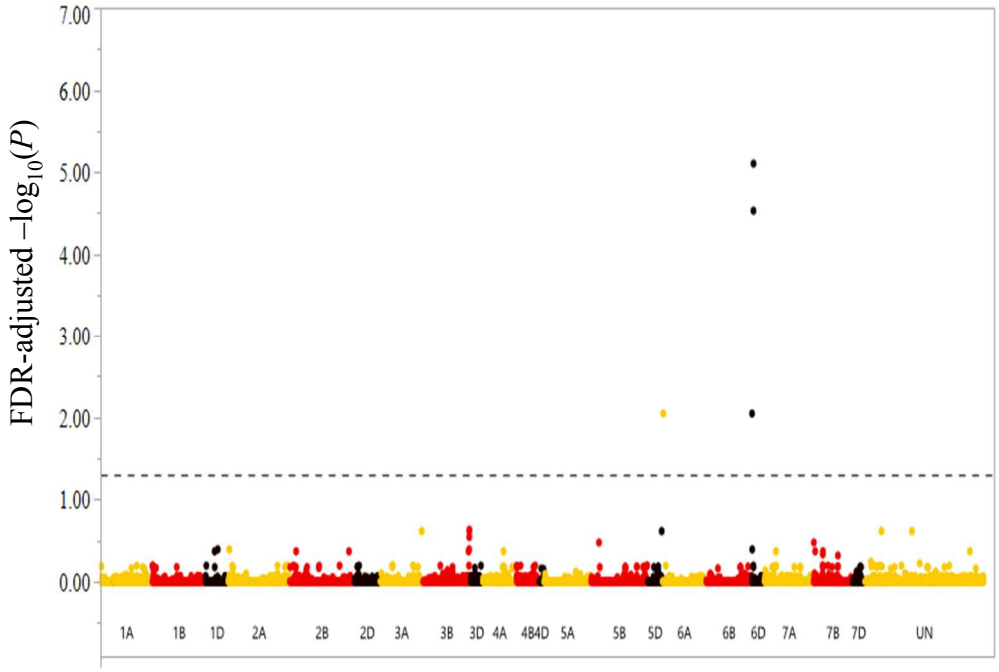

Fig. 2.3 Manhattan plot showing associations between 19,281 SNP markers and DB normalized incidence best linear unbiased estimates (BLUEs) across 246 bread wheat accessions; the horizontal dashed line indicates an FDR-adjusted significance threshold of $P = 0.05$; A-,B- and D-genome SNP markers are represented by yellow, red and black dots, respectively.

CHAPTER 3:

AGRONOMIC AND GENETIC ASSESSMENT OF TERMINAL DROUGHT TOLERANCE IN TWO-ROW SPRING BARLEY

ABSTRACT

Barley, *Hordeum vulgare* L., ranks fourth in global cereal grain production and is an important crop for animal feed, malting and human consumption. Identification of two-row barley germplasm with drought tolerance can increase genetic diversity of two-row barley germplasm and facilitate future barley breeding efforts. The present study evaluated 480 two-row spring barley accessions from the USDA National Small Grains Collection across two years of irrigated and terminal drought trials for grain yield, test weight, protein content, thousand kernel weight, and kernel size. Twenty accessions were identified that showed stable high yield, high test weights, and low protein content across trials. Ten of these accessions were cultivars or breeding lines originating primarily from North Dakota, and the other twenty accessions were landraces originating primarily from central Asia. An additional 10 accessions were identified with stable high yield, high test weights and high protein across trails. Genome-wide association mapping with 6,366 SNP markers revealed 15 drought-stable genetic loci significantly (FDR-adjusted $P < 0.05$) associated with at least one agronomic trait across and within treatments. One locus, on chromosome 2H between 27.2 and 29.8 Mbp, was significantly associated with heading date, plant height, and kernel size across treatments in this study, and the *PPD-H1* mutation in a previous study. Genetic loci on chromosomes 2H and 3H were significantly associated with increased test weight, and loci on chromosomes 3H and 5H were significantly associated with decreased grain

protein content across treatments. Accessions and SNP markers significantly associated with agronomic trait stability across terminal drought and irrigated environments can assist development of drought tolerant barley germplasm.

INTRODUCTION

Barley, *Hordeum vulgare* L., is an important cereal crop with over 147 million t harvested globally in 2017 (FAOSTAT, 2019). Barley's tolerance to drought, cold, and salinity have contributed to wide-scale cultivation for use as animal feed, malt, and for human food (Ullrich, 2010; Xie, et al., 2018). Recent genomic analyses of barley collections indicate that major contributors to barley population structure are geographic origin, annual growth habit, and row type (Muñoz-Amatriaín, et al., 2014; Wang, et al., 2017; Milner, et al., 2019). Growth habit is differentiated by sowing date in the autumn (winter) or spring (spring), whereas row type distinguishes between lateral floret sterility (two-row) and fertility (six-row). Six-row barley has the potential to produce more seeds and a higher yield. However, two-row types tend to produce larger, more uniform seeds which are often preferred by the malting industry.

Barley must meet strict standards for use as malt. To meet U.S. number 1 grade, two-row malting barley shipments must have a test weight $\geq$ 64.35 kg hl^{-1} and <5% thin kernels with widths <2.18 mm (USDA, 2013). Important malting parameters include grain protein content, kernel plumpness, thousand kernel weight, malt extract, wort β-glucan, α-amylase, diastatic power, soluble/total protein, and free amino nitrogen (Mohammadi, et al., 2015;

Belcher, et al., 2018; AMBA, 2019). These agronomic and malting traits can be significantly influenced by environmental conditions. For example, applying drought and heat stress to two-row spring barley cultivars during heading, i.e. Zadoks (1974) stage 55, resulted in large yield losses up to 95%, and reduced grain and malting quality (Law, 2019; Mahalingam and Bregitzer, 2019). In another study (Morgan and Riggs, 1981) spring barley cultivars subjected to terminal drought at heading had lower yields, fewer plump kernels with widths >2.2 mm, lower thousand kernel weights, higher grain protein content, and reduced malting quality compared to irrigated controls.

Global mean surface temperatures have been steadily increasing (Hansen, et al., 2006; Hansen, et al., 2010) and exacerbate extreme drought and heat events (Xie, et al., 2018). Fresh water availability is inextricably linked with climate extremes as heat and drought require additional water allocation to agricultural production which accounts for roughly 92% of fresh water use (Hoekstra and Mekonnen, 2012). Based on current production constraints, as the number of extreme weather events increase, agricultural outputs including malt barley production are predicted to substantially shrink while prices for these products increase (Xie, et al., 2018). Selecting drought and heat tolerant crops and varieties and identifying genomic regions associated with extreme-event tolerance has been proposed as a way to mitigate these effects in agricultural systems (Varshney, et al., 2018).

Germplasm collections provide a wealth of unexplored diversity and harbor alleles that could be useful for drought and heat tolerance. Collection accessions paired with expanded genetic information and physical annotations (Mascher, et al., 2017) can provide a platform for novel trait and gene discovery through genome-wide association studies

(GWAS) (Milner, et al., 2019). The National Small Grains Collection (NSGC) in Aberdeen, Idaho contains the US collection of small grains including over 33,000 barley accessions. A previous GWAS study conducted by Muñoz-Amatriaín, et al. (2014) identified 1,860 barley accessions of diverse origin, growth habit and row type within the NSGC that were characterized as the informative Core or iCore. Their analysis identified SNP markers associated with row type, hull cover, and heading date that were coincident with previously reported mutations. These results indicated that the iCore could be effectively utilized for GWAS.

In the present study 480 spring habit, two-row barley accessions were selected from the NSGC iCore and grown as plots under normal irrigation and terminal drought treatments over two growing seasons in Aberdeen, ID. The objectives of the study were to: 1) identify terminal drought tolerant accessions and 2) locate genomic regions associated with agronomic trait stability in these drought-tolerant accessions through GWAS.

MATERIALS AND METHODS

Plant materials

Four-hundred and eighty two-row, spring growth habit barley accessions were selected from the iCore (Muñoz-Amatriaín, et al., 2014). These 480 accessions were comprised of 84 breeding lines, 211 cultivars, five genetic stocks, 140 landraces, and 40 of uncertain breeding status which originated in 72 countries (Supplemental File A3.1) Additional information on the accessions can be found online at the Germplasm Resources Information

Network (GRIN website = www.ars-grin.gov). Five barley checks were used including four modern malting varieties: 'Harrington' (Harvey and Rossnagel, 1984), 'Lacey' (Rasmusson, et al., 2001), 'Pinnacle' released by the North Dakota State University Research Foundation in 2007, and 'Tradition' released by Busch Agricultural Resources, Inc. in 2003, and one feed variety 'Baronesse' released by Western Plant Breeders, Inc. in 1991.

Field design

Trials were sown on 17 April 2014 and 19 April 2017 at the University of Idaho Aberdeen Research and Extension Center in Aberdeen, Idaho, U.S. (42°57'36'' N, 112°49'12'' W, elevation 1342 m). The two treatments, irrigated (IR) and terminal drought (DR), were planted adjacent to one another and separated by four plots of spring wheat as borders. Accessions and checks were sown in 1.5 m by 3.05 m plots and planted in seven rows at a rate of approximately 364,500 kernels ha^{-1}. Each treatment was arranged in an augmented complete block design (Federer and Raghavarao, 1975), in which the 480 unique accessions were planted only once across twelve sub-blocks, with 40 accessions in each sub-block. Five check varieties were planted once per sub-block, but replicated in all twelve sub-blocks to account for spatial variation within each treatment. Each of the sub-blocks consisted of 45 plots arranged in a rectangle 9 plots by 5 plots and each treatment was two sub-blocks by six sub-blocks.

Trial soils were Declo-loam (coarse-loamy, mixed, superactive, mesic Xeric Haplocalcids) with 0 to 2% slopes, pH of 8.1, and high water holding capacity (USDA Web Soil Survey, https://websoilsurvey.sc.egov.usda.gov/App/WebSoilSurvey.aspx. Accessed 30

May 2019). Soil samples were taken at 10 cm and 20 cm depth in three randomly selected blocks within each treatment, and nitrogen was applied in a pelleted 46-0-0 formulation of urea to adjust the nitrogen target rate within each treatment. Both the IR and DR treatments had target yields of 6.5 t ha^{-1}, which required 202 kg of available N ha^{-1}.

Spring sown barley grown in Aberdeen, Idaho typically reaches anthesis, Zadoks growth stage 60 (Zadoks, et al., 1974), in late June and is harvested in late July. Aberdeen's arid, continental climate has warm and dry summers and provides a conducive environment for terminal drought studies. The Aberdeen Experiment Center mean July temperatures were 22°C in both 2014 and 2017, and total July precipitation was 6 mm in 2014 and 4 mm in 2017 (US Bureau of Reclamation, Cooperative Agricultural Weather Network 'AGRIMET': http://www.usbr.gov/pn/agrimet/webarcread.html accessed 29 May 2019).

Irrigation for all treatments was scheduled using AGRIMET weather data and targeted a soil moisture profile half way between field capacity and wilting point in the active crop root zone, or 50% management allowable depletion (Qureshi and Neibling, 2009). In 2014 water was applied using an irrigation drip system spaced every two rows in each plot, and in 2017 water was applied using an aluminum pipe system with overhead sprinklers spaced every 12.8 m across each treatment. IR treatments were irrigated weekly until 50% of plots had reached physiological maturity, Zadoks 85. DR treatments were also irrigated weekly; however, irrigation was terminated when 50% of the plots reached anthesis, Zadoks 60. Plots were harvested after ripening, Zadoks 93, on 30 July 2014, and on 16 August 2017.

In 2014 the DR treatment received three fewer irrigations totaling 128 mm less water than the IR treatment, while in 2017 the DR treatment received two fewer irrigations accounting for 97 mm less water than the IR treatment (Supplemental File A3.2). Using AGRIMET estimations of water holding capacity based on crop and soil types, the DR treatments reached continuous 50% maximum allowable soil water depletion approximately 11 days before the IR treatment in both years (Supplemental File A3.2).

Agronomic trait evaluation

In all trials, individual plots were harvested with a Wintersteiger Classic small plot combine equipped with a Harvest Master system (Wintersteiger Inc., Ried im Innkreis, Austria). Yield (YLD) was calculated from raw grain weight and converted to kg ha^{-1}. Test weight (TWT) was measured after the grains had dried to <10% moisture using a dry pint kettle and balance in accordance with USDA Grain Grading Standards (USDA, 2013) and converted to kg hl^{-1}. Days to heading (HD) was measured in days from the planting date until 50% of the heads in each plot had emerged, Zadoks 55. Plant height (HT) was measured in cm at physiological maturity, Zadoks 85, from the soil surface to the top of the spike. At physiological maturity a qualitative rating scale (Erickson, et al., 1982) was used to determine the severity of relative lodging (LODGE) in each plot with 0 = no lodging/upright and 9 = fully lodged/prostrate on the ground and lodging incidence (LODGED%) on a scale between 0-100%. Protein percentage (PRO) was calculated using a 20 g sample processed with a near-infrared Perten Grain Analyzer 9100 (Perten Instruments, Hägersten, Sweden) as described by the manufacturer. Individual grain kernel dimensions

including width in mm (GW), length in mm (GL), diameter in mm (GD), volume in mm^3 (GV), width/length ratio (GWL), and thousand kernel weight in g (TKW) were measured using a 5 g sample processed through a Foss Tecator GrainCheck™ 2313 (Foss Analytical, Hillerød, Denmark) grain analyzer as described by the manufacturer. All traits were measured in both environments and years except the lodging traits were only measured in 2017, and the PRO and kernel sizes were not measured in the 2014 IR treatment.

Molecular markers

Single nucleotide polymorphism (SNP) marker acquisition and curation for the entire iCore panel was previously described (Muñoz-Amatriaín, et al., 2014) and downloaded from T3, the Triticeae Toolbox website available at http://triticeaetoolbox.org/barley/ (Accessed 29 May 2019). SNPs were excluded if minor allele frequency (MAF) was <4%, or missing data >10%; heterozygous calls were also removed. A final group of 480 two-row, spring habit barley accessions was selected and 6,366 SNP markers were aligned with the physical barley annotation (Mascher, et al., 2017) and used for subsequent marker-trait associations.

Statistical analysis

All statistical analyses were conducted using JMP® Genomics v. 9.0. A mixed model with factors Year, Block, Treatment, Check, Accession, Treatment *Accession, were used to extract best linear unbiased predictors (BLUPs) using the restricted maximum likelihood (REML) method for all agronomic traits across and among treatments (Wolfinger, et al., 1997). Year, Block, and Accession were random effects, while Treatment

and Check were fixed effects in the model. Correlation coefficient estimates between agronomic trait BLUPs across and among treatments were calculated using a Pearson's correlation procedure (Pearson, 1895). BLUP values for YLD, TWT, and PRO were ranked from 1 to 480 across treatments and within each treatment. Two accession summaries were generated based on these three trait rankings. One summary was for those interested in malting quality: the highest values for YLD and TWT were given the highest rank, while the lowest BLUP values for PRO were given the highest rank. The other summary was for those interested in food quality: the highest values for YLD, TWT and PRO were given the highest ranks. Trait rankings were combined into a single score in order to select those accessions with the best ranks across all treatments. A drought susceptibility index (DSI) was calculated, as described by Li et al. (2012), derived from Fischer and Maurer (1978), for each accession across all measured traits (Supplemental File A3.1) to assess drought stability.

Genome-wide linkage disequilibrium (LD) was calculated as r^2 values between each marker within chromosome groups. An identity-by-state (IBS) familial relationship matrix (K matrix) and heat map (Fig. 3.1A) were generated to explore potential sub-populations (Gower, 1971). STRUCTURE (Pritchard, et al., 2000) and STRUCTURE HARVESTER (Earl and vonHoldt, 2012) were used to determine optimal subpopulation number and identify the subpopulation with the highest Δk value. In STRUCTURE, burn-in iterations and Markov chain Monte Carlo replications were set to 10,000, the admixture correlated model was selected, and five replicate iterations were performed. A principal component analysis with 10 principal components (PCs) was generated (Q matrix) to explore population

stratification (Fig. 3.1B), and the resulting scree plot was used to estimate the optimal number of PCs in the model (Price, et al., 2006).

Various association models were formally tested using Bayesian information content (BIC) assessments (Burnham and Anderson, 2004). Tested models included a general linear model (GLM) with no corrections for K or Q (the naïve model), a GLM that corrected for population stratification with 2, 3 or 5 PCs, and mixed linear models (MLMs) controlling for both familial relationships, as a random effect, and population stratification with 2, 3 or 5 PCs as a fixed effect (Yu, et al., 2006). All models correcting for familial relationships performed better and had lower BIC values than the naïve model. A MLM with a kinship covariate matrix and 3 PCs had the lowest BIC value and was therefore chosen for further marker-trait association analysis. An FDR multiple testing correction (Benjamini and Hochberg, 1995) was used to adjust all reported SNPs and control for potential type I errors (Supplementary File 3). SNPs significantly (FDR-adjusted $P \leq 0.05$) associated with any agronomic trait BLUP across or within treatments are listed in Supplemental File A3.3. To assess potentially linked SNPs, the most significant marker in each putative QTL were included one at a time in the MLM as covariates. Markers in high LD with the covariate marker were no longer significantly associated with agronomic traits and were grouped with the covariate SNP QTL.

RESULTS

Treatment effects among accessions

The DR treatment received fewer irrigations than the IR treatment, and stayed continuously below the estimated 50% water holding capacity for 11 days longer than the IR treatment in both 2014 and 2017 (Supplementary File A3.2). A mixed model was used to test for significant parameter effects and random effect variance components. The model indicated a significant treatment effect $P \leq 0.001$ associated with most of the measured agronomic traits (Table 3.1). Specifically, a reduction in YLD, TWT, GW and an increase in PRO HT and GL was observed in the DR treatments when compared with the IR treatment (Table 3.1). Heading date and lodging severity were the two traits that did not show a significant treatment effect (Table 3.1). Variance components generated from the mixed model suggested that genetic variability within the accessions accounted for more than 95% of the observed phenotypic variation for every measured trait (Table 3.2). Year and the Accession*Treatment interaction did not have a significant effect on any trait, and Block had a significant effect on some of the measured traits including YLD, PRO, HD, HT, LODGE, GV and TKW (Table 3.2).

Correlations among traits

BLUPs and significance values were extracted for each of the 480 accessions across treatments and within each treatment (Supplemental File A3.1). Trait correlations were assessed using the extracted accession BLUPs within each treatment and across treatments. GW and GWL were positively correlated with YLD (coefficients of 0.43 and 0.48,

respectively) and TWT (coefficients of 0.16 and 0.45, respectively) across treatments (Table 3.3) and within each treatment (Supplemental File A3.4). LODGE% and GL were negatively correlated with YLD (coefficients of -0.60 and -0.28, respectively) and TWT (coefficients of -0.22 and -0.40) across treatments (Table 3.3) and within each treatment (Supplemental File A3.4). GD and GV were highly correlated, 0.98 and 0.74, respectively, with GL and were excluded from Table 3, but are shown in Supplemental File A3.4.

Accession identification based on phenotyping data

Of the five checks, Pinnacle tended to outperform the other four barley checks across years and environments. It had the highest test weights and seed dimensions, but also lodged less severely and headed earlier than the other checks (Table 3.4). Harrington tended to underperform the other checks with low test weights and seed dimensions, and had more severe lodging and later heading dates (Table 3.4). These trends were exacerbated in the DR treatment, where Pinnacle had significantly higher TWT, GW, and TKW than the other check accessions (Table 3.4).

Accessions with the highest YLD and TWT and lowest PRO were identified by ranking BLUP values for these traits across and within each treatment as described previously in the Methods. Accessions with the highest rankings for these traits were breeding lines or cultivars from the United States or Europe (Table 3.5); whereas accessions originating from Central Asia and South American had the highest rankings within the landraces (Table 3.5). Table 3.6 highlights ten accessions with high YLD, TWT and high PRO content and could be used for food barley development. The highest ranking

accessions in this category were from Europe: PI 321770, PI 428113, PI 371102 and Ethiopia PI 296455 (Table 3.6). Breeding lines and cultivars composed the majority of the accessions with highest YLD rankings, but landraces composed a substantial number of accessions with the highest TWT and PRO and lowest DSI values (Supplemental File A3.1).

PI 643339 was the highest ranked accession based on YLD, TWT and low PRO within and across treatments (Table 3.5). This accession is a breeding line from North Dakota State University with resistance to barley leaf rust (*Puccina hordei* Otth.), and contains the leaf rust resistance gene *Rph15*. Malting barley breeders may consider using PI 643339 as a source of drought tolerance and leaf rust resistance in their programs. The second highest ranking accession across treatments was PI 643267, and may also be a good line for breeding programs, but it is not listed as having disease resistance in GRIN. The highest ranking landrace was PI 422233 from Yemen (Table 3.5). GRIN data indicates that this accession may have extreme resistance to barley stripe mosaic virus, and therefore would be a good source of resistance to this disease in addition to drought tolerance in barley breeding programs. CIho 14395 from Armenia was the second highest ranking landrace in this study, and also was resistant to barley stripe mosaic virus in GRIN.

Population structure

Accession identity across 6,366 SNPs and between the 480 accessions ranged from 48% to 99%, with a mean identity of 63% and median of 62%. Genome-wide LD decayed from a median r^2 of 0.28 at 0 Mbp to 0.16 at 1 Mbp, indicating a nearly 50% decline within the first Mbp. Ward hierarchical clustering based on an IBS matrix and PCA (Fig. 3.1) were

used to group accessions based on SNP relatedness. The variation explained by the first three principal components was 10.7%, 4.4% and 4.0% by PC1, PC2 and PC3, respectively. STRUCTURE and STRUCTURE HARVESTER were used to determine the optimal number of subpopulations. Subpopulation identity based on SNP data generally matched geographic origin descriptor data found on GRIN. More than 50% of the Asian landraces were in subpopulation 1, while 80% of the African landraces were in subpopulation 2 (Table 3.7). A majority (96%) of the accessions in subpopulation 4 were U.S. breeding lines from North Dakota, while the lone landrace in subpopulation 4 was PI 422233, originating from Yemen, which was also the highest yielding landraces in this study (Table 3.5). Subpopulation 5 was composed primarily (83%) of accessions from Europe, while subpopulation 3 had representation from all UN regions (Table 3.7). Overall, the highest YLD and TWT and lowest PRO across treatments were in subpopulation 4, while the inverse was observed in subpopulation 2 (Table 3.8).

Genome-wide trait-associations

There were 590 SNPs significantly (FDR-adjusted $P \leq 0.05$) associated with at least one trait and one treatment (Supplemental File A3.3). Forty-four of these SNPs were significant in both IR and DR treatments and across treatments and composed 15 marker-trait association loci (Table 3.9). These loci had significant SNPs with FDR-adjusted negative log *P*-values ranging from 1.3 to 14.4 and R^2 values ranging from 0.03 to 0.15. Four of the loci were associated with TWT, three were associated PRO and four were associated with GW. Locus groups 3 and 5 were associated with the lowest *P*-values and

minor allele haplotypes from these groups were associated with high YLD and TWT and low PRO (Tables 3.9 and 3.10). Locus group 9 was associated with high YLD and very low PRO, but also a reduced TWT (Table 3.9). Several of the locus groups were associated with specific subpopulations, for example, 93% and 81% of the accessions associated with the minor alleles in locus groups 4 and 15, respectively, were from subpopulation 2 (Table 3.11).

DISCUSSION

Germplasm collections harbor underexploited genetic resources that can help mitigate future global agriculture production constraints when used in breeding for heat and drought tolerance. NSGC spring habit, two-row barley accessions were selected from the previously developed iCore and planted as plots under regular irrigation and terminal drought conditions. Evapotranspiration data generated from AGRIMET provided the basis for determining irrigation timing for the various treatments. Significant treatment effects were observed using this method (Table 3.1), however, the use of soil moisture probes spread throughout the field would have provided additional accuracy for determining soil water content and could provide supplementary guidance for irrigation timing.

BLUP values were extracted from each accession across and within treatments. An augmented complete block design was used with check cultivars randomly interspersed throughout each block to account for spatial variation in the field (Federer and Raghavarao, 1975). This design maximized the number of accessions that could fit in each treatment, and allowed for the quick determination of accession ranking for each trait under each specified

environment. However, the BLUPs are not easy to relate to other published accession data. For instance, an accession with a negative yield or test weight BLUP would indicate a poor performing individual, but this negative value would mean little outside this set of accessions and treatments.

Trait correlations were similar to those previously reported in barley (Sharma, et al., 2018). For instance, GL was negatively correlated with YLD and GWL in the present study (-0.28 and -0.79, respectively across trials) and -0.11 and -0.82, respectively, in Sharma et al. (2018). Similarly, in both studies GW and GWL were positively associated with YLD at 0.43 and 0.48, respectively. Another GWAS study (Matthies, et al., 2014) also reported a strong positive correlation between grain sieve fraction above 2.8 mm and high TWT in European malting barley cultivars. Selecting barley lines with uniformly high width would be a beneficial strategy for selecting high yields and test weights. Removing barley lines with high lodging severity from a breeding program would also have the effect of increasing YLD and TWT in the remaining lines. Conversely, some accessions that might otherwise have high YLD or TWT would not be detected if lodging severity was high.

High test weights $\geq$ 64.35 kg hl^{-1} and low protein content $\leq$ 12% are useful agronomic benchmarks for two-row malting barley and are associated with price premiums (USDA, 2013; AMBA, 2019). In this study accessions with high YLD, high TWT and low PRO were identified in Table 3.5 by ranking them for each of these attributes and then combining the ranks for a single rank value across and within treatments. Twenty accessions were identified that had the highest yield and test weights and lowest protein content (Table 3.5). Most of the highest ranking accessions across these traits, including the highest ranking

accession, PI 643338, were breeding lines from North Dakota, which are already likely in the genetic background of many U.S. cultivars. Some of the other accessions in this group might not be present in advanced breeding material and therefore would be useful germplasm for breeding programs. For example, PI 313113 from Germany was introduced in the US National Plant Germplasm System (NPGS) in 1966 and ranked highly for YLD, TWT and PRO. Similarly, PI 467811 from Austria, was ranked highly across the treatments, shows net blotch resistance based on GRIN data, and may have potential for improving two-row barley germplasm.

The highest ranking landrace, PI 422233 from Yemen, was grouped in the same subpopulation as the North Dakota breeding lines and its genetic background may already be represented in U.S. cultivars (Table 3.5). CIho 14395 from Armenia was added to the U.S. NPGS in 1927 and may have unique drought-tolerance attributes that are not currently available in adapted germplasm. PI 573687 from Georgia was in subpopulation 1 and was the fourth highest ranking landrace and had low stripe rust severity based on GRIN data.

Other trait weighting schemes would allow selection for additional accessions. The American Malting Barley Association (2019) recommends production of >90% plump kernels with widths >2.38 mm and <3% thin kernels with <2.18 mm widths. CIho 16658, a genetic stock with large lateral florets, but a late heading date, had the second highest kernel widths across treatments and in the IR and DR treatments. PI 151787, an Ethiopian landrace received by the U.S. NPGS in 1945, had exceptionally high GW and GWL ratios across treatments and could also be selected for kernel plumpness. PI 195319, a landrace from

Guatemala, although not a high yielding accession, had the highest test weight across all treatments and within each treatment.

Barley used for food is often selected for high protein content (Rogers, et al., 2017). Ten accessions with high YLD, TWT and high PRO content are listed in Table 3.6. PI 321770 from Slovenia, PI 428413 from France and PI 296455 from Ethiopia could be important accessions for development of high protein and food quality barley (Table 3.6). Numerous other Ethiopian landraces had high levels of protein. Specifically, CIho 14978 from Tigray province, had the highest protein content across treatments and could perhaps be used to increase protein content in barley grown for livestock or human consumption. However, protein content was strongly and negatively correlated with yield; perhaps incremental progress could be made in breaking this negative linkage using previously unexploited resources such as PI 296455 or CIho 14978. Additional food and malt quality characteristics including grain β-glucan, malt extract, wort, α-amylase, diastatic power, soluble/total protein, and free amino nitrogen could be measured on selected accessions within this panel. Specifically, landraces with exceptional agronomic trait performance would make good candidates for additional malt or food quality assessment.

One major genetic locus for heading date, locus 3, was identified on chromosome 2H between 27.2 and 29.8 Mbp (Table 3.9). This locus was associated with increased YLD and TWT (Table 3.10) across treatments, and 65% of the accessions with the minor allele for this locus were in subpopulation 3 (Table 3.11). This 2H locus is located at the same physical position as SNP markers associated with heading date previously identified in the iCore (Muñoz-Amatriaín, et al., 2014) on chromosome 2H at 29.1 Mbp and linked with

PPD-H1, a photoperiod response regulator involved in long day flowering time (Turner, et al., 2005). Spring accessions in the Muñoz-Amatriaín et al (2014) study were planted in the fall in Corvallis, Oregon, while in the present study the accessions were planted in the spring in Aberdeen, Idaho. Variations in planting date and treatment conditions did not seem to affect the robust positive effects of this locus which may warrant inclusion in barley breeding programs.

Another heading date locus reported from the iCore was on chromosome 2H at 519.1 Mbp (Muñoz-Amatriaín, et al., 2014) and was tightly linked with *EPS2*, an earliness per se flowering determinant mutation (Comadran, et al., 2012). We detected a significant locus on chromosome 2H at 559.1 Mbp that was associated with many agronomic traits, but not HD. It is likely that these two loci are distinct, and the HD locus reported by Muñoz-Amatriaín et al. (2014) was not detected in the present study because of the different growing environments and absence of the six-row accessions.

Two of the agronomic trait genetic loci aligned with previously reported agronomic hotspots identified in a nested association mapping study (Sharma, et al., 2018) measuring YLD, TKW, GW, GL, grain number per ear, grain area, and grain roundness in 25 wild barley (*Hordeum vulgare* subsp. *spontaneum* (K. Koch) Thell.) accessions backcrossed into the adapted spring barley cultivar 'Barke'. Locus 3 on chromosome 2H at 29.1 Mbp, and locus 15 between 576.3 and 591.6 Mbp were associated with grain size in the present study and also the Barke nested association mapping study. Additional barley agronomic trait loci that were within 5 Mbp of the loci in this study were detected in a GWAS of 174 European malting cultivars including those on 1H associated with TKW, and 5H associated with PRO

(Matthies, et al., 2014). One previously reported SNP associated with grain protein content in elite North American six-row malting barley, SCRI_RS_147762, was located on 5H at 559.2 Mbp (Belcher, et al., 2018). This is approximately 86 Mbp proximal to the 5H locus associated with PRO in this study, and is likely a distinct locus that controls grain protein content.

Previously reported loci associated with agronomic traits in barley that were also detected in the current study indicate the potential durability of these genetic regions across environments, and would be important regions to target for barley cultivar development. Specifically, the locus on 2H at 29.1 Mbp was significantly associated with HD, HT and other grain size traits and was correlated with higher YLD and TWT across treatments in this study and other studies (Muñoz-Amatriaín, et al., 2014; Sharma, et al., 2018). Unreported genetic loci 6 and 9 that are associated with increased TWT and low PRO across treatments in the present study may be novel, and bi-parental populations could be developed to validate these loci and explore the effects in additional detail.

CONCLUSIONS

Global temperature increases correspond with an increase in the number and frequency of extreme weather events that negatively affect barley yield and quality. Here we highlight 30 two-row spring barley accessions from the NSGC iCore that can tolerate terminal drought conditions and show stable agronomic traits across drought treatments. Fifteen genetic loci associated with important agronomic traits across and within drought treatments were

identified. Loci on chromosomes 2H, 3H and 7H were associated with higher TWT and loci on chromosomes 3H and 5H were associated with lower PRO. Accession trait and genetic information can be used to facilitate the improvement of barley cultivars that are resilient to drought stress and fulfill an important goal of future barley breeding efforts.

REFERENCES

AMBA. 2019. Malting barley breeding guidelines ideal commercial malt criteria. American Malting Barley Association, Inc. Publication.

Belcher, A.R., A. Cuesta-Marcos, K.P. Smith and P.M. Hayes. 2018. TCAP FAC-WIN6 elite barley GWAS panel QTL. II. malting quality QTL in elite North American facultative and winter six-rowed barley identified via GWAS. Crop Sci 58: 120-132.

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

Burnham, K.P. and D.R. Anderson. 2004. Multimodel inference: understanding AIC and BIC in model selection. Sociol Methods Res 33: 261-304.

Comadran, J., B. Kilian, J. Russell, L. Ramsay, N. Stein, M. Ganal, et al. 2012. Natural variation in a homolog of *Antirrhinum CENTRORADIALIS* contributed to spring growth habit and environmental adaptation in cultivated barley. Nature Genet 44: 1388.

Earl, D.A. and B.M. vonHoldt. 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour 4: 359-361.

Erickson, D., D. Meyer and A. Foster. 1982. The effect of genotypes on the feed value of barley straws. J Animal Sci 55: 1015-1026.

FAOSTAT. 2019. Production and yield quantities of cereal grains. Food and agriculture organization of the United Nations.

Federer, W.T. and D. Raghavarao. 1975. On augmented designs. Biometrics: 29-35.

Fischer, R. and R. Maurer. 1978. Drought resistance in spring wheat cultivars. I Grain yield responses. Aust J Agric Res 29: 897-912.

Gower, J.C. 1971. A general coefficient of similarity and some of its properties. Biometrics 27: 857-871.

Hansen, J., R. Ruedy, M. Sato and K. Lo. 2010. Global surface temperature change. Rev of Geophys 48.

Hansen, J., M. Sato, R. Ruedy, K. Lo, D.W. Lea and M. Medina-Elizade. 2006. Global temperature change. Proc Nat Acad Sci USA 103: 14288-14293.

Harvey, B. and B. Rossnagel. 1984. Harrington barley. Canadian J Plant Sci 64: 193-194.

Hoekstra, A.Y. and M.M. Mekonnen. 2012. The water footprint of humanity. Proc Nat Acad Sci USA 109: 3232-3237. doi: .

Law, S.R. 2019. Sobering challenges faced by the brewing industry in a warming world. Physiol. Plant. 165: 131-133.

Li, P., J. Chen and P. Wu. 2012. Evaluation of grain yield and three physiological traits in 30 spring wheat genotypes across three irrigation regimes. Crop Sci 52: 110-121.

Mahalingam, R. and P. Bregitzer. 2019. Impact on physiology and malting quality of barley exposed to heat, drought and their combination during different growth stages under controlled environment. Physiol. Plant. 165: 277-289.

Mascher, M., H. Gundlach, A. Himmelbach, S. Beier, S.O. Twardziok, T. Wicker, et al. 2017. A chromosome conformation capture ordered sequence of the barley genome. Nature 544: 427-433.

Matthies, I.E., M. Malosetti, M.S. Röder and F. Eeuwijk. 2014. Genome-wide association mapping for kernel and malting quality traits using historical European barley records. PloS ONE 9: e110046.

Milner, S.G., M. Jost, S. Taketa, E.R. Mazón, A. Himmelbach, M. Oppermann, et al. 2019. Genebank genomics highlights the diversity of a global barley collection. Nature Genet 51: 319-326.

Mohammadi, M., T.K. Blake, A.D. Budde, S. Chao, P.M. Hayes, R.D. Horsley, et al. 2015. A genome-wide association study of malting quality across eight U.S. barley breeding programs. Theor and Appl Genet 128: 705-721.

Morgan, A.G. and T.J. Riggs. 1981. Effects of drought on yield and on grain and malt characters in spring barley. J Sci Food Agric 32: 339-346.

Muñoz-Amatriaín, M., A. Cuestra-Marcos, J.B. Endelman, J. Comadran, J.M. Bonman, H.E. Bockelman, et al. 2014. The USDA barley core collection: genetic diversity,

population structure, and potential for genome-wide association studies. PLoS ONE 9 (4): e94688.

Pearson, K. 1895. VII. Note on regression and inheritance in the case of two parents. Proc Royal Soc Lond 58: 240-242.

Price, A.L., R.M. Patterson, R.M. Plenge, M.E. Weinblatt, N.A. Shadick and D. Reich. 2006. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 38: 904-909.

Pritchard, J.K., M. Stephens and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics 155: 945-959.

Qureshi, Z.A. and H. Neibling. 2009. Response of two-row malting spring barley to water cutoff under sprinkler irrigation. Agric Water Manag 96: 141-148.

Rasmusson, D., K. Smith, R. Dill-Macky, E. Schiefelbein and J. Wiersma. 2001. Registration of 'Lacey' barley. (Registration Of Cultivars). Crop Sci 41: 1991-1992.

Rogers, C.W., G. Hu and R. Mikkelsen. 2017. Grain yield, quality, and nutrient concentrations of feed, food, and malt barley. Commun Soil Sci Plan 48: 2678-2686.

Sharma, R., F. Draicchio, H. Bull, P. Herzig, A. Maurer, K. Pillen, et al. 2018. Genome-wide association of yield traits in a nested association mapping population of barley reveals new gene diversity for future breeding. J Exp Bot 69: 3811-3822.

Turner, A., J. Beales, S. Faure, R.P. Dunford and D.A. Laurie. 2005. The pseudo-response regulator Ppd-H1 provides adaptation to photoperiod in barley. Science 310: 1031-1034.

Ullrich, S.E. 2010. Barley: production, improvement, and uses. John Wiley & Sons.

USDA. 2013. Grain Inspection Handbook, Book II, Grain Grading Procedures. US Department of Agriculture, Marketing and Regulatory Programs, Grain Inspection, Packers and Stockyards Administration, Federal Grain Inspection Service, Washington, DC.

Varshney, R.K., V.K. Singh, A. Kumar, W. Powell and M.E. Sorrells. 2018. Can genomics deliver climate-change ready crops? Curr Opin Plant Biol 45: 205-211.

Wang, R., Y. Leng, S. Ali, M. Wang and S. Zhong. 2017. Genome-wide association mapping of spot blotch resistance to three different pathotypes of Cochliobolus sativus in the USDA barley core collection. Mol Breed 37: 44.

Wolfinger, R.D., W.T. Federer and O. Cordero-Brana. 1997. Recovering information in augmented designs, using SAS PROC GLM and PROC MIXED. Agron J 89: 856-859.

Xie, W., W. Xiong, J. Pan, T. Ali, Q. Cui, D. Guan, et al. 2018. Decreases in global beer supply due to extreme drought and heat. Nat Plants 4: 964-973.

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

Zadoks, J.C., T.T. Chang and C.F. Konzak. 1974. A decimal code for the growth stages of cereals. Weed Res 14: 415-421.

TABLES AND FIGURES

Table 3.1 Trait estimates across treatments and within treatments, and treatment significance across accessions and years.

Trait[a]	Across treatments[b]	Irrigated	Terminal drought
YLD	5.48***	5.84	5.12
TWT	65.1***	65.6	64.5
PRO	13.4***	12.1	13.5
HD	63.5 ns	63.1	63.8
HT	91.9***	90.4	93.4
LODGE	3.0 ns	2.6	3.4
LODGE%	23.3***	19.9	26.8
GW	3.5***	3.6	3.5
GL	9.0***	9.0	9.1
GD	5.80**	5.81	5.83
GV	47.5**	49.4	47.3
GWL	0.40***	0.41	0.39
TKW	41.5***	44.7	40.6

[a]YLD is yield in metric tons ha^{-1}; TWT is test weight in kg hl^{-1}; PRO is total grain protein percentage; HD is heading date in days; HT is height in cm; LODGE is lodging severity where 0 = no lodging and 9 = fully lodged; LODGE% is lodging incidence between 0-100%; GW is grain width in mm; GL is grain length in mm; GD is grain diameter in mm; GV is grain volume in mm^3; GWL is the grain width by grain length ratio; TKW is thousand kernel weight in g

[b]Mean trait estimate across all treatments; treatment significance: ns = not significant, *≤0.05, **≤0.01, and ***≤0.001

Table 3.2 Restricted maximum likelihood variance component estimates and significance[a] for 480 accessions across two years and 12 blocks.

Trait[b]	Variance component					
	Year	Block	Accession	Accession*Treatment	Residual	Total
YLD	0.40 ns	0.11*	38.84***	0.03 ns	0.94	40.32
TWT	0 ns	0.37 ns	397.56***	0.22 ns	10.67	408.82
PRO	2.62 ns	0.13*	124.49***	0 ns	1.14	128.39
HD	14.03 ns	0.50*	1802.10***	0.50 ns	3.71	1820.85
HT	45.31 ns	23.16*	6651.58***	2.81 ns	106.42	6826.47
LODGE	.	0.45*	166.89***	0.69 ns	1.83	169.86
LODGE%	.	44.4*	35001.8***	143.6 ns	258.6	35448.6
GW	0.01 ns	0.01 ns	1.70***	0.01 ns	0.01	1.72
GL	0.02 ns	0.01 ns	36.30***	0.01 ns	0.04	36.37
GD	0.01 ns	0.01 ns	8.74***	0.01 ns	0.01	8.76
GV	5.90 ns	0.32*	2505.89***	0.10 ns	4.8	2517.02
GWL	0 ns	0 ns	0.06***	0 ns	0	0.06
TKW	10.69 ns	0.79*	1943.96***	0.38 ns	8.14	1963.96

[a]Variance component significance: ns = not significant, *≤0.05, **≤0.01, and ***≤0.001
[b]YLD is yield in metric tons ha^{-1}; TWT is test weight in kg hl^{-1}; PRO is total grain protein percentage; HD is heading date in days; HT is height in cm; LODGE is lodging severity where 0=no lodging and 9=fully lodged; LODGE% is lodging incidence between 0-100%; GW is grain width in mm; GL is grain length in mm; GD is grain diameter in mm; GV is grain volume in mm^3; GWL is the grain width by grain length ratio; TKW is thousand kernel weight in g

Table 3.3 Trait[a] correlations and significance[b] across all treatment BLUPs.

Trait	YLD	TWT	PRO	HD	HT	LODGE	LODGE%	GW	GL	GWL	TKW
YLD	1										
TWT	0.21 ***	1									
PRO	-0.58 ***	-0.02 ns	1								
HD	0.01 ns	-0.27 ***	0.09 *	1							
HT	-0.12 **	-0.05 ns	0.14 **	0.56 ***	1						
LODGE	-0.53 ***	-0.27 ***	0.33 ***	0.09 *	0.08 ns	1					
LODGE%	-0.60 ***	-0.22 ***	0.39 ***	-0.05 ns	0.02 ns	0.89 ***	1				
GW	0.43 ***	0.16 ***	-0.42 ***	-0.37 ***	-0.14 **	-0.40 ***	-0.38 ***	1			
GL	-0.28 ***	-0.40 ***	-0.01 ns	-0.32 ***	-0.26 ***	0.32 ***	0.41 ***	0.11 ***	1		
GWL	0.48 ***	0.45 ***	-0.25 ***	0.04 ns	0.13 *	-0.53 ***	-0.59 ***	0.48 ***	-0.79 ***	1	
TKW	0.12 **	0.14 ***	-0.21 ***	-0.60 ***	-0.37 ***	-0.11 *	-0.01 ns	0.71 ***	0.62 ***	-0.09 *	1

[a] YLD is yield in metric tons ha^{-1}; TWT is test weight in kg hl^{-1}; PRO is total grain protein percentage; HD is heading date in days; HT is height in cm; LODGE is lodging intensity where 0=no lodging and 9=fully lodged; LODGE% is lodging incidence between 0-100%; GW is grain width in mm; GL is grain length in mm; GD is grain diameter in mm; GV is grain volume in mm^3; GWL is the grain width by grain length ratio; TKW is thousand kernel weight in g

[b] Significance is denoted below each correlation coefficient as: ns = not significant, *≤0.05, **≤0.01 and ***≤0.001

Table 3.4 Barley check trait[a] means[b] across environments and years.

Trait	Irrigated					Terminal drought				
	Baronesse	Harrington	Lacey	Pinnacle	Tradition	Baronesse	Harrington	Lacey	Pinnacle	Tradition
YLD	6.3a	5.4a	6.1a	6.3a	6.4a	5.6a	4.9a	5.6a	5.5a	5.2a
TWT	67.18a	64.71a	65.12a	66.44a	65.06a	65.9ab	62.8c	64.9b	66.8a	62.5c
PRO	11.6b	12.4a	12.0ab	11.0c	11.6bc	13.1ab	13.9a	13.3ab	12.3b	13.4ab
HD	63.8ab	64.9a	60.4c	61.2bc	63.6ab	65.2a	66.3a	60.6c	61.8bc	64.2ab
HT	86.7a	89.1a	93.3a	90.3a	93.2a	87.3b	95.1	94.3ab	93.6ab	97.5a
LODGE (0-9)	1.8b	3.1a	2.7ab	1.7b	1.9b	3.5a	3.2a	3.5a	2.1a	3.8a
LODGE (%)	8.8bc	29.2a	19.6ab	3.8c	9.6bc	20.0ab	24.6ab	28.3ab	9.2b	34.6a
GW	3.7abc	3.7ab	3.5c	3.8a	3.6bc	3.5b	3.5b	3.4c	3.7a	3.3d
GL	9.2a	9.2a	8.4c	9.1a	8.6b	9.3a	9.2b	8.6d	9.3a	8.8c
GD	5.9a	5.9a	5.5b	5.9a	5.6b	5.9b	5.9c	5.6d	6.0a	5.6d
GV	51.6b	50.7b	45.0c	55.6a	43.6c	49.2a	47.8b	43.9c	54.7a	40.7d
GWL	0.41d	0.41cd	0.42ab	0.43a	0.41bc	0.38b	0.39b	0.40a	0.40a	0.38b
TKW	47.2b	43.8c	40.7d	52.4a	40.1d	41.9b	39.6bc	38.3c	48.5a	34.8d

[a]YLD is yield in metric tons ha^{-1}; TWT is test weight in kg hl^{-1}; PRO is total grain protein percentage; HD is heading date in days; HT is height in cm; LODGE is lodging severity where 0 = no lodging and 9 = fully lodged; LODGE% is lodging incidence between 0-100%; GW is grain width in mm; GL is grain length in mm; GD is grain diameter in mm; GV is grain volume in mm^3; GWL is the grain width by grain length ratio; TKW is thousand kernel weight in g

[b]Means followed by a common letter are not significantly different by Tukey's HSD at $P \leq 0.05$

Table 3.5 Accessions with the highest overall BLUP trait[a] values for highest YLD and TWT and low PRO across all treatments.

Overall Rank[b]	IR Rank	DR Rank	Accession	Status	COUNTRY	STATE	YLD	TWT	PRO	DSI[c] YLD
1	1	1	PI 643339	BREEDING	United States	North Dakota	17.8	95.6	-6.2	1.0
2	4	2	PI 643267	BREEDING	United States	North Dakota	14.0	96.4	-5.0	1.7
3	6	34	PI 643310	BREEDING	United States	North Dakota	18.8	84.3	-2.6	2.0
4	22	7	PI 313113	CULTIVAR	Germany	Baden-Wurttemberg	15.1	90.5	0.3	1.8
5	5	35	PI 467811	CULTIVAR	Austria	Vienna	13.4	86.7	-5.0	2.5
6	3	5	PI 643314	BREEDING	United States	North Dakota	13.7	98.4	1.3	2.3
7	24	43	PI 643376	BREEDING	United States	North Dakota	15.4	85.6	-0.3	0.8
8	8	9	PI 643274	BREEDING	United States	North Dakota	13.4	90.4	0.6	3.3
9	11	8	PI 643383	BREEDING	United States	North Dakota	16.9	79.2	-6.9	2.7
10	33	14	PI 428497	CULTIVAR	France	Yvelines	15.2	78.3	-3.3	1.8
14	37	11	PI 422233	LANDRACE	Yemen		12.2	79.5	-0.9	-1.2
38	31	65	CIho 14395	LANDRACE	Armenia		9.7	82.3	7.0	3.2
64	126	36	PI 467486	LANDRACE	Austria	Upper Austria	9.9	79.8	9.9	-0.1
66	208	60	PI 573687	LANDRACE	Georgia		10.0	83.1	11.2	-5.5
71	82	91	PI 48641	LANDRACE	Iran		9.4	90.8	11.9	2.8
80	103	123	PI 639300	LANDRACE	Kazakhstan	Alma-Ata	7.6	67.9	-3.9	2.0
87	80	161	PI 639299	LANDRACE	Kazakhstan	Alma-Ata	9.7	64.6	3.2	2.4
88	102	61	PI 250861	LANDRACE	Iran	Khorasan	5.3	83.8	7.0	-0.9
112	334	62	PI 328907	LANDRACE	Turkey		6.6	73.5	9.0	-0.3
120	84	144	PI 436149	LANDRACE	Chile	La Araucania	9.6	78.2	14.8	2.2

[a]Traits are reported for BLUP values across treatments: YLD is yield in metric tons ha^{-1}; TWT is test weight in kg hl^{-1}; PRO is total grain protein percentage

[b]Accession rank was calculated by assigning each accession a rank based on the highest yield, highest test weight values and lowest protein value then adding the ranks together for a single value across and among treatments

[c]Accession drought susceptibility index (DSI) calculated using the formula provided by Li et al. (2012); numbers close to 1 indicate yield stability across treatments

Table 3.6 Accessions with the highest overall BLUP trait[a] values for YLD and TWT and high PRO across all treatments.

Overall Rank[b]	IR Rank	DR Rank	Accession	Status	COUNTRY	STATE	YLD	TWT	PRO	DSI-YLD[C]
1	18	1	PI 321770	CULTIVAR	Slovenia		10.8	83.2	18.4	2.1
2	15	6	PI 428413	CULTIVAR	France	Yvelines	9.8	80.6	19.4	2.4
3	9	94	PI 296455	LANDRACE	Ethiopia	Tigray	6.3	87.3	27.5	4.2
4	5	48	PI 371102	LANDRACE	Switzerland	Valais	4.1	102.6	36.8	7.0
5	1	181	PI 399482	CULTIVAR	Netherlands	South Holland	16.1	69.6	17.5	4.9
6	6	14	PI 357317	BREEDING	Denmark	Roskilde	8.9	76.0	20.5	4.1
7	3	15	PI 392458	BREEDING	South Africa	Cape Province	6.6	81.0	25.5	2.9
8	30	157	PI 41250	UNCERTAIN	Russian Federation	Leningrad	8.9	85.8	15.8	4.0
9	12	72	PI 342139	LANDRACE	Turkey	Usak	4.7	80.5	56.6	-0.5
10	17	40	PI 290234	CULTIVAR	Hungary	Heves	6.2	78.9	26.3	4.0

[a]Traits are reported for BLUP values across treatments: YLD is yield in metric tons ha^{-1}; TWT is test weight in kg hl^{-1}; PRO is total grain protein percentage

[b]Accession rank was calculated by assigning each accession a rank based on the highest yield, test weight and protein values then adding the ranks together for a single value across and among treatments

[c]Accession drought susceptibility index (DSI) calculated using the formula provided by Li et al. (2012); numbers close to 1 indicate yield stability across treatments

Table 3.7 Geographic region of accessions within each subpopulation; the total number of accessions in each subpopulation is listed with number of landraces shown in parenthesis.

Origin[a]	Subpopulation					
	1	2	3	4	5	All
Africa	4 (2)	20 (20)	27 (2)	0	5 (1)	56 (25)
Americas	8 (2)	0	21 (1)	25	18 (6)	72 (9)
Asia	29 (23)	5 (5)	15 (11)	1 (1)	6 (1)	56 (41)
Europe	12 (5)	4	61 (40)	0	199 (14)	276 (59)
Oceania	0	1	2	0	11	14
Unknown	2 (2)	1 (1)	3 (3)	0	0	6 (6)
Total	55 (34)	31 (26)	129 (57)	26 (1)	239 (22)	480 (140)

[a]Accession region within each subpopulation is based on UN M49 regional statistical use criteria

Table 3.8 Mean trait[a] values[b] within each subpopulation across all treatments.

Trait	Subpopulation					
	1	2	3	4	5	All
YLD	1.37d	0.91d	3.48c	12.23a	7.36b	5.48
TWT	68.4b	58.4c	62.2bc	79.0a	65.2bc	65.1
PRO	16.4ab	21.0a	16.0b	0.6d	11.8c	13.4
HD	33.2c	37.4bc	56.6b	17.8c	82.4a	63.5
HT	49.5b	56.9b	104.5a	63.5b	102.4a	91.9
LODGE	11.0a	8.7ab	5.7b	-12.8d	0.73c	3.0
LODGE%	177.7a	157.5a	74.0b	-229.2d	-29.5c	23.3
GW	3.0c	3.3bc	3.5b	6.0a	3.4bc	3.5
GL	11.7b	18.1a	11.4b	6.6c	6.2c	9.0
GWL	0.26c	0.11d	0.30c	0.75a	0.47b	0.40
TKW	55.2b	82.5a	50.6b	92.8a	22.4c	41.5

[a]YLD is yield in metric tons ha^{-1}; TWT is test weight in kg ha^{-1}; PRO is total grain protein percentage; HD is heading date in days; HT is height in cm; LODGE is lodging intensity where 0=no lodging and 9=fully lodged; LODGE% is lodging severity between 0-100%; GW is grain width in mm; GL is grain length in mm; GWL is the grain width and grain length ratio; TKW is thousand kernel weight in g

[b]Means followed by a common letter are not significantly different by Tukey's HSD at $P \leq 0.05$

Table 3.9 Genetic loci significantly (FDR-adjusted $P \leq 0.05$) associated with barley agronomic traits across treatments and in the irrigated and terminal drought treatments.

Locus Index	SNP Index[a]	Chr.[b]	Range (Mbp)[c]	SNP[d]	SNPs[e]	FDR-adjusted -log$_{10}$(P)	R2[f]	MAF[g]	Trait(s)[h]
1	12_30191	1H	522.4	A/**G**	1	3.29	0.05	0.15	GW, GL, GWL
2	SCRI_RS_48925	1H	538.5	**A**/C	1	4.12	0.07	0.38	PRO
3	BK_12	2H	27.2-29.8	**T**/G	9	14.4	0.15	0.15	HD, HT, GL, GWL
4	12_21476	2H	559.1	**T**/G	1	3.21	0.05	0.06	TWT, GL, GD, GV
5	12_11414	3H	28.8	**T**/G	1	1.75	0.04	0.09	GL
6	SCRI_RS_115045	3H	38.3	**A**/G	1	9.42	0.12	0.41	TWT, PRO, GL, GD, GV, GWL, TKW
7	SCRI_RS_235065	3H	538.2	**T**/C	1	6.67	0.06	0.14	GL, GD, GV, GWL
8	SCRI_RS_180891	4H	9.7	T/**C**	1	3.71	0.07	0.32	GW, GL, GD, GV, TKW
9	SCRI_RS_50995	5H	644.0-648.4	A/**G**	15	2.53	0.06	0.07	PRO
10	12_30626	6H	542.83	**T**/C	1	2.65	0.06	0.20	TWT, GL, GV, GD, GWL
11	11_10256	7H	217.3	A/**G**	1	5.51	0.05	0.05	GW, GL, GV, GD, TKW
12	12_30492	7H	264.3	A/**C**	1	1.75	0.04	0.20	GL, GD, GV
13	12_10362	7H	457.2	T/**C**	1	1.46	0.03	0.09	GL, GD, GV, TKW
14	12_30996	7H	517.6-554.2	**A**/G	3	1.34	0.04	0.09	GW, GL, GD, GV, GWL, TKW
15	12_30335	7H	576.3-591.6	A/**C**	6	2.07	0.05	0.05	TWT, GW, GL, GD, GV, GWL, TKW

[a] Single nucleotide polymorphism (SNP) index in each locus with the lowest FDR-adjusted *P*-value associated with the first trait listed; additional associated SNPs are reported in Supplementary File 3

[b]Physical chromosome location of each locus with 99 or 100% identity based on the physical barley annotation (Mascher et al., 2017)

[c]Physical locus regions in Megabase pairs (Mbp) based on the physical annotation of barley (Mascher et al., 2017)

[d]SNP with minor allele in bold and underlined

[e]Other covariate SNPs in the locus with the index SNP

[f]Proportion of phenotypic variation explained by each SNP

[g]Minor allele frequency (MAF)

[h]Traits include: YLD as yield in metric tons ha^{-1}; TWT as test weight in kg hl^{-1}; PRO as total grain protein percentage; HD as heading date in days; HT as height in cm; LODGE as lodging intensity where 0=no lodging and 9=fully lodged; LODGE% as lodging severity between 0-100%; GW as grain width in mm; GL as grain length in mm; GWL as the grain width and grain length ratio; TKW as thousand kernel weight in g

Table 3.10 Mean trait[a] values associated with the minor allele haplotype for each marker-trait locus across and within treatments.

Locus	Across treatments			Irrigated		Terminal drought		
Index	YLD[b]	TWT[c]	PRO[d]	YLD	TWT	YLD	TWT	PRO
1	0.65	60.9	19.9	2.41	62.9	3.57	60.5	19.0
2	5.38	64.3	12.3	5.93	65.1	5.2	63.8	12.5
3	6.44	75.1	14.9	6.88	68.7	6.07	73.1	13.6
4	1.70	54.9	18.4	2.70	61.3	4.25	53.5	18.5
5	1.32	58.9	18.9	2.95	62.8	3.85	57.3	18.5
6	6.93	65.8	11.4	6.84	66.3	5.88	64.9	11.7
7	5.92	63.3	11.9	6.07	65.0	5.39	62.6	12.3
8	4.85	64.4	13.7	5.59	64.9	4.69	63.7	14.0
9	7.10	60.4	8.3	7.29	62.7	5.73	62.5	8.8
10	4.07	66.9	16.0	4.93	65.6	4.65	65.6	15.6
11	0.22	75.2	22.6	1.95	66.4	3.40	80.0	20.7
12	5.53	74.3	13.0	5.82	68.2	5.43	75.4	12.6
13	3.43	58.2	14.8	4.06	63.2	4.61	57.9	14.9
14	3.40	59.1	17.5	4.11	62.3	4.66	58.9	18.0
15	2.95	60.0	15.1	4.14	62.5	4.11	62.0	14.8

[a]YLD is yield in metric tons ha^{-1}; TWT is test weight in kg hl^{-1}; PRO is total grain protein percentage
[b]Mean YLD across treatments is 5.48; YLD-IR treatment is 5.84 and YLD-DR treatment is 5.12
[c]Mean TWT across treatments is 65.1; TWT-IR treatment is 65.6 and TWT-DR treatment is 64.5
[d]Mean PRO across treatments is 13.4; PRO-IR treatment is 12.1 and PRO-DR treatment is 13.5

Table 3.11 Proportion of accessions represented by each subpopulation within each marker-trait loci minor allele haplotype.

Locus Index	Subpopulation				
	1	2	3	4	5
1	0.36	0.40	0.24	0	0
2	0.20	0.09	0.27	0.09	0.34
3	0.14	0.03	0.65	0	0.19
4	0.03	0.93	0.03	0	0
5	0.17	0.71	0.07	0.02	0.02
6	0.06	0.11	0.16	0.12	0.54
7	0.07	0	0.52	0.26	0.14
8	0.12	0	0.41	0	0.47
9	0.20	0.05	0.05	0	0.70
10	0.34	0	0.49	0.01	0.16
11	0.61	0.22	0.13	0	0.04
12	0.24	0.05	0.34	0.25	0.12
13	0.02	0.40	0.17	0.02	0.38
14	0.06	0.56	0.28	0.06	0.03
15	0.06	0.81	0.13	0	0

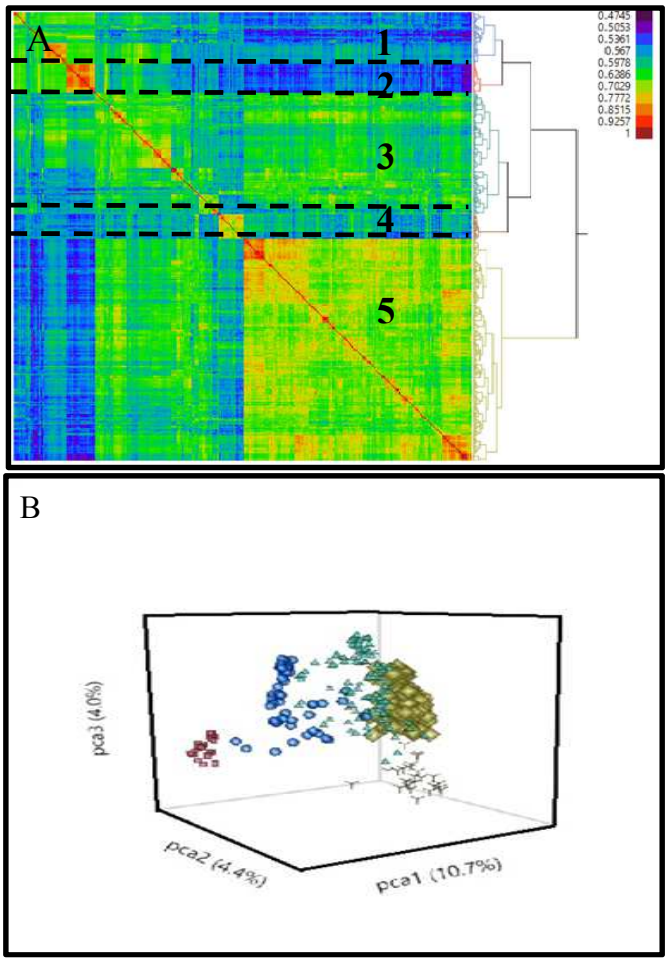

Fig. 3.1 Five two-row spring habit barley subpopulations visualized by **A** Ward hierarchical clustering based on an identity-by-state matrix and **B** principal component analysis with three dimensions where five subpopulations are denoted as 1: blue circles, 2: red squares, 3: green triangles, 4: brown Ys, and 5: yellow diamonds.

APPENDICIES

CHAPTER 2 APPENDICIES

Table A2.1 Previously reported common bunt (CB) and dwarf bunt (DB) resistance loci.

Chromosome	Position (cM)	Disease	Resistance Source	Reference
1A	74-76	DB CB	IDO444; PI 476212	Chen et al. (2016); Mourad et al. (2018)
1BS	19-20	CB	'AC Domain'; 'Blizzard'; PI 476212	Fofana et al. (2008); Wang et al. (2009)
1BS	45	CB	'Trintella'	Dumalasová et al. (2012)
1BS	43.8	CB	'Carberry'	Singh et al. (2016)
1BS	47-55	CB	'CDC Go'	Zou et al. (2017)
1B	76-96	CB	'AC Domain'	Fofana et al. (2008); Bhatta et al. (2018); Mourad et al. (2018)
2A		CB		Bhatta et al. (2018)
2BS	13-15.2	DB	IDO444; PI 476212	Chen et al. (2016)
2BL		CB		Bhatta et al. (2018); Mourad et al. (2018)
3AL	200-206	CB	'CDC Go'	Zou et al. (2017); Mourad et al. (2018)
3BS		CB		Mourad et al. (2018)
3DL		CB		Bhatta et al. (2018)
4AL		CB		Bhatta et al. (2018); Mourad et al. (2018)
4BS	62.3-89.4	CB	'Carberry'	Singh et al. (2016)
4DS	7.44	CB	'Carberry'	Singh et al. (2016)
5AL		CB		Mourad et al. (2018)
5BS	0	CB	'Trintella'	Dumalasová et al. (2012)
5BL		CB		Mourad et al. (2018)
5DL		CB		Mourad et al. (2018)
6AL		CB		Mourad et al. (2018)
6BL		CB		Mourad et al. (2018)

6DS	9.7	CB	'AC Cadillac'; 'AC Taber'; PI 178383	Menzies et al. (2006); Singh et al. (2016)
6DL	125-132	CB	PI 178383; PI 554099	Steffen et al. (2017)
7AS	44	CB	'Trintella'	Dumalasová et al. (2012); Bhatta (2018)
7AL	117-120	CB	'AC Domain'	Fofana et al. (2008); Bhatta et al. (2018); Mourad et al. (2018)
7BS	9.6	CB	'McKenzie'	Knox et al. (2013)
7BS	13	CB	'Trintella'	Dumalasová et al. (2012); Mourad et al. (2018)
7BL		CB		Bhatta et al. (2018); Mourad et al. (2018)
7DS	1	DB	IDO444; PI 476212	Chen et al. (2016)
7DL	46.6	CB	'Carberry'	Singh et al. (2016)
7DL		CB		Mourad et al. (2018)

Table A2.2 Restricted maximum likelihood variance component estimates and significance values for 292 wheat accessions across four trials.

Effect	Variance ratio	Variance component	Standard error	*P*-value	Percent
Accession	5.02	1601.08	141.43	<.0001	74.91
Trial	0.24	76.21	63.32	0.2287	3.57
Accession*Trial	0.44	141.21	27.76	<.0001	6.61
Residual		318.75	21.97		14.91
Total		2137.25	155.86		100

Table A2.3 Number, mean, standard deviation and range of BLUE values for each marker-trait haplotype group combination.

Marker-trait group	Number	Mean	Standard Deviation	Min.	Max
6D1	210	40.5	39.2	0.1	122.1
6D2	33	16.3	19.8	3.8	82.1
6D1+6D2	32	16.1	20.1	3.8	82.1

Table A2.4 Proportion of 246 bread wheat accessions with resistant allele haplotypes represented by each subpopulation.

DB Marker-trait association	Subpopulation					
	1	2	3	4	5	6
DB-6D1	0.11	0.15	0.22	0.11	0.15	0.24
DB-6D2	0.09	0.12	0.06	0.48	0.03	0.21

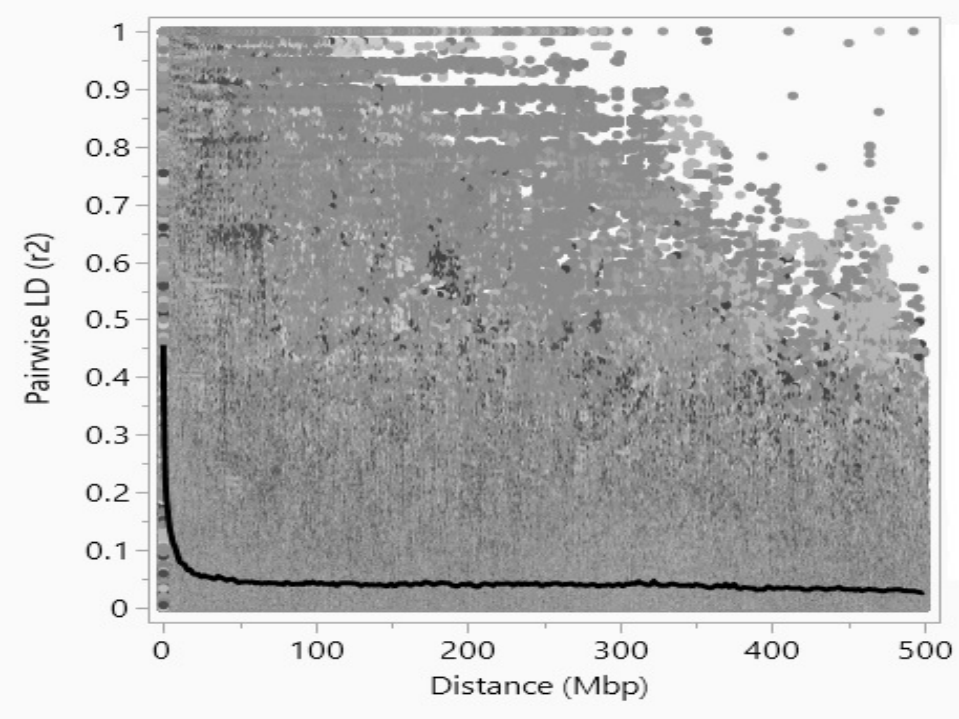

Figure A2.1 Genome-wide linkage disequilibrium (LD) scatter plot showing correlations (r^2) between intrachromosomal SNP marker pairs as a function of inter-marker physical position; a smoothing spline (black line) with lambda equal to 10,000 is fit to the data.

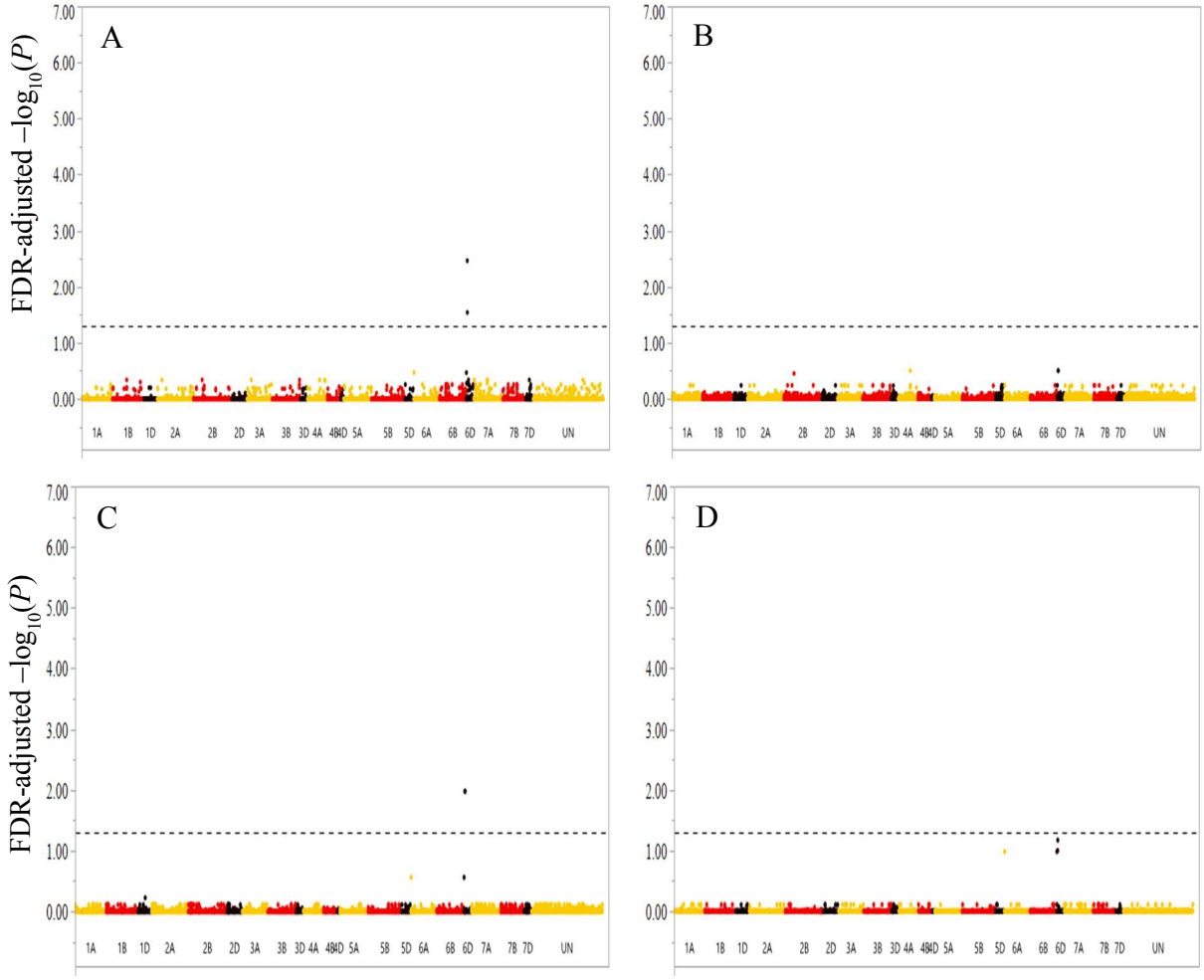

Figure A2.2 Manhattan plots showing FDR-adjusted associations between 19,281 SNP markers and DB normalized incidence in **A**: GRIN, **B**: 2017, **C**: 2018 and **D**: 2019 data sets across 246 wheat accessions; the horizontal dashed line indicates an FDR-adjusted significance threshold of $P = 0.05$; A-,B- and D-genome SNP markers are represented by yellow, red and black dots, respectively.

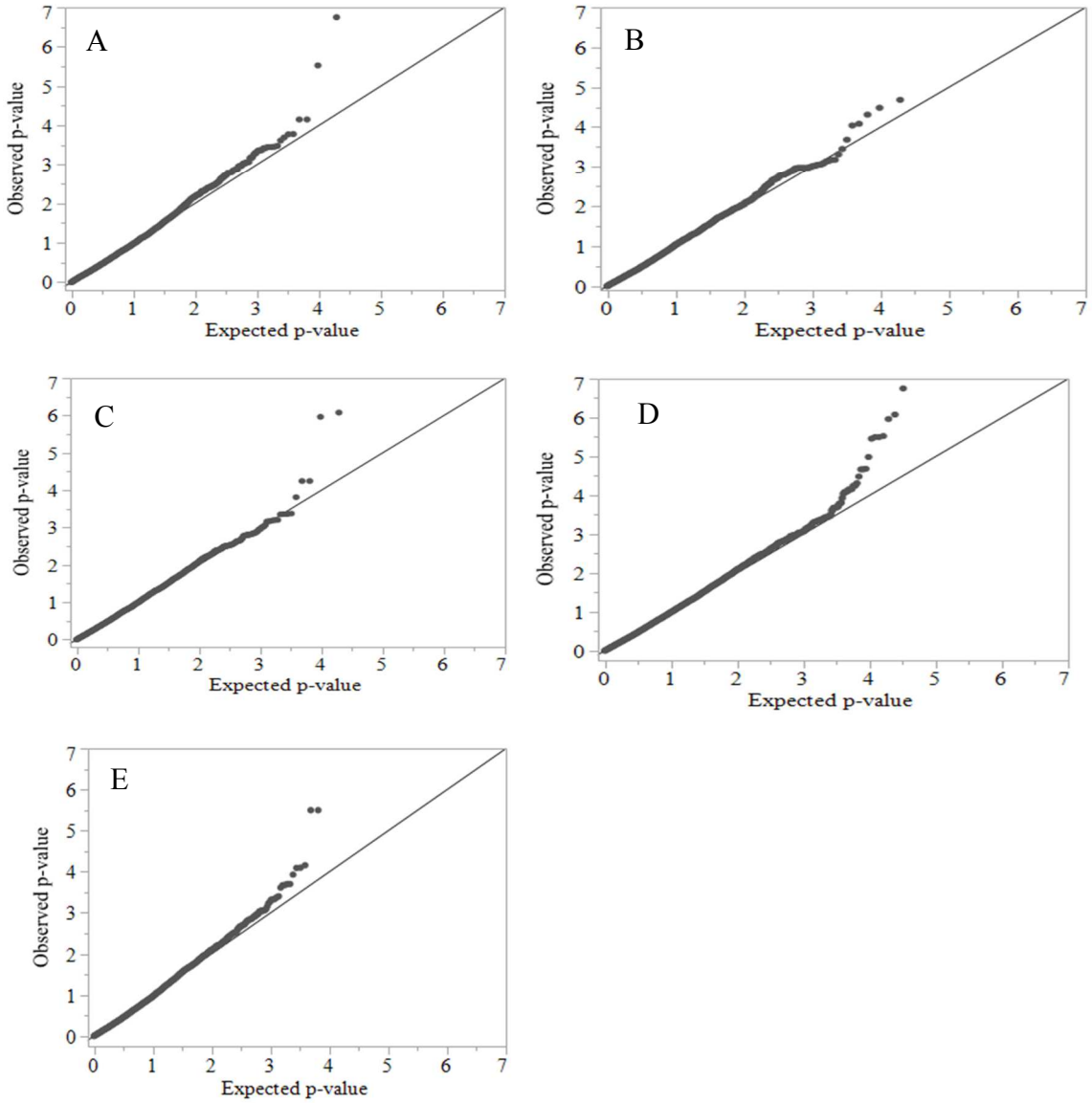

Figure A2.3 Quantile-quantile plots showing expected vs observed negative $\log_{10}$(P) values generated from the GWAS for 19,281 SNPs in **A**: GRIN, **B**: 2017, **C**: 2018, **D**: 2019 and **E**: BLUE data sets across 246 wheat accessions.

Supplementary File A2.1 Accessions and their associated descriptors and phenotypes.

Supplementary File A2.2 All 19,281 SNP markers, their associated annotation information, $-\log_{10}(P)$, FDR-adjusted $-\log_{10}(P)$ and R^2 values for each DB NI data set.

Supplementary File A2.3 Four SNP markers that were significant at an FDR-adjusted ($P < 0.05$) in at least one trial mean or BLUE with associated annotation information, $-\log_{10}(P)$ values, FDR-adjusted $-\log_{10}(P)$ values and R^2 values for each DB NI data set.

Supplementary File A2.4 All 246 accessions with corresponding trial means, BLUE values and marker-trait group haplotypes detected in this study and three QTL reported in previous studies.

CHAPTER 3 APPENDICIES

Supplementary File A3.1 Accession ID, subpopulation designation, improvement status, origin and BLUP values for each measured trait across and within treatments.

Supplementary File A3.2 2014 and 2017 AGRIMET soil water availability from seedling emergence to harvest for the irrigated and terminal drought treatments.

Supplementary File A3.3 All SNP markers used in this study, their corresponding negative log10 *P*-values, FDR-adjusted *P*-values and R2 association values, and SNP markers significantly (FDR-adjusted $P < 0.05$) associated with two-row spring barley agronomic traits grown under irrigated and terminal drought conditions.

Supplementary File A3.4 Two-row spring barley agronomic trait correlation coefficients and probabilities across, within and amongst treatments.