Genomic Prediction and Genetic Dissection of Yield-Related Traits in Soft Red Winter Wheat

Brian Phillip Ward

Dissertation submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
In
Crop and Soil Environmental Sciences

Carl A. Griffey, Chair
M.A. Saghai Maroof
Wade Thomason
Jason Holliday

April 10, 2017
Blacksburg, VA

Keywords: wheat, *Triticum aestivum*, genomic prediction, genomic selection, quantitative trait locus, genome-wide association study, yield, grain yield

Copyright 2017 Brian P. Ward

Genomic Prediction and Genetic Dissection of Yield-Related Traits in Soft Red Winter Wheat

Brian P. Ward

ABSTRACT (Academic)

In multiple species, genome-wide association (GWA) studies have become an increasingly prevalent method of identifying the quantitative trait loci (QTLs) that underlie complex traits. Despite this, relatively few GWA analyses using high-density genomic markers have been carried out on highly quantitative traits in wheat. We utilized single-nucleotide polymorphism (SNP) data generated via a genotyping-by-sequencing (GBS) protocol to perform GWA on multiple yield-related traits using a panel of 329 soft red winter wheat genotypes grown in four environments. In addition, the SNP data was used to examine linkage disequilibrium and population structure within the testing panel. The results indicated that an alien translocation from the species *Triticum timopheevii* was responsible for the majority of observed population structure. In addition, a total of 50 significant marker-trait associations were identified. However, a subsequent study cast some doubt upon the reproducibility and reliability of plant QTLs identified via GWA analyses. We used two highly-related panels of different genotypes grown in different sets of environments to attempt to identify highly stable QTLs. No QTLs were shared across panels for any trait, suggesting that QTL-by-environment and QTL-by-genetic background interaction effects are significant, even when testing across many environments. In light of the challenges involved in QTL mapping, prediction of phenotypes using whole-genome marker data is an attractive alternative. However, many evaluations of genomic prediction in crop species have utilized univariate models adapted from animal breeding. These models cannot directly account for genotype-by-environment interaction, and hence are often not suitable for use with lower-heritability traits assessed in multiple environments. We sought to test genomic prediction models capable of more *ad-hoc* analyses, utilizing highly unbalanced experimental designs consisting of individuals with varying degrees of relatedness. The results suggest that these designs can successfully be used to generate reasonably accurate phenotypic predictions. In addition, multivariate models can dramatically increase predictive accuracy for some traits, though this depends upon the quantity and characteristics of genotype-by-environment interaction.

Genomic Prediction and Genetic Dissection of Yield-Related Traits in Soft Red Winter Wheat

Brian P. Ward

ABSTRACT (General Audience)

Quantitative traits are those traits that can display a wide range of variability within a population of individuals. These traits are influenced by the interaction of many different genes, and are also influenced by the environment to varying degrees. Traditionally, geneticists who studied quantitative traits had to rely on statistical models, while the biological causes of variation in the expression of these traits remained largely unknown. However, the advent of DNA marker technology granted geneticists the ability to identify specific regions of the genome that highly influence quantitative traits. Many studies have since attempted to find these *quantitative trait loci* (QTLs) across a wide range of traits and species. However, we are faced with something of a paradox when we attempt to find QTLs. Theory tells us that an idealized, truly quantitative trait arises due to the effects of many genes, each with an infinitesimal effect on the trait in question. Therefore, the more quantitative a trait, the fewer QTLs we should expect to find. In addition, QTLs may not be reliable, due to the effects of different environments and different genetic backgrounds within a population. A more recent trend involves using all available marker data simultaneously to predict a particular line's performance. This method entails ignoring the genomic underpinnings of a trait, and instead focusing solely on its expression, much like classical quantitative genetics. The obvious downside of this method is that it cannot be used to increase our understanding of what is giving rise to the variations in the trait's expression that we observe. The studies described in this dissertation were designed to 1) test whether we could identify QTLs for highly quantitative yield-related traits in winter wheat, 2) test the reliability of identified QTLs, and 3) use the DNA marker data to instead generate predictions of line performance. The results indicate that while we can identify QTLs for highly quantitative traits in winter wheat, these QTLs may not be very reliable. Therefore, predictive models may be a good alternative to identifying QTLs, and these methods can be readily implemented within breeding programs.

DEDICATION

I would like to dedicate this dissertation to my parents, Michael and Marcia, my sister Jena, and my partner Katie. They have always stuck by me through all life's trials and tribulations. Needless to say, I wouldn't be writing this if it weren't for their love and support.

ACKNOWLEDGMENTS

I owe a great deal of thanks to the many individuals who I was privileged to work with during graduate school. Without their help, this dissertation could never have come together.

First and foremost, I would like to thank my advisor, Dr. Carl Griffey. Dr. Griffey taught me most of what I know about plant breeding, and always sought to emphasize real-world applications over theory. In addition, Dr. Griffey helped with much of the data collection for my project out in the field; I have never heard of another advisor who embodies "leading by example" to quite the same degree.

In addition, I would like to thank everyone affiliated with the Virginia Tech Small Grains Breeding group whom I had the privilege of working with during my time at Virginia Tech, including Wynse Brooks, Limei Liu, Subas Malla, Luciana Rosso, Anthony Christopher, John Seago, Jon Light, Steve Potorff, Neal Carpenter, Jordan Ulrich, Kyle Brasier, Nick Meier, Greg Berger, Tiffany Sykes, Emily Wright, and Camron Clark. They, more than anyone else, helped with all of the logistics and data collection that made these projects possible.

I would also like to thank the staff at the Virginia Tech EVAREC station in Warsaw, VA, including Bob Pitman, Mark Vaughn, David "Scooter" Dunaway, Lyn Barrack, Mary Beahm, and Robin Markham. They helped with much of the data collection that occurred in Warsaw, and kept me apprised of conditions "on the ground" there.

Additionally, I would like to thank all of the friends who I made in Blacksburg along the way. Emma Flemmig, Rob Norris, Sarah Jonson, and Ben Owsley were all housemates of mine during a period of my life I now refer to with the phrase "overcoming adversity." Ben Vernasco is the universal little brother; John Lewis serendipitously entered my life much like a neighbor character in a sitcom. Neal Carpenter and Kyle Brasier shared in much of my requisite grad school angst along the way.

Finally, I would like to also thank the members of my committee; Dr. Wade Thomason, Dr. M.A. Saghai-Maroof, and Dr. Jason Holliday. Their critiques of my writing greatly helped in improving the drafts of this dissertation.

ATTRIBUTIONS

Many colleagues and collaborators aided in the collection, analysis, and interpretation of data included in my dissertation. Descriptions of their contributions are below.

Chapter I: Genome-Wide Association Studies for Yield-Related Traits in Soft Red Winter Wheat Grown in Virginia

Frederic Kolb, PhD is currently a professor in the Department of Crop Sciences at the University of Illinois in Urbana, IL. Dr. Kolb's breeding program contributed germplasm used in the study.

David Van Sanford, PhD is currently a professor in the Department of Plant and Soil Sciences at the University of Kentucky in Lexington, KY, and was a co-PI for the USDA-NIFA *Triticeae* Coordinated Agricultural Project (TCAP). Dr. Van Sanford's breeding program contributed germplasm used in the study.

Clay Sneller, PhD is currently a professor in the Department of Horticulture and Crop Science at the Ohio State University in Wooster, OH, and was a co-PI for the USDA-NIFA *Triticeae* Coordinated Agricultural Project (TCAP). Dr. Sneller developed the original design for the TCAP Allele-Based Breeding study. The study described in Chapter I focuses on a subset of this larger experiment.

Gina Brown-Guedira, PhD is a professor in the Department of Crop and Soil Sciences at North Carolina State University in Raleigh, NC, director of the USDA Eastern Regional Small Grains Genotyping Lab, and was a co-PI for the USDA-NIFA *Triticeae* Coordinated Agricultural Project (TCAP). Dr. Brown-Guedira's lab generated the genotyping-by-sequencing genotypic data used in this chapter.

Priyanka Tyagi, PhD is a postdoctoral research scholar in the USDA Eastern Regional Small Grains Genotyping Lab in Raleigh, NC. Dr. Tyagi generated the genotyping-by-sequencing data used in this chapter.

Carl Griffey, PhD is currently the W.G. Wysor professor of crop breeding and genetics at Virginia Tech. Dr. Griffey was a co-PI for the USDA-NIFA *Triticeae* Coordinated Agricultural Project (TCAP), and a PI for several state-supported grants that funded this work. He provided guidance on experiment design, crop management and data analysis, in addition to providing editorial comments.

Chapter II: Use of Multivariate Genomic Selection Models in Unbalanced Early-Generation Wheat Yield Trials

Frederic Kolb, PhD is currently a professor in the Department of Crop Sciences at the University of Illinois in Urbana, IL. Dr. Kolb's breeding program contributed germplasm used in the study.

David Van Sanford, PhD is currently a professor in the Department of Plant and Soil Sciences at the University of Kentucky in Lexington, KY, and was a co-PI for the USDA-NIFA *Triticeae*

Coordinated Agricultural Project (TCAP). Dr. Van Sanford's breeding program contributed germplasm used in the study.

Clay Sneller, PhD is currently a professor in the Department of Horticulture and Crop Science at the Ohio State University in Wooster, OH, and was a co-PI for the USDA-NIFA *Triticeae* Coordinated Agricultural Project (TCAP). Dr. Sneller developed the original design for the TCAP Allele-Based Breeding study. The study described in Chapter I focuses on a subset of this larger experiment.

Gina Brown-Guedira, PhD is a professor in the Department of Crop and Soil Sciences at North Carolina State University in Raleigh, NC, director of the USDA Eastern Regional Small Grains Genotyping Lab, and was a co-PI for the USDA-NIFA *Triticeae* Coordinated Agricultural Project (TCAP). Dr. Brown-Guedira's lab generated the genotyping-by-sequencing genotypic data used in this chapter.

Priyanka Tyagi, PhD is a postdoctoral research scholar in the USDA Eastern Regional Small Grains Genotyping Lab in Raleigh, NC. Dr. Tyagi generated the genotyping-by-sequencing data used in this chapter.

Carl Griffey, PhD is currently the W.G. Wysor professor of crop breeding and genetics at Virginia Tech. Dr. Griffey was a co-PI for the USDA-NIFA *Triticeae* Coordinated Agricultural Project (TCAP), and a PI for several state-supported grants that funded this work. He provided guidance on experiment design, crop management and data analysis, in addition to providing editorial comments.

Chapter III: Genome-Wide Association Studies in Two Panels of Elite Soft Winter Wheat Lines

Mao Huang, PhD is currently a postdoctoral research scholar at the Ohio State University in Wooster, OH. Dr. Huang processed phenotypic data for the elite panel, and generated trait stability estimates for both the elite panel and yield validation panel.

David Van Sanford, PhD is currently a professor in the Department of Plant and Soil Sciences at the University of Kentucky in Lexington, KY, and was a co-PI for the USDA-NIFA *Triticeae* Coordinated Agricultural Project (TCAP). Dr. Van Sanford's breeding program collected phenotypic data in all Kentucky environments included in the study.

Anne McKendry, PhD is currently a professor in the Division of Plant Sciences at the University of Missouri in Columbia, MO, and was a co-PI for the USDA-NIFA *Triticeae* Coordinated Agricultural Project (TCAP). Dr. McKendry's breeding program collected phenotypic data in all Missouri environments included in the study.

Clay Sneller, PhD is currently a professor in the Department of Horticulture and Crop Science at the Ohio State University in Wooster, OH, and was a co-PI for the USDA-NIFA *Triticeae* Coordinated Agricultural Project (TCAP). Dr. Sneller's breeding program collected phenotypic data in all Ohio environments included in the study.

Gina Brown-Guedira, PhD is a professor in the Department of Crop and Soil Sciences at North Carolina State University in Raleigh, NC, director of the USDA Eastern Regional Small Grains Genotyping Lab, and was a co-PI for the USDA-NIFA *Triticeae* Coordinated Agricultural Project (TCAP). Dr. Brown-Guedira's lab generated the genotyping-by-sequencing genotypic data used in this chapter.

Priyanka Tyagi, PhD is a postdoctoral research scholar in the USDA Eastern Regional Small Grains Genotyping Lab in Raleigh, NC. Dr. Tyagi generated the genotyping-by-sequencing data used in this chapter.

Carl Griffey, PhD is currently the W.G. Wysor professor of crop breeding and genetics at Virginia Tech. Dr. Griffey was a co-PI for the USDA-NIFA *Triticeae* Coordinated Agricultural Project (TCAP), and a PI for several state-supported grants that funded this work. He provided guidance on experiment design, crop management and data analysis, in addition to providing editorial comments.

TABLE OF CONTENTS

LIST OF TABLES

Chapter I

Chapter II

Chapter III

Appendices

LIST OF FIGURES

Chapter I

Chapter II

Chapter III

Appendices

CHAPTER I

Genome-Wide Association Studies for Yield-Related Traits in Soft Red Winter Wheat Grown in Virginia

B. Ward, F. Kolb, D. Van Sanford, G. Brown-Guedira, P. Tyagi, C. Sneller, and C. Griffey

Abstract

Grain yield remains the trait that is of paramount importance in the breeding of all cereals. In wheat (*Triticum aestivum*), grain yield has steadily increased since the Green Revolution, though the current rate of increase is not forecasted to keep pace with demand due to growing world population and affluence. In addition, recent reports suggest that the rate of yield increases may be plateauing in some regions. Despite this, genome-wide association studies (GWAS) for yield and yield-related traits using high-density markers remain rare in wheat, and even more so in winter wheat. A genome-wide-association study was carried out on a population of 324 soft red winter wheat lines across a total of four rain-fed environments in the state of Virginia using single-nucleotide polymorphism (SNP) marker data generated by genotyping-by-sequencing (GBS). Two separate mixed linear models were used to identify significant marker-trait associations (MTAs). The first was a single-locus model (GCTA), utilizing a leave-one-chromosome-out approach to estimating kinship. The second was a sub-setting kinship multi-locus method (FarmCPU). The GCTA model identified seven significant MTAs for various yield-related traits, while the FarmCPU model identified 49 significant MTAs. The results indicate promising avenues for increasing grain yield by exploiting variation in traits relating to the number of grains per unit area, as well as phenological traits influencing stay-green characteristics of genotypes.

Introduction

Worldwide, wheat has the fourth-highest production of all crops, with a net production value that is second-highest of any crop (Food and Agriculture Organization of the UN, 2013). In addition, wheat maintains the highest harvested acreage of any crop worldwide (USDA

Foreign Agricultural Service, 2015). Current best estimates by the United Nations forecast world population reaching close to 10 billion by 2050 – an increase of roughly 40% over the current population (United Nations Department of Economic and Social Affairs, Population Division, 2015). However, worldwide demand for food production is expected to increase by a higher percentage over the same period, due mainly to increasing demand for calories from inefficient food sources (e.g. livestock). Indeed, global per capita meat production increased by more than 60% in the last forty years of the 20th century (Tilman et al., 2002). Hence, worldwide cereal production will have to increase by an estimated 50% over the period ending in 2050 to keep pace with demand (Bruinsma, 2011). Such a scenario requires continuing genetic gains in yield potential of approximately 1.1% per year.

Sharma, et al. (2012) estimated a historical average increase in grain yields in spring wheat of 0.65% per year when analyzing data across 15 years and 919 environments, while a recent study involving winter wheat in the Eastern United States estimated yearly increases in grain yield between 0.56% and 1.41%, depending upon environment (Green et al., 2012). However, evidence suggests that anthropogenic climate change will have an adverse effect on yield trends in the future. Some troubling recent reports suggest that wheat grain yield is plateauing in some regions. This is especially evident in Europe, Japan, and the Indian subcontinent (Lin and Huybers, 2012). While a large portion of the wheat yield plateau in Europe is explained by policy changes implemented within the European Union, climate change has been shown to have a statistically significant effect as well (Moore and Lobell, 2015). As a cool-season crop, wheat is expected to suffer relatively high losses due to current warming trends over the next few decades (Lobell et al., 2011).

Dissecting Grain Yield

Thus, there is justified concern over the ability of wheat grain yields to keep pace with demand in the coming decades. Increasing the rate of genetic gains over time for quantitative traits such as grain yield will require either an increase in direct selection efficiency, or a better understanding of how grain yield may best be increased through functional dissection of yield *per se* and related component traits. Yield *per se* may be functionally divided into sub-components in various ways. Donald developed a hypothesis-driven method of plant breeding he termed "ideotype breeding" (1968), in which he argued for a rational approach to improving performance for various traits, as opposed to simpler methods that he termed "defect removal" and "selection for yield". Critically, Donald's model suggested that selection for yield *per se* without any regard to *how* yield was increased might be unsustainable in the long term, and that plant breeders would be better served leveraging plant physiology theory to improve other traits, thereby improving yield indirectly.

Prior to this work, Grafius (1956) had developed a simple multiplicative model for explaining grain yield in oats (*Avena sativa*) as a function of panicle and grain traits, whereby yield is represented as the volume of a rectangular parallelepiped with component traits as each of its three dimensions:

$$YLD = PPA \times KPP \times AKW \quad (1)$$

Where yield potential (YLD), is a function of panicles per unit area (PPA), average kernels per panicle (KPP), and average kernel weight (AKW). However, a reliable estimate of kernels per unit area can combine the panicles per unit area and average kernels per panicle terms from above to yield a simplified equation:

$$YLD = KPA \times AKW \quad (2)$$

Where yield potential (YLD), is a function of kernels per unit area (KPA) and average kernel weight (AKW).

Alternatively, yield may be partitioned into both its biological aspects (total biomass), and those specific portions of biomass that are of interest to the breeder. Niciporovic (1956) defined "biological yield" as the total biomass of a plant at time of harvest (though often only the above-ground biomass is considered). However, in practice only a portion of this total yield is retained at harvest, whether it be grain, fiber, or fruit. This portion, which is of interest to the breeder, the producer and the consumer, he termed the "economic yield." Building upon this partitioning of yield, Donald (1962) defined the term "harvest index" as the ratio of harvested seed weight to total above-ground biomass to quantify the economic yield in crops that are harvested for grain. Thus the grain yield of crops may simply be represented as the proportion of biological yield that goes towards grain production:

$$YLD = HI \times BIOM \quad (3)$$

Where yield potential (YLD) is a function of harvest index (HI), and above-ground biomass (BIOM). Given this relationship, it follows that a plant breeder may increase grain yield by increasing harvest index, biomass, or both. However, there is a limit to how much harvest index may be increased, as grain production relies on the presence of adequate amounts of stem and leaf tissue for photosynthesis and physical support. In many crops, including wheat, increases in grain yield over the course of the 20th century were primarily brought about by increases in harvest index. Maize (*Zea mays*) is an exception, as typical harvest index values were reaching theoretical maximums early in the 20th century, and subsequent gains in yield came via gains in biomass (Hay, 1995). In wheat, the widespread introgression of *Rht* dwarfing genes into most

breeding germplasm brought about the greatest increases in harvest index. Austin (1980) suggested an upper limit of ~60% for harvest index in wheat.

Thus, future yield gains in wheat will primarily be realized by increasing biomass. Biomass itself may be defined as a product of light interception and radiation use efficiency, such that we arrive at the following equation for yield potential:

$$YLD = HI \times LI \times RUE \quad (4)$$

Where yield potential (YLD) is a function of harvest index (HI), light interception (LI), and radiation use efficiency (RUE). Intuitively, RUE can be defined as biomass production per unit of absorbed solar energy.

Light interception may be further decomposed into the amount of photosynthetically active radiation (PAR) absorbed, and the duration of absorption. Reynolds et al. (2005) note that the amount of light interception is already nearly maximized in many wheat cultivars during the period after canopy closure and prior to leaf senescence. This leaves only two avenues for further increases in light interception: faster canopy establishment through increased early-season vigor or later senescence, i.e. the "stay-green" trait (Silva et al., 2001).

As for increasing radiation use efficiency, one possibility is increasing the rate of photosynthesis, or decreasing net photorespiration. Notably, there has been little genetic gain in photosynthetic rate over time, and in many cases there has actually been a decrease in this trait value over generations (Richards, 2000). However, while increases in photosynthetic capacity and the rate of photosynthesis offer some potential for future yield gains, a wide body of literature suggests that wheat is primarily sink-limited with respect to photosynthetic assimilates (reviewed in Borras et al. [2004]). For instance, the manual removal of several seeds from spikes partway through the grain-filling period does not generally lead to any significant increase in

mass of the remaining seeds (Slafer and Savin, 1994). In addition, the late-season decline in starch accumulation in seeds associated with physiological maturity is primarily triggered by a drop in the synthetic capacity of the endosperm, rather than any decline of photosynthetic assimilates available (Jenner and Rathjen, 1975). Hence the greatest gains in radiation use efficiency in the immediate future may arise from efforts to increase the quantity of assimilates that can be translocated into the endosperm, rather than increasing the quantity of assimilates available.

Supporting this strategy, several studies have suggested that maximizing the number of seeds per unit area is critical for maximizing yield (Reynolds et al., 2005). This would presumably provide a ready solution for circumventing the current sink limitations imposed during grain fill. The two possible routes for increasing seeds per unit area are to: 1) increase the average number of seeds per head, or 2) increase the number of heads per unit area. Of course if harvest index is already near its limit, either of these two approaches will necessarily entail an increase in biomass.

Genome-Wide Association Studies

Genome-wide association studies (GWAS) and linkage mapping are the two predominate methods employed in plant breeding for associating phenotypic variation with underlying genetic variation. Of these two methods, GWAS is generally less laborious and faster to implement, as testing may be carried out on existing panels of individuals, rather than on specific populations which must be generated via crossing and then inbred to a sufficient level of homozygosity. Panels may be assembled from highly diverse and unadapted germplasm, or from more adapted germplasm used in breeding programs. GWAS also offers higher resolution due to many more

ancestral gene recombinations within the testing panel, as opposed to only one or a few meiotic recombinations in a linkage mapping population. However, allele frequency is a primary factor limiting power in GWAS studies; marker-trait associations (MTAs) will be difficult or impossible to detect if causal variants are rare within the testing population (Myles et al., 2009).

Some early GWAS papers in plants came to the surprising conclusion that relatively small sample sizes and low marker densities were adequate to reliably detect MTAs (Atwell et al., 2010), as contrasted to human studies, in which low power is often inevitable even with many thousands of individuals and millions of markers. This is attributable to the unique evolutionary histories of many crop plants whereby recent domestication has resulted in genetic bottlenecks, generally causing extensive linkage disequilibrium (LD) and high minor allele frequencies. Thus many crop GWA studies can achieve reasonable power with fewer resources, the limitation being that mapping resolution will be limited in cases where LD is high and effective population size is low (Hamblin et al., 2011). However, GWAS in plants brings its own set of challenges; most notably increased false-positive risk due to complex genetic histories and population structure (Atwell et al., 2010).

In the process of adapting GWAS from its original implementation in human studies to use in plants, maize geneticists developed the mixed linear model method to address the confounding effects of kinship and population stratification, which may be highly pronounced in plant populations that have undergone artificial selection (Yu et al., 2006). Since that time, mixed linear models have become the standard for performing GWAS in populations with high levels of relatedness (whether these relationships are cryptic in nature, or else known *a priori*), and many variations of these models have been developed. Compressed mixed linear models (CMLM) seek to improve power by assigning individuals to clusters based upon kinship (Zhang

et al., 2010). The multi-locus mixed model (MLMM) fits SNPs that are highly associated with phenotype as covariates, using a stepwise regression method (Segura et al., 2012).

A recent development has been the advent of mixed models that utilize a subset of the total markers to estimate kinship, which can have the effect of increasing statistical power. The Fast-LMM-Select algorithm was the first implementation of such a model, in which a linear regression of SNPs on the tested phenotype is carried out prior to running the LMM to identify SNPs that are highly associated with the phenotype (Listgarten et al., n.d.). This technique makes use of the equivalence between a linear mixed model using a given set of markers to estimate genetic relationships, and a Bayesian linear regression using these same SNPs fitted as covariates. Only markers identified as highly associated with the phenotype in the preliminary linear regression stage are used to construct the kinship matrix, the exception being when an associated marker is in high LD with the marker being tested for association, in which case it is discarded from the kinship matrix. SNPs that are used to construct the kinship matrix may be referred to as pseudo quantitative trait nucleotides (pseudo-QTNs). The SUPER algorithm (Q. Wang et al., 2014), implemented in version 2 of the popular suite of GWAS tools GAPIT (Tang et al., 2016), generalizes the Fast-LMM-Select method for better applicability in a wider range of organisms by empirically determining LD between tested markers and pseudo-QTNs, rather than inferring LD using a distance-based cutoff as in Fast-LMM-Select. Finally, the Fixed and random model Circulating Probability Unification (FarmCPU) algorithm utilizes two models in an iterative fashion: a fixed-effect model which incorporates a multi-locus mixed model fitting the tested SNP with associated SNPs as covariates, and a random-effect model that is used to estimate kinship from the associated SNPs (Liu et al., 2016). Despite these recent advances, there has been disagreement as to whether kinship matrix-sub-setting methods adequately control for

false positives, particularly in cases of complex population stratification (Widmer et al., 2014; Yang et al., 2014). However, more recent models have included genotypic principle components to mitigate test statistic inflation due to population stratification (Tucker et al., 2014).

Genotyping-by-Sequencing

Genotyping-by-sequencing (GBS) is one of several complexity-reduction genotyping techniques in which genomic DNA is subjected to multiple restriction digests, with the generated fragments then being bridge amplified and sequenced on next-generation sequencing platforms. In addition, the use of a methylation-sensitive restriction enzymes can exclude many epigenetically-silenced genomic regions (e.g. repetitive elements) from the generated fragments (Elshire et al., 2011). This development has been especially critical for the implementation of GBS protocols in wheat, as studies of chromosome 3B suggest that between 80% and 85% of the wheat genome may consist of repetitive elements (Choulet et al., 2014, 2010). Methylation-sensitive GBS has successfully been adapted for use in wheat and barley (Poland et al., 2012).

DNA markers in previous wheat studies have not been ordered physically on the chromosomes. Genetic maps have been constructed for simple-sequence repeat (SSR) markers (Somers et al., 2004) as well as increasingly dense SNP microarrays (Cavanagh et al., 2013; S. Wang et al., 2014). Thus far, the large size of the wheat genome has hindered efforts at traditional sequencing by constructing bacterial artificial chromosome (BAC) clones. To date, only chromosome 3B has been sequenced using BAC clones (Choulet et al., 2014). In addition, the high degree of repetitive elements present in the wheat genome has made it resistant towards attempts at sequencing using the short read-lengths characteristic of next-generation sequencing platforms. The development of a draft genome is critical, as it allows for in-depth automated

annotation and prediction of genomic elements (Yandell and Ence, 2012). Recent work has sought to create increasingly accurate draft genomes in wheat despite the many logistical hurdles present. Brenchley et al. (2012) published the first survey sequence of the 'Chinese Spring' bread wheat genome using next-generation sequencing technology. This work was subsequently improved upon by efforts to create a proper draft genome by sequencing each individual chromosome arm in isolation (with the exception of chromosome 3B, which was sequenced whole) (IWGSC, 2014).

Several recent studies have developed methods to create assemblies of next-generation sequencing (NGS) fragments using genetic mapping techniques (Mascher and Stein, 2014). One such technique, termed POPSEQ, was first demonstrated in barley (Mascher et al., 2013), and has since successfully been used in wheat to both anchor whole-genome assemblies (Chapman et al., 2015) and to order wheat GBS markers (Edae et al., 2015). However, a limiting factor is the recombination rate within the segregating population; rates of recombination vary substantially by chromosome position, and recombination may be nearly absent in peri-centromeric regions of the genome (Mascher and Stein, 2014). A recent improvement is the adaptation of new methods for genome-wide mapping of chromatin structure (Lieberman-Aiden et al., 2009), termed "Hi-C" mapping, to further enhance the accuracy of *de novo* assembly of large scaffolds (Putnam et al., 2016). The incorporation of Hi-C into the wheat genome assembly pipeline should bring enhanced spatial resolution beyond that of POPSEQ alone.

Ultimately, the identification of QTLs in many crop species, whether performed via linkage mapping or association studies, currently suffers from a lack of data curation. This applies to both the archiving of results , and the archiving of the raw phenotypic, genotypic, and environmental data used to generate results (Zamir, 2013). The development of a draft genome

will help to unambiguously identify various MTAs and their associated haplotypes (Welter et al., 2014).

Previous GWAS Findings

GWAS has been used extensively in human studies over several decades, and is increasing being used in plant quantitative genetics research. Despite this increasing prevalence, relatively few GWAS analyses for yield and yield-related traits have been conducted in wheat, and of these, fewer still have been conducted in winter wheat germplasm (Breseghello and Sorrells, 2006; Dodig et al., 2012; Neumann et al., 2011; Tadesse et al., 2015). None of these studies used marker data generated via GBS; many used low-density simple sequence repeat (SSR) markers, or else proprietary Diversity Arrays Technology (DArT) markers, making translation between genetic and physical chromosome positions difficult or impossible. Association studies have been more common in spring wheat, though the majority of these have been candidate-gene studies, with genome-wide studies being much more limited (Sukumaran et al., 2014). Finally, it has been more common to perform GWAS in crop species using assembled diversity panels, rather than elite germplasm in current use by breeding programs (Spindel et al., 2015). GWAS in elite germplasm is typically limited to the identification of smaller-effect QTLs, as QTLs of major effect will have likely already become fixed within the mapping population (Salvi and Tuberosa, 2015). Nevertheless, GWAS using panels of elite germplasm remain useful due to their higher relevance to the process of cultivar development (Spindel et al., 2015).

Several GWA studies have recently been carried out in wheat. One such study by Edae et al. (2014) was conducted in spring wheat in irrigated and rain-fed environments in Greeley, CO

and Melkassa, Ethiopia using a mapping panel of 285 to 294 lines (depending upon environment) and 1,863 DArT markers. They found a total of eleven significant MTAs for the traits green leaf area, leaf width, plant height, spikelets per spike, spike number per unit area, and a drought susceptibility index.

Sukumaran et al. (2014) performed GWAS using the International Maize and Wheat Improvement Center (CIMMYT) wheat association mapping initiative (WAMI) panel, consisting of 287 elite spring wheat lines. This panel was grown over four years in Ciudad Obregón, Mexico, and GWAS was performed using 18,704 SNP markers from the 90K Illumina iSelect SNP array (S. Wang et al., 2014). Sukumaran et al. found 31 significant MTAs for various yield-related traits. Most notably, they identified two pleiotropic regions including one on chromosome 6A affecting yield, thousand-kernel weight, plant height, chlorophyll index at grain fill, and canopy temperature at grain fill; and one on chromosome 5A affecting yield, thousand-kernel weight, and grain number m^{-2}.

Tadesse et al. (2015) evaluated 120 elite winter wheat genotypes for yield and grain quality traits under rain-fed and irrigated conditions over two years in Syria. These lines were genotyped with 3,051 DArT markers, of which 1,586 had known genetic positions. They found several QTLs for the traits yield, days to heading, plant height, thousand kernel weight, test weight, and grain protein content, in addition to QTLs for many grain quality traits. Notably, they did not identify any QTLs for yield *per se* that were stable across years or across rain-fed/irrigated treatments.

Ain et al. (2015) grew 123 historic Pakistani wheat cultivars in rain-fed conditions in Islamabad, Pakistan across three years. GWAS was performed using 14,960 SNPs from the 90K

iSelect array. They detected a total of 44 significant MTAs for eight different yield-related traits, including five MTAs for yield *per se*.

Finally, in a series of papers, Zanke et al. (2014; 2014; 2015) utilized GWAS to study heading date, plant height, and thousand kernel weight in a panel of 358 European winter wheat lines grown across eight environments. They performed GWAS using a set of 635 SSR markers as well as 7,769 SNP markers from the 90K iSelect array. Utilizing the best-linear unbiased estimators (BLUEs) from across all environments, for heading date they found 10 significant SSR and 51 significant SNP marker trait associations. The majority of significant SNP-MTAs were located on chromosome 5B. For plant height, they found 153 significant SSR-MTAs and 280 significant SNP-MTAs. However, combining closely-linked markers, they estimated 109 distinct loci for the SSR data, and 87 for the SNPs. Finally, for thousand-kernel weight, they found two significant SSR-MTAs (on chromosomes 4A and 6D) and seven significant SNP-MTAs (on chromosomes 3B and 5A).

Materials and Methods

Germplasm Selection

A total of 185 lines were included in each year of the study. The majority of lines changed between each year. For each year, 31 lines were sourced from Illinois, 30 from Kentucky, 2 from Missouri, and 122 from Virginia. Among the total of 329 lines, 41 were tested in both years; the others were only tested in the first or second year. Five checks were included in the study, including ‘Bess’, ‘Branson’, IL00-8250, ‘Roane’, and ‘Shirley’. All lines included in the study are listed in **Table B.1**. With the exception of checks and several older cultivars, the

majority of lines were either F_4 or higher filial generation. During processing of the genotypic data, 5 lines were removed during quality filtering, leaving 324 lines used for analysis.

Experimental Design and Field Management

Experimental plots were planted in the 2013-14 and 2014-15 winter wheat growing seasons. In Virginia, genotypes were planted in a generalized randomized complete block design (GRCBD) across two locations at Kentland Farm near Blacksburg, VA (Guernsey/Hayter silt loams, 37.1965° N, 80.5718° W, 531 m elevation) and the Eastern Virginia Agricultural Research and Extension Center (EVAREC) in Warsaw, VA (Kempsville sandy loam, 37.9879º N, 76.7770º W, 40 m elevation). Two randomized replications were planted at each location.

For the 2013-14 test at Warsaw, and the 2014-15 tests at Blacksburg and Warsaw, each experimental unit consisted of a seven-row plot with a length of 2.74 m, width of 0.91 m, row-spacing of 15.2 cm, and a harvested area of 2.49 m^2. Plots planted in Blacksburg for the 2013-14 season were smaller, with a length of 1.98 m, width of 0.91 m, row-spacing of 15.2cm, and a harvested area of 1.80 m^2. However, at both locations, plot areas were adjusted to 4.18 m^2 to account for inflated yield values caused by border effects. All plots were sown with 70 g of seed. Seed was treated with Raxil® MD fungicide (0.48% tebuconazole/0.64% metalaxyl; Bayer CropScience) at a rate of 2.95mL a.i. per kg of seed, and Gaucho® 600 flowable insecticide (48.7% imidacloprid; Bayer CropScience) at a rate of 0.7mL a.i. per kg of seed. At each location, seed was planted to roughly coincide with the average date of first frost (see **Table A.1**).

At Blacksburg and Warsaw, several tiller counts representative of the test area as a whole were used to calculate ideal nitrogen application rates at Zadok's growth stage 25 (Zadoks et al.,

1974) in the spring, and plant tissue tests were used to calculate ideal nitrogen application rates at GS30, per standard regional recommendations from the Virginia Cooperative Extension Service (Alley et al., 1993). All plots at Blacksburg and Warsaw were treated with Palisade® 2EC growth regulator (trinexapac-ethyl; Syngenta Crop Protection) to minimize lodging. In addition, plots in each environment were treated with Tilt® fungicide (propiconazole, Syngenta Crop Protection) throughout the growing season, Prosaro® fungicide (prothioconazole/tebuconazole, Bayer CropScience) near heading date, Harmony® Extra SG herbicide (thifensulfuron-methyl/tribenuron-methyl, DuPont), and Starane® Ultra broadleaf herbicide (fluroxypyr 1-methylheptyl ester , Dow AgroSciences) as needed. The exact dates and rates of chemical applications for each environment included in the study are listed in **Table A.1.**

Phenotyping

Table 1.1 lists the phenotypic traits that were assessed across all environments, with their abbreviations and units of measure. For the 2014-2015 growing season only, seedling emergence was estimated for plots in Blacksburg and Warsaw by averaging the count of seedlings at the two-leaf stage (GS12) from two 0.348 m samples taken from two inner rows. There were no significant differences in seedling numbers between lines where seed originated from VA or IL, while the number of seedlings per line was lower for KY lines due to seed source differences. Normalized Difference Vegetative Index (NDVI) was measured for each plot at GS25 as described by Phillips et al. (2004) using a Greenseeker® Handheld crop sensor (Trimble® Agriculture, Sunnyvale, CA).

Flag leaf angle was assessed visually at boot stage (GS40) and the average for each plot was recorded on a 1 to 9 scale (1 corresponding to completely erect, 9 corresponding to a 180°

curve in flag leaves). Heading date was recorded as the Julian date at which 50% of plants within a plot had extruded heads from the boot.

After plants had reached physiological maturity (GS90), a single 0.914 m cutting of all above-ground plant material was taken from one of the three inner rows of each plot and placed within a paper bag. All cuttings were stored in a sheltered environment for several days to allow for equilibration to ambient moisture levels. Each bag was weighed to derive an estimate of above-ground biomass m^{-1} of row. Subsequently, the number of heads per cutting were counted manually to derive an estimate of heads m^{-2}. Cuttings were then threshed on a plot combine (Wintersteiger NA Inc., Salt Lake City, UT) with settings optimized to recover as much threshed seed as possible. Threshed seed was weighed to derive an estimate of seed weight m^{-1} of row. Harvest index was calculated as the ratio of seed weight to total above-ground biomass. The total number of seeds threshed from each cutting were then counted on a Count-A-Pak optical seed counter (Seedburo® Equipment, Des Plaines, IL) to derive an estimate of grains m^{-2}. In addition, thousand-kernel weight was defined as the net weight of the threshed seed sample divided by the number of seeds present * 1,000.

Plant height was averaged from two measurements within each plot, and was recorded as the distance from the soil surface to the tip of the heads (excluding any awns if present). Lodging was measured on a 0 to 9 scale (0 corresponding to no lodging, 9 corresponding to complete lodging). Plots were harvested using a Wintersteiger plot combine. Moisture content and test weight (grain volume weight) of harvested grain was measured using a GAC® 2500-AGRI grain analysis computer (Dickey-John® Corporation, Auburn, IL). Grain yield was calculated at 13.5% moisture equivalence.

Grain ash, crude fiber, fat, starch, and protein were estimated via near-infrared (NIR)

spectroscopy for subsamples from each plot using an XDS Rapid Content Analyzer (FOSS NIR Systems, Laurel, MD). Fifteen grain samples from each location were sent to Cumberland Valley Analytical Services (Hagerstown, MD) for wet-chemistry analysis of protein, starch and dry matter in order to generate calibration curves for the NIR data.

Modelling of Phenotypic Response

Each location/year combination was considered as a unique environment in order to model phenotypic response across environments. For each trait, the following random effects model was fit using the *lme4* package (Bates et al., 2015) in the R statistical computing environment (R Core Team, 2015):

$$Y_{ijk} = \mu + G_i + E_j + R_k(E_j) + GE_{ij} + \varepsilon_{ijk} \quad (5)$$

Where the phenotypic response (Y_{ijk}) is a function of the overall mean (μ), the ith genotype (G_i), the kth replication (R_k) nested within the jth environment (E_j), the genotype-environment interaction (GE_{ij}) and the residual error (ε_{ijk}). For each trait, variance components for all effects were estimated, and entry-mean heritability (H^2) was calculated. In addition, genotypic best-linear unbiased predictors (BLUPs) were calculated for use as the phenotypic input for the subsequent GWAS analyses.

Genotyping

Genomic DNA was isolated from fresh green leaf tissue using a cetyltrimethylammonium bromide (CTAB) extraction protocol (Saghai-Maroof et al., 1984). Genotyping-by-sequencing was performed at USDA Agricultural Research Service (ARS) facilities using a *PstI-MseI* double digest of genomic DNA. The SNP calling was performed using TASSEL-GBS in

TASSEL 5.2.24 (Bradbury et al., 2007; Glaubitz et al., 2014). The Burrows-Wheeler aligner (Li and Durbin, 2009) was used to align SNPs to the International Wheat Genome Sequencing Consortium's whole genome assembly v0.4. In addition to the GBS genotyping described above, several major gene loci highly associated with agronomic performance were genotyped using SSR markers and LGC® KASP™ SNP genotyping assays. **Table 1.3** lists the mean allelic effects of each KASP marker on each analyzed trait.

SNP Quality Filtering and Imputation

Prior to imputation of missing genotypes, the genotypic datasets for the 2013-2014 and 2014-2015 material were jointly filtered to remove SNPs with missing data frequencies >50%, heterozygous call frequencies >15%, and minor allele frequency <5%. In addition, all unaligned SNPs were removed. After the initial filtering, missing data in the genotypic dataset was imputed using LinkImpute (Money et al., 2015). LinkImpute implements a nearest-neighbor algorithm using both the *k* nearest individuals and the *l* SNPs in highest LD with the specific missing SNP genotype that must be imputed. LinkImpute was used with its default settings, which optimize the number of nearest individuals and SNPs via data masking simulations at 10,000 randomly selected genotypes. After imputation, the dataset was once again filtered to remove SNPs with minor allele frequencies <5%. The imputed genotypic dataset was finally filtered in PLINK 1.9 (Chang et al., 2015) to remove all but one SNP in clusters separated by <64bp, as this is the tag size used in the TASSEL-GBS SNP-calling pipeline, (i.e. all SNPs located on the same tag should have the same genotype prior to imputation). In addition to the positional filtering, PLINK was used to remove all but one SNP in groups of SNPs in perfect LD ($r^2 > 0.99$) using a 200-SNP sliding window, advancing by 5 SNPs with each step.

Population Structure and Linkage Disequilibrium

Prior to performing GWAS, population structure was examined via principle component analysis (PCA) of the filtered and imputed genotypic data using the SNPRelate package (Zheng et al., 2012) in R. Linkage disequilibrium was calculated on a pairwise basis for all intra-chromosomal SNPs in the genotypic dataset, yielding 5,582,695 comparisons. The LD decay was plotted for the A, B, and D genomes separately by randomly selecting 20,000 pairwise comparisons from each genome. Then LD was plotted for each separate chromosome of the B genome using 20,000 randomly-selected pairwise comparisons from each chromosome. In addition, inter-chromosomal LD was calculated between chromosomes within each genome. Inter-chromosomal LD was also calculated between homeologous chromosomes in the A, B and D genomes. In each case, 1,000 SNPs were randomly selected to form inter-chromosomal pairs, yielding a total of 117,408 comparisons. For intra-chromosomal SNPs, r^2 values for pairwise LD comparisons were plotted against physical distance, and a second-degree locally-weighted scatterplot smoothing (LOESS) curve was fit to the data (Cleveland, 1979). For LD estimates from non-linked (i.e. inter-chromosomal) loci, the 98th percentile of the LD distribution was defined as the linkage-disequilibrium critical value. All r^2 values exceeding this value were assumed to have been caused by genetic linkage (Breseghello and Sorrells, 2006).

Genome-Wide Association Analysis

For each trait, genome-wide association analysis was performed using the Genome-Wide Complex Trait Analysis (GCTA) software (Yang et al., 2011), using a leave-one-chromosome-out (LOCO) method in which a separate genetic relationship matrix (GRM) is estimated from

SNP data for each chromosome. Specifically, the LOCO approach entails excluding all SNPs located on the chromosome of the SNP undergoing testing when estimating the GRM. For each trait, permutation testing was performed to empirically determine a significance threshold by randomly shuffling phenotypic data and included principle components in unison, performing the GCTA-LOCO analysis on the randomly-reordered data, recording the lowest observed p-value, and repeating this process 1,000 times. The GCTA model failed to converge for the traits heading date and thousand-kernel weight and, therefore, GWAS for these traits was performed using the rrBLUP package (Endelman, 2011) in R.

In addition, GWAS was run on each trait using the Fixed and random model Circulating Probability Unification (FarmCPU) model (Liu et al., 2016) in R. Once again, permutation testing was performed for 1,000 iterations for each trait. In addition, to enhance our confidence in QTLs with p-values exceeding the significance threshold in FarmCPU, we implemented a bootstrapping method developed by Wallace, et al. (2016), in which 10% of the phenotypic observations were replaced with missing data for a total of 100 runs of the model. Subsequently, for each trait the resample model inclusion probability (Valdar et al., 2009) was calculated for each SNP by determining the fraction of bootstraps in which its p-value exceeded the permutation-derived significance threshold. The value 0.05 was chosen as a lower threshold for the RMIP as it coincided with the point of inflection in the RMIP density curve (data not shown).

For each model, the first four principal components were included to model population structure, based upon visual examination of the Scree plot for variance explained by each PC, and clustering of lines shown in biplots of the first few PCs. However, for the traits HI and HT in the FarmCPU model, and GSQM in the GCTA model, a strong signal identified when including

the first three PCs was seemingly nullified when including the fourth PC. Therefore, these traits were run including the first three PCs only.

SNP Translation Effect Prediction and Gene Annotation

As the latest WGAv04 genome assembly is not yet annotated, we were unable to use WGAv04 SNP positions to directly ascertain genomic features near to or overlapping SNPs involved in significant SNP-trait associations. Therefore, 100bp of flanking sequence on either side of each significant SNP identified in the GWAS was aligned against the previous wheat genome assembly (TGACv1, generated by the Earlham Institute, formerly the Centre for Genome Analysis [TGAC]; (Clavijo et al., 2016)). The TGACv1 assembly consists of unordered scaffolds sorted by chromosome arm, with a scaffold N50 of 88.8Kb. The Ensembl Variant Effect Predictor (McLaren et al., 2010) was used to classify SNPs as being either intergenic, intronic, exonic, or else residing in 3' or 5' untranslated regions. In addition, the predicted effects of exonic SNPs on protein translation were classified as synonymous, missense, or nonsense. For each intergenic SNP, distance was calculated between the SNP position and the closest end position of the nearest gene, if at least one gene was located on the same scaffold as the SNP. Peptide sequences for the transcripts of all genes contained within the set of scaffolds containing significant SNPs were used to perform a protein-protein BLAST search (Altschul et al., 1997, 1990; Camacho et al., 2009) against a database containing all plant proteins downloaded from Ensembl Plants v33, filtered to remove proteins of unknown or putative functions.

Results

General Line Performance and Trait Correlations

The variance components and entry-mean heritability across all four environments included in the study are shown in **Table 1.2**. Significant differences among genotypes were recorded for all traits except for canopy temperature depression (CTD) at anthesis, CTD during grain fill, and flag leaf angle at boot stage. Therefore, the latter mentioned traits were not assessed in the second year of the study. In addition, lodging was not extensive enough within all locations and years to enable reliable estimation of its effects across environments, so this trait was excluded from further analysis. Trait heritability ranged from 0.41 for BIOM to 0.94 for HD and TKW. Heritability for YLD was slightly higher than expected at 0.83. A naïve estimation of allele effects for the KASP markers assaying genes of known function revealed that the stem rust (*Puccinia graminis*) resistance gene *Sr36* and the sucrose-synthase gene *TaSus2-2B* produced many significant differences among genotypes for multiple traits (**Table 1.3**). Gene *Sr36* is located on a 2G:2B alien translocation originating from *Triticum timopheevi* (A^mA^mGG) (Brown-Guedira et al., 1996; Nyquist, 1962). Gene *TaSus2-2B* is located on the short arm of chromosome 2B, and is one of the three sucrose synthase *Sus2* orthologs, located on chromosomes 2A, 2B, and 2D (Jiang et al., 2011). Two common haplotypes for *TaSus2-2B* include *Hap-H* (high seed weight) and *Hap-L* (low seed weight). Cabrera et al. (2015) found that the simple sequence repeat (SSR) marker *Xwmc477* used to test for the 2G:2B translocation was in perfect LD with *TaSus2-2B*, suggesting that the haplotypes of this gene can be used to diagnostically test for the presence or absence of the translocation. Interestingly, the presence of the *Rht-B1b* and *Rht-D1b* dwarfing alleles produced significant effects of opposite signs for

many traits, including BIOM, SSQM, TKW, and TWT. Effects were also of opposite signs for HT and YLD, though in both of these cases the effects of *Rht-B1b* were not significant.

Figure 1.1 depicts the pairwise Pearson correlation coefficients among all traits included in the study. As expected, phenological traits (e.g. HD, FLS, MAT) and traits relating to the intervals between phenological stages (e.g. HP, FLSG) demonstrated a high degree of positive correlation. Conversely, the traits GSQM and TKW demonstrated a strong negative correlation. Yield was most highly correlated with the traits FLS, MAT, and GW (positive), and grain protein (negative).

Population Structure and Linkage Disequilibrium

Principle component analysis of the processed GBS-SNP data revealed that population structure was not as stratifying as expected, with the first principle component explaining ~7% of the total variance. No distinct point of inflection was observed in either the plot for variance explained by each principle component (**Figure 1.2a**), or the cumulative variance explained by principle component (**Figure 1.2b**). Lines included in the panel formed two distinct clusters in the biplot of the first two principle components. Further examination revealed that these two clusters were largely delineated by the presence or absence of *Sr36* and *TaSus2-2B* as determined by KASP assay (**Figure 1.3**; similar results for *TaSus2-2B* not shown). Interestingly, neither the 1B:1R nor the 1A:1R alien translocations from rye (*Secale cereale*) produced any discernable clustering of genotypes (data not shown).

Linkage disequilibrium decay plots demonstrated significant LD extending out to large physical distances. Notably, intra-chromosomal LD in the B genome extended much farther than in either the A or D genomes (**Figure 1.4**). Within the B genome, chromosome 2B displayed the

most extensive LD, chromosomes 4B and 5B displayed intermediate LD, and chromosomes 1B, 3B, 6B, and 7B displayed low LD more similar to the overall LD patterns observed in the A and D genomes (**Figure 1.5**). A plot of genotypic LD between the KASP *Sr36* marker and all other SNPs located on chromosome 2B revealed extensive LD extending across the entire chromosome (**Figure 1.6**). Plotting distance between adjacent SNPs vs. LD revealed that LD between adjacent SNPs was largely a function of SNP density. However, there were exceptions, with several high-LD, low SNP-density regions occurring on chromosomes 4A, 6A, 7A, and 1D (**Figure 1.7**). There do not appear to be significant differences in LD between inter-chromosomal SNPs for homeologous versus non-homeologous chromosomes.

Genome-Wide Association Studies

In general, GCTA identified far fewer significant MTAs than FarmCPU. In total, GCTA identified 30 significant MTAs at the empirically-determined 0.05 significance threshold. However, many of these were located within high-LD blocks and, therefore, clustered at the same putative QTLs. Discounting those SNPs believed to co-localize to identical QTLs yielded seven unique trait-QTL associations for FLS, FLSG, GSQM, MAT, SSQM, and TWT (**Table 1.4**). In contrast, FarmCPU identified 68 MTAs at the same significance threshold. Two of these MTAs involved *Ppd* alleles. Removing those SNPs with RMIPs below 0.05 and the two *Ppd* alleles yielded a total of 49 significant MTAs for the traits FLS, FLSG, GSQM, HD, HI, HP, HT, MAT, SSQM, STARCH, TKW, TWT, and YLD (**Table 1.5**). Several SNPs affected multiple traits, and hence the 49 significant MTAs involved 46 unique SNPs. As these SNPs were fitted as covariates by the FarmCPU model based partially upon LD analysis, it is assumed that the MTAs identified for each trait represent separate QTLs. SNPs with low RMIP values were tested

for genotypic LD with other SNPs identified as significant, but no significant SNP with a RMIP value below the 0.05 threshold was found to be in high LD with any SNP above the threshold.

SNP Translation Effect Prediction and Gene Identification

The GCTA and FarmCPU analyses identified a total of 49 unique SNPs involved in marker-trait associations. All of these SNPs were located on separate scaffolds in the TGACv1 assembly. Of the 49 total SNPs, 35 (71%) were intergenic, and of these intergenic SNPs, 10 were labeled as upstream or downstream proximal variants (i.e. within 5Kb of the start or end of a gene). Of the 14 SNPs located within genes, 3 occurred within 3' untranslated regions, 5 were intronic, and 4 were exonic. In addition, one SNP (S4A_739598141) was putatively located within two overlapping genes on opposite strands. However, in both genes this SNP had no predicted effect on the translated peptide (i.e. the SNP was intronic in one gene and led to a synonymous substitution in the other gene). In addition, at this time the presence of a true overlap of these two genes cannot be validated. Of the 5 exonic SNPs (including S4A_739598141), 3 were missense variants, 1 was the aforementioned synonymous variant due to S4A_739598141, and 1 created a premature stop codon. For intergenic SNPs, the median distance to the closest end of the closest gene was 26.3Kb. Predicted translational effects of SNPs and distances to closest genes are given in **Table C.1**.

A total of 127 genes producing 157 unique transcripts were located on the scaffolds containing the significant MTAs generated by both the GCTA and FarmCPU models. However, genes were not uniformly distributed across scaffolds containing significant MTAs. Only 36 of the 49 total SNPs involved in significant MTAs were located on scaffolds containing one or more genes. Of the 127 total genes, 5 produced noncoding RNA transcripts. **Table D.1** lists the

results of a protein-protein BLAST search of all transcripts located on the set of scaffolds containing the significant MTAs from both models against all proteins of known function in Ensembl Plants v33. The results are filtered to include only the single best match for each transcript.

Discussion

Within the testing panel used for this study, the 2G:2B translocation from *T. timopheevi* appears to be the driving force behind the last vestiges of population structure, as well as the more prominent LD patterns observed in the B genome and on chromosome 2B specifically (**Figures 1.4 and 1.5**). Measuring the genotypic r^2 values between *Sr36* and every other SNP present on chromosome 2B revealed that the 2G:2B translocation involved nearly all of chromosome 2B, and that it introduced extensive LD across the entire chromosome that is still in the process of decaying (**Figure 1.6**). The findings regarding LD in the B genome are in contrast to the findings of Chao et al. (2010), who reported lower levels of LD in the B genome of both winter and spring wheat. However, it is not known how frequently the 2G:2B translocation occurred in the lines they tested. It is also notable that the 1B:1R translocation did not produce similarly high LD in chromosome 1B. In fact, chromosomes 1B, 3B, and 6B all exhibited LD decay patterns that were similar to the patterns of the pooled A and D genome chromosomes. It is likely that LD due to the 1B:1R and 1A:1R translocations has decayed to a greater extent than that due to the 2G:2R translocation in the germplasm used in this study.

The presence of several high-LD yet low SNP-density regions on chromosomes 4A, 6A, 7A and 1D (**Figure 1.7**) suggests the presence of selection sweep due to linkage drag with a highly-important locus, or else a possible recent translocation event. None of the significant

MTAs identified within this study fell within any of these low SNP-density/high-LD regions. If these regions represent haplotype blocks in linkage with a highly important allele, it is possible that the causative allele may be fixed within the population.

Population structure was generally negligible within the testing panel. This finding is in line with those of previous studies examining population structure in elite European winter wheat germplasm (Reif et al., 2011; Würschum et al., 2013). This suggests extensive past admixture among the lines included in the population, which is as expected given the frequent germplasm exchanges that are typical of public small grains breeding programs. In the testing panel used in this study, the *T. timopheevi* 2G:2B translocation explained some stratification between lines, though the first genotypic principle component still only explained ~7% of total variation, suggesting extensive past admixture among lines included in the panel. In contrast, Sukumaran et al. (2014) found that a panel of elite spring germplasm clustered into two distinct sub-populations explained by the presence or absence of the rye 1B:1R translocation.

In addition, Sukumaran et al. (2014) found that the 1B:1R translocation explained significant differences among the two subpopulations for most of the traits included in the study (e.g. grain yield, grain number, grain weight, plant height, and several phenological traits). The 1A:1R and 1B:1R translocations have been associated with desirable disease and insect resistance traits, as well as drought and general environmental stress resistance. However, the 1B:1R translocation has been associated with lateness, and effects on yield appear to due to these translocations appear to be more variable, depending upon environment and genetic background (Ehdaie et al., 2003; Moreno-Sevilla et al., 1995; Singh et al., 1998). In the present study, many of the previously-characterized loci of agronomic importance interrogated with KASP-SNP assays had significant allelic effects for multiple traits (**Table 1.3**). Loci that significantly

affected many traits included the 1A:1R and 1B:1R translocations, the *Ppd-B1* and *Ppd-D1* photosensitivity genes, the *Rht-B1* and *Rht-D1* dwarfing genes, the *Vrn-A1* vernalization gene, as well as the *Sr36* stem rust resistance gene and *TaSus2-2B* sucrose synthase gene (both thought in this case to indicate the presence/absence of the 2G:2B *T. timopheevi* translocation). However, despite their significant effects on many traits, these loci were generally not among the significant MTAs identified in the current study, the exceptions being *Ppd-B1* and *Ppd-D1*, which were identified as being significantly associated with FLSG and FLS, respectively, by the FarmCPU model. In addition, one significant GWAS hit, SNP S1B_39294256 affecting the trait STARCH, appears to have been introduced on the 1B:1R translocation, as it was in high LD (r^2 = 0.87) with the KASP marker for this translocation. This would suggest that most loci interrogated with KASP-SNP assays were sufficiently confounded with population structure and/or kinship so as to be effectively nullified in the GWA analyses.

A note on the two *Rht* dwarfing alleles included in the KASP assays is warranted here, as these two alleles (*Rht-B1b* and *Rht-D1b*) occurred mostly in repulsion within the test population. Of the 324 lines included in the panel, 22 had neither the *Rht-B1b* nor the *Rht-D1b* dwarfing alleles, one line was heterozygous for both alleles, and one line was homozygous for both alleles. At face value, the high degree of repulsion between *Rht-B1* and *Rht-D1* led to the somewhat odd finding that the presence of the *Rht-B1b* allele increased height and decreased yield (albeit non-significantly) when allelic effects were calculated for each of the KASP assays individually (**Table 1.3**). Subsequent ANOVAs revealed that lines with either *Rht-B1b* or *Rht-D1b* were significantly shorter and yielded significantly higher than lines that were wild-type for both genes at an alpha level of 0.05. Yield was significantly higher for lines with only *Rht-D1b* vs. those with only *Rht-B1b*, though height was not significantly different between these two groups.

(These analyses excluded lines that were heterozygous for either allele, and the single line that was homozygous for both alleles). This perhaps serves as a good cautionary tale when interpreting single-locus tests such as classical GWAS methods, as inter-locus interactions such as linkage, LD, and epistasis can significantly alter the phenotypic expression of a single locus' effects. Multi-locus methods such as FarmCPU are expected to at least partially ameliorate the effects of linkage and LD, though accounting for epistasis remains a computationally daunting task.

As previously mentioned, the FarmCPU model identified many more significant MTAs than the single-locus model implemented in GCTA. All but one of the seven significant associations identified by the GCTA model were also identified by FarmCPU. Cases in which a QTL was identified by the same SNP in both models include S5B_396479359 for the trait GSQM and S1A_583587147 for the trait TWT. For other traits, the same QTLs were identified by both models, but via different SNPs within the same high-LD blocks. These include a QTL for FLSG at ~64Mb on chromosome 7B, a QTL for SSQM at ~3Mb on chromosome 3A, and a pleiotropic QTL affecting FLS and MAT at ~58.5Mb on chromosome 7D. The QTL on chromosome 7D was also significant for the trait HD in the FarmCPU results, though this was not the case for the GCTA results. The only QTL identified via the GCTA model which was not identified by FarmCPU was labeled by SNP S6B_181128808 for the trait FLS. The fact that 17 out of a total of 68 significant FarmCPU results were removed after bootstrapping due to having RMIP values below 0.05 suggests that caution is warranted when interpreting FarmCPU results. Bootstrapping was easy to implement and ran reasonably fast on modest hardware for the genotypic datasets used herein.

There was evidence for pleiotropic effects relating to a number of different SNPs and traits in both the GCTA and FarmCPU results (**Tables 4 and 5**). The sole region of pleiotropic effect identified by GCTA was a QTL located at ~58.5Mb on chromosome 7D. This QTL was also identified by FarmCPU. For both models the SNP S7D_58449294 within this QTL was most highly associated with the trait FLS, while the SNP S7D_58589271 was most highly associated with MAT. This SNP was also associated with HD in the results of the FarmCPU model, though this was not the case for the GCTA results. Several other pleiotropic QTLs were identified by FarmCPU but not GCTA. These included the SNP S2D_35084672, significantly associated with the traits HD and MAT, with very similar effect sizes of -0.55 and -0.52 days for each respectively. Caution is warranted regarding both this QTL, and the one affecting phenological traits on chromosome 7D. In both cases, these QTLs had similar effects on both HD, as well as traits relating to late-season maturity (FLS, MAT). Therefore, it is likely that both of these QTLs have little if any effect on grain fill duration. The SNP S4B_626390000 was significantly associated with the traits HI and STARCH. However, this SNP had a low effect size for both of these traits (-0.01% and -0.23% respectively). In addition to the QTL located on 7D, there were several more examples of SNPs in close proximity to each other affecting different traits. The SNP marker S5B_394707451 was significantly associated with HI (though with a negligible estimated effect size of -0.004), while S5B_396479359 was associated with GSQM (with an effect size of -319 grains m^{-2}). The genotypic r^2 value between these two SNPs was 0.94, strongly suggesting that they both reside within the same QTL. The SNP marker S7A_673387152 was associated with GSQM, while S7A_673436887 was associated with TKW. Predictably, the effect sizes for this QTL were of opposite signs for the two traits (+302 grains

sqm^{-2} for GSQM, and -0.553 grams for TKW), reflecting the negative correlation between these two traits.

While the majority of significant SNP-MTAs identified by both models resided in intergenic spaces, several SNP-MTAs were located within genes, and 5 were exonic. The most severe predicted translational consequence for any of the SNP-MTAs identified by either model was a premature stop codon formation caused by SNP S5B_34721398, which affects whole-grain starch content. Interestingly, this SNP resides within a gene (TRIAE_CS42_5BS_TGACv1_424234_AA1387820) that may be an ortholog of the disease-resistance protein *RPM1*. *RPM1* is a classical peripheral plasma-membrane disease resistance protein displaying nucleotide-binding site (NBS) and leucine-rich repeat (LRR) motifs in *Arabidopsis thaliana* (Boyes et al., 1998). The SNP S4A_726716318, affecting yield *per se*, was intronic within TRIAE_CS42_4AL_TGACv1_288738_AA0956970, a potential ortholog of *RPS2*, which is likewise a NBS-LRR resistance protein in *A. thaliana* (Bent et al., 1994). In *A. thaliana*, both *RPM1* and *RPS2* detect disturbance of the *RIN4* protein by the *Pseudomonas syringae* effectors *avrRpm1* and *avrB*, initiating a cell hypersensitive defense response (Kim et al., 2005), though there is no evidence to suggest that the possible orthologs of these proteins in wheat target *P. syringae* effectors specifically. The significance of the association of these disease-resistance proteins with the traits grain starch content and grain yield, respectively, is as yet unknown. However, at least in the case of the SNP affecting yield on chromosome 4A, one potential explanation may be the fitness cost that the expression of many resistance genes entails (Tian et al., 2003).

Several other SNPs led to predicted missense amino acid substitutions. These included S3B_695966897 (affecting test weight), located within a gene with a conserved domain similar

to the human signal peptide pedtidase (*SPPL2B*), which is an aspartic protease (Weihofen et al., 2002). Numerous aspartic proteases have been identified in plants, and are known to be involved in senescence mechanics (Roberts et al., 2012). At least one study has found an upregulation of an aspartic protease in wheat flag leaves undergoing senescence (Gregersen and Holm, 2007), while an *in vitro* study in tobacco found that the *CND41* protease affected senescence by degrading denatured Rubisco (Kato et al., 2004). The SNP S6D_127384672 affected heading date, and was located within a putative heavy metal ATPase. Heavy metal ATPases are known to be involved in zinc and cadmium transport and detoxification in rice (Takahashi et al., 2012). Finally, SNP S7D_58449294 affected flag leaf senescence, and was located within a protein containing a pentatricopeptide repeat (PPR). The exact biological role of PPR proteins remains unknown, though numerous experiments in *A. thaliana*, rice, and maize suggest that they play an important role in post-transcriptional processing within chloroplasts and mitochondria (Schmitz-Linneweber and Small, 2008). This suggests an enticing hypothesis linking this particular PPR protein with senescence mechanics within plastids, though this association remains conjectural for the time being.

The data and results demonstrated that physical proximity is not always a reliable predictor of LD. This was evident for the SNPs S6A_614373502 (associated with the trait SSQM), and S6A_614660970 (associated with TWT). The physical distance between these two SNPs was ~287 Kb, which was within the size range of two other pleiotropic QTLs revealed in this study. However, the genotypic r^2 value between these two SNPs was 0.16. While this is above the LD threshold empirically derived from non-linked loci ($r^2 = 0.06$), it is likely still too low to confidently declare that these loci inhabit the same QTL.

It is somewhat surprising that only a single identified QTL affected both FLS and MAT, as these traits are among the most highly correlated of all that were included in the study. It is especially surprising that while the SNP marker S2D_35084672 was associated with HD and MAT in the FarmCPU results, it was not associated with FLS, as this trait is likewise related to physiological development timing, is of similar heritability to HD and MAT, and has a timing of data collection falling between the two.

Collectively, the results identify some ancillary traits that are likely to be useful for increasing grain yield. The two phenological traits FLS and MAT, both relating to timing of senescence, had both high positive correlation with yield and higher heritability than yield, suggesting that selection for stay-green characteristics could be a viable and easily-implemented strategy to increase yield. Previous literature had concluded that the trait GSQM was critical for increasing yield (Reynolds et al., 2005), and in the present study this trait demonstrated high positive correlation with yield, as well as heritability roughly equal to that of yield (0.81 for GSQM vs. 0.83 for YLD). However, dissecting this trait into two further component traits, SPH and SSQM, demonstrated that the trait SPH was the main factor driving the relevance of GSQM in determining yield, with both highly significant correlation with yield and high heritability. Conversely, SSQM was not significantly correlated with yield. However surprisingly, while this study did reveal significant MTAs for GSQM and SSQM, this was not the case for SPH. In this study, grain protein content (WCPROT) was negatively correlated with yield, a finding that has been repeated across many studies (Cox et al., 1985; Groos et al., 2003; Terman et al., 1969). Several results from the current study indicated compensation between traits, suggesting that indirect selection for yield based upon ancillary traits could be difficult. For instance, the traits GSQM, SPH, and SSQM were all negatively correlated with TKW, suggesting that efforts to

select for higher grain number could be nullified by smaller average grain size. In addition, SPH and SSQM were negatively correlated, suggesting a tradeoff between head size and number of heads per plant. Slafer et al. (1996) suggested that there is some peril in blindly selecting for increased grains per unit area, as corresponding decreases in average grain weight will tend to negate any possible yield gains. They also suggested that this effect could potentially be mitigated by coupling selection for grains per unit area with selection for phenological characteristics to increase grain filling duration, in particular photoperiod and temperature response during stem elongation, and temperature response during grain fill. While the present study did not directly measure these traits, the identification of numerous QTLs affecting both GSQM and various traits relating to stay-green characteristics suggests promising avenues for further research.

The distribution of harvest index values in the testing panel (**Table 1.2**) suggests a simpler path to increased yields; the mean harvest index across all environments for the tested germplasm was 0.43, while the maximum value was 0.53. Both of these values are well below the theoretical peak harvest index value in wheat of 0.6 proposed by Austin et al. (1980). The FarmCPU results (**Table 1.5**) demonstrate the presence of 3 QTLs affecting harvest index, located on chromosomes 4B and 5B. In particular, one of the QTLs on 5B is pleiotropic, also affecting the trait GSQM, and the effects of this QTL on each trait are of the same sign. While harvest index is a difficult trait to select for directly, measurement of this trait on select breeding material could prove useful in increasing yield.

The current lack of a genetic map for the markers used in this study makes comparison to the results of previous studies difficult. No previous GWAS study in wheat has had a whole-genome assembly available to physically anchor markers. Results were collected from several

recent GWAS studies focusing on yield-related traits in wheat (Ain et al., 2015; Sukumaran et al., 2014; C. Zanke et al., 2014; C. D. Zanke et al., 2014; Zanke et al., 2015), focusing on markers in which a significant MTA for the same trait was found on the same chromosome as the study herein. Several recent studies (Edae et al., 2014; Tadesse et al., 2015) were excluded due to the use of proprietary DArT markers. The flanking sequences for a total of 26 markers from previous studies were aligned against both the TGACv1 and WGAv04 assemblies; however, none of these markers co-localized to the same scaffold as the significant markers from this study in either assembly. Nevertheless, we cannot exclude the possibility that significant MTAs identified in this study are in high LD with significant MTAs identified in previous studies, despite residing on separate scaffolds in the TGACv1 assembly, as this study included several examples of SNPs in high LD with each other residing on separate scaffolds. Hopefully, the release of the forthcoming whole-genome assembly will help to resolve this matter.

Conclusions

The significant MTAs reported in this study indicate that there is still some degree of genetic variation in the tested elite germplasm that may be exploited for yield gains. In particular, the combination of identified QTLs affecting traits relating to grains per unit area and phenological development seems to offer promise for increasing the former while avoiding the penalizing effect of lower average grain weights, as suggested in previous literature (Slafer et al., 1996). The findings of this study also indicate, at least for the soft winter wheat germplasm panel used, that increasing harvest index remains a viable approach to increasing yield. However, this raises some important technical and logistical considerations – direct selection for many of the traits included in this study, including harvest index, remains laborious. Therefore, breeders must

exercise prudent judgement in selecting which material to collect data on. Breeding schemes utilizing genomic selection may help in this regard, by limiting data collection on certain traits to a model training population. This study also identifies some potential targets for future *in vitro* studies to ascertain the biological functions of several genes in wheat.

References

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

Alley, M.M., Brann, D.E., Stromberg, E.L., Hagood, E.S., Herbert, A., Jones, E.C., Griffith, W.K., 1993. Intensive Soft Red Winter Wheat Production: A Management Guide.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410. doi:10.1016/S0022-2836(05)80360-2

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389–3402. doi:10.1093/nar/25.17.3389

Atwell, S., Huang, Y.S., Vilhjálmsson, B.J., Willems, G., Horton, M., Li, Y., Meng, D., Platt, A., Tarone, A.M., Hu, T.T., Jiang, R., Muliyati, N.W., Zhang, X., Amer, M.A., Baxter, I., Brachi, B., Chory, J., Dean, C., Debieu, M., de Meaux, J., Ecker, J.R., Faure, N., Kniskern, J.M., Jones, J.D.G., Michael, T., Nemri, A., Roux, F., Salt, D.E., Tang, C., Todesco, M., Traw, M.B., Weigel, D., Marjoram, P., Borevitz, J.O., Bergelson, J., Nordborg, M., 2010. Genome-wide association study of 107 phenotypes in Arabidopsis thaliana inbred lines. Nature 465, 627–631. doi:10.1038/nature08800

Austin, R.B., 1980. Physiological limitations to cereal yields and ways of reducing them by breeding, in: Hurd, R.G., Biscoe, P.V., Dennis, C., Association of Applied Biologists (Eds.), Opportunities for Increasing Crop Yields. Pitman, Boston, pp. 3–19.

Bates, D., Mächler, M., Bolker, B., Walker, S., 2015. Fitting Linear Mixed-Effects Models Using lme4. J. Stat. Softw. 67. doi:10.18637/jss.v067.i01

Bent, A.F., Kunkel, B.N., Dahlbeck, D., Brown, K.L., Schmidt, R., Giraudat, J., Leung, J., Staskawicz, B.J., 1994. RPS2 of Arabidopsis thaliana: a leucine-rich repeat class of plant disease resistance genes. Science 265, 1856–1860. doi:10.1126/science.8091210

Borrás, L., Slafer, G.A., Otegui, M.E., 2004. Seed dry weight response to source–sink manipulations in wheat, maize and soybean: a quantitative reappraisal. Field Crops Res. 86, 131–146. doi:10.1016/j.fcr.2003.08.002

Boyes, D.C., Nam, J., Dangl, J.L., 1998. The Arabidopsis thaliana RPM1 disease resistance gene product is a peripheral plasma membrane protein that is degraded coincident with the hypersensitive response. Proc. Natl. Acad. Sci. 95, 15849–15854. doi:10.1073/pnas.95.26.15849

Bradbury, P.J., Zhang, Z., Kroon, D.E., Casstevens, T.M., Ramdoss, Y., Buckler, E.S., 2007. TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23, 2633–2635. doi:10.1093/bioinformatics/btm308

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

Breseghello, F., Sorrells, M.E., 2006. Association Mapping of Kernel Size and Milling Quality in Wheat (Triticum aestivum L.) Cultivars. Genetics 172, 1165–1177. doi:10.1534/genetics.105.044586

Brown-Guedira, G.L., Badaeva, E.D., Gill, B.S., Cox, T.S., 1996. Chromosome substitutions of Triticum timopheevii in common wheat and some observations on the evolution of polyploid wheat species. Theor. Appl. Genet. 93, 1291–1298. doi:10.1007/BF00223462

Bruinsma, J., 2011. The resources outlook: by how much do land, water and crop yields need to increase by 2050?, in: Looking Ahead in World Food and Agriculture: Perspectives to 2050. Food and Agriculture Organization of the United Nations, Rome, Italy.

Cabrera, A., Guttieri, M., Smith, N., Souza, E., Sturbaum, A., Hua, D., Griffey, C., Barnett, M., Murphy, P., Ohm, H., Uphaus, J., Sorrells, M., Heffner, E., Brown-Guedira, G., Sanford, D.V., Sneller, C., 2015. Identification of milling and baking quality QTL in multiple soft wheat mapping populations. Theor. Appl. Genet. 128, 2227–2242. doi:10.1007/s00122-015-2580-3

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T.L., 2009. BLAST+: architecture and applications. BMC Bioinformatics 10, 421. doi:10.1186/1471-2105-10-421

Cavanagh, C.R., Chao, S., Wang, S., Huang, B.E., Stephen, S., Kiani, S., Forrest, K., Saintenac, C., Brown-Guedira, G.L., Akhunova, A., See, D., Bai, G., Pumphrey, M., Tomar, L., Wong, D., Kong, S., Reynolds, M., da Silva, M.L., Bockelman, H., Talbert, L., Anderson, J.A., Dreisigacker, S., Baenziger, S., Carter, A., Korzun, V., Morrell, P.L., Dubcovsky, J., Morell, M.K., Sorrells, M.E., Hayden, M.J., Akhunov, E., 2013. Genome-wide comparative diversity uncovers multiple targets of selection for improvement in hexaploid wheat landraces and cultivars. Proc. Natl. Acad. Sci. U. S. A. 110, 8057–8062. doi:10.1073/pnas.1217133110

Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., Lee, J.J., 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience 4, 7. doi:10.1186/s13742-015-0047-8

Chao, S., Dubcovsky, J., Dvorak, J., Luo, M.-C., Baenziger, S.P., Matnyazov, R., Clark, D.R., Talbert, L.E., Anderson, J.A., Dreisigacker, S., Glover, K., Chen, J., Campbell, K., Bruckner, P.L., Rudd, J.C., Haley, S., Carver, B.F., Perry, S., Sorrells, M.E., Akhunov, E.D., 2010. Population- and genome-specific patterns of linkage disequilibrium and SNP variation in spring and winter wheat (Triticum aestivum L.). BMC Genomics 11, 727. doi:10.1186/1471-2164-11-727

Chapman, J.A., Mascher, M., Buluç, A., Barry, K., Georganas, E., Session, A., Strnadova, V., Jenkins, J., Sehgal, S., Oliker, L., Schmutz, J., Yelick, K.A., Scholz, U., Waugh, R., Poland, J.A., Muehlbauer, G.J., Stein, N., Rokhsar, D.S., 2015. A whole-genome shotgun approach for assembling and anchoring the hexaploid bread wheat genome. Genome Biol. 16, 26. doi:10.1186/s13059-015-0582-8

Choulet, F., Alberti, A., Theil, S., Glover, N., Barbe, V., Daron, J., Pingault, L., Sourdille, P., Couloux, A., Paux, E., Leroy, P., Mangenot, S., Guilhot, N., Le Gouis, J., Balfourier, F., Alaux, M., Jamilloux, V., Poulain, J., Durand, C., Bellec, A., Gaspin, C., Safar, J., Dolezel, J., Rogers, J., Vandepoele, K., Aury, J.-M., Mayer, K., Berges, H., Quesneville, H., Wincker, P., Feuillet, C., 2014. Structural and functional partitioning of bread wheat chromosome 3B. Science 345, 1249721. doi:10.1126/science.1249721

Choulet, F., Wicker, T., Rustenholz, C., Paux, E., Salse, J., Leroy, P., Schlub, S., Le Paslier, M.-C., Magdelenat, G., Gonthier, C., Couloux, A., Budak, H., Breen, J., Pumphrey, M., Liu, S., Kong, X., Jia, J., Gut, M., Brunel, D., Anderson, J.A., Gill, B.S., Appels, R., Keller, B., Feuillet, C., 2010. Megabase level sequencing reveals contrasted organization and evolution patterns of the wheat gene and transposable element spaces. Plant Cell 22, 1686–1701. doi:10.1105/tpc.110.074187

Clavijo, B.J., Venturini, L., Schudoma, C., Accinelli, G.G., Kaithakottil, G., Wright, J., Borrill, P., Kettleborough, G., Heavens, D., Chapman, H., Lipcombe, J., Barker, T., Lu, F.-H., McKenzie, N., Raats, D., Gonzalez, R.H.R., Coince, A., Peel, N., Percival-Alwyn, L., Duncan, O., Trösch, J., Yu, G., Bolser, D., Naamati, G., Kerhornou, A., Spannagl, M., Gundlach, H., Haberer, G., Davey, R.P., Fosker, C., Palma, F.D., Phillips, A., Millar, A.H., Kersey, P.J., Uauy, C., Krasileva, K.V., Swarbreck, D., Bevan, M.W., Clark, M.D., 2016. An improved assembly and annotation of the allohexaploid wheat genome identifies complete families of agronomic genes and provides genomic evidence for chromosomal translocations. bioRxiv 080796. doi:10.1101/080796

Cleveland, W.S., 1979. Robust Locally Weighted Regression and Smoothing Scatterplots. J. Am. Stat. Assoc. 74, 829–836. doi:10.1080/01621459.1979.10481038

Cox, M.C., Qualset, C.O., Rains, D.W., 1985. Genetic Variation for Nitrogen Assimilation and Translocation in Wheat. I. Dry Matter and Nitrogen Accumulation. Crop Sci. 25, 430–435. doi:10.2135/cropsci1985.0011183X002500030002x

Dodig, D., Zoric, M., Kobiljski, B., Savic, J., Kandic, V., Quarrie, S., Barnes, J., 2012. Genetic and Association Mapping Study of Wheat Agronomic Traits Under Contrasting Water Regimes. Int. J. Mol. Sci. 13, 6167–6188. doi:10.3390/ijms13056167

Donald, C.M., 1968. The breeding of crop ideotypes. Euphytica 17, 385–403. doi:10.1007/BF00056241

Donald, C.M., 1962. In Search of Yield. J. Aust. Inst. Agric. Sci. 28, 171–178.

Edae, E.A., Bowden, R.L., Poland, J., 2015. Application of Population Sequencing (POPSEQ) for Ordering and Imputing Genotyping-by-Sequencing Markers in Hexaploid Wheat. G3 GenesGenomesGenetics 5, 2547–2553. doi:10.1534/g3.115.020362

Edae, E.A., Byrne, P.F., Haley, S.D., Lopes, M.S., Reynolds, M.P., 2014. Genome-wide association mapping of yield and yield components of spring wheat under contrasting moisture regimes. Theor. Appl. Genet. 127, 791–807. doi:10.1007/s00122-013-2257-8

Ehdaie, B., Whitkus, R.W., Waines, J.G., 2003. Root Biomass, Water-Use Efficiency, and Performance of Wheat–Rye Translocations of Chromosomes 1 and 2 in Spring Bread Wheat "Pavon." Crop Sci. 43, 710–717. doi:10.2135/cropsci2003.7100

Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J.A., Kawamoto, K., Buckler, E.S., Mitchell, S.E., 2011. A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. PLoS ONE 6, e19379. doi:10.1371/journal.pone.0019379

Endelman, J.B., 2011. Ridge Regression and Other Kernels for Genomic Selection with R Package rrBLUP. Plant Gen 4, 250–255. doi:10.3835/plantgenome2011.08.0024

Food and Agriculture Organization of the UN, 2013. FAOSTAT Food and Agricultural Commodities Production [WWW Document]. URL http://faostat3.fao.org/ (accessed 3.30.16).

Glaubitz, J.C., Casstevens, T.M., Lu, F., Harriman, J., Elshire, R.J., Sun, Q., Buckler, E.S., 2014. TASSEL-GBS: A High Capacity Genotyping by Sequencing Analysis Pipeline. PLOS ONE 9, e90346. doi:10.1371/journal.pone.0090346

Grafius, J.E., 1956. Components of Yield in Oats: A Geometrical Interpretation1. Agron J 48, 419–423. doi:10.2134/agronj1956.00021962004800090010x

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

Gregersen, P.L., Holm, P.B., 2007. Transcriptome analysis of senescence in the flag leaf of wheat (Triticum aestivum L.). Plant Biotechnol. J. 5, 192–206. doi:10.1111/j.1467-7652.2006.00232.x

Groos, C., Robert, N., Bervas, E., Charmet, G., 2003. Genetic analysis of grain protein-content, grain yield and thousand-kernel weight in bread wheat. Theor. Appl. Genet. 106, 1032–1040. doi:10.1007/s00122-002-1111-1

Hamblin, M.T., Buckler, E.S., Jannink, J.-L., 2011. Population genetics of genomics-based crop improvement methods. Trends Genet. 27, 98–106. doi:10.1016/j.tig.2010.12.003

Hay, R.K.M., 1995. Harvest index: a review of its use in plant breeding and crop physiology. Ann. Appl. Biol. 126, 197–216. doi:10.1111/j.1744-7348.1995.tb05015.x

IWGSC, 2014. A chromosome-based draft sequence of the hexaploid bread wheat (Triticum aestivum) genome. Science 345. doi:10.1126/science.1251788

Jenner, C., Rathjen, A., 1975. Factors Regulating the Accumulation of Starch in Ripening Wheat Grain. Funct. Plant Biol. 2, 311–322.

Jiang, Q., Hou, J., Hao, C., Wang, L., Ge, H., Dong, Y., Zhang, X., 2011. The wheat (T. aestivum) sucrose synthase 2 gene (TaSus2) active in endosperm development is associated with yield traits. Funct. Integr. Genomics 11, 49–61. doi:10.1007/s10142-010-0188-x

Kato, Y., Murakami, S., Yamamoto, Y., Chatani, H., Kondo, Y., Nakano, T., Yokota, A., Sato, F., 2004. The DNA-binding protease, CND41, and the degradation of ribulose-1,5-bisphosphate carboxylase/oxygenase in senescent leaves of tobacco. Planta 220, 97–104. doi:10.1007/s00425-004-1328-0

Kim, M.G., da Cunha, L., McFall, A.J., Belkhadir, Y., DebRoy, S., Dangl, J.L., Mackey, D., 2005. Two Pseudomonas syringae type III effectors inhibit RIN4-regulated basal defense in Arabidopsis. Cell 121, 749–759. doi:10.1016/j.cell.2005.03.025

Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25, 1754–1760. doi:10.1093/bioinformatics/btp324

Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., Sandstrom, R., Bernstein, B., Bender, M.A., Groudine, M., Gnirke, A., Stamatoyannopoulos, J., Mirny, L.A., Lander, E.S., Dekker, J., 2009. Comprehensive mapping of long range interactions reveals folding principles of the human genome. Science 326, 289–293. doi:10.1126/science.1181369

Listgarten, J., Lippert, C., Kadie, C.M., Davidson, R.I., Eskin, E., Heckerman, D., n.d. Improved linear mixed models for genome-wide association studies. Nat. Methods 9, 525. doi:10.1038/nmeth.2037

Liu, X., Huang, M., Fan, B., Buckler, E.S., Zhang, Z., 2016. Iterative Usage of Fixed and Random Effect Models for Powerful and Efficient Genome-Wide Association Studies. PLOS Genet 12, e1005767. doi:10.1371/journal.pgen.1005767

Lobell, D.B., Schlenker, W., Costa-Roberts, J., 2011. Climate Trends and Global Crop Production Since 1980. Science 333, 616–620. doi:10.1126/science.1204531

Mascher, M., Muehlbauer, G.J., Rokhsar, D.S., Chapman, J., Schmutz, J., Barry, K., Muñoz-Amatriaín, M., Close, T.J., Wise, R.P., Schulman, A.H., Himmelbach, A., Mayer, K.F., Scholz, U., Poland, J.A., Stein, N., Waugh, R., 2013. Anchoring and ordering NGS contig assemblies by population sequencing (POPSEQ). Plant J. 76, 718–727. doi:10.1111/tpj.12319

Mascher, M., Stein, N., 2014. Genetic anchoring of whole-genome shotgun assemblies. Front. Genet. 5. doi:10.3389/fgene.2014.00208

McLaren, W., Pritchard, B., Rios, D., Chen, Y., Flicek, P., Cunningham, F., 2010. Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. Bioinforma. Oxf. Engl. 26, 2069–2070. doi:10.1093/bioinformatics/btq330

Money, D., Gardner, K., Migicovsky, Z., Schwaninger, H., Zhong, G.-Y., Myles, S., 2015. LinkImpute: Fast and Accurate Genotype Imputation for Nonmodel Organisms. G3 GenesGenomesGenetics 5, 2383–2390. doi:10.1534/g3.115.021667

Moreno-Sevilla, B., Baenziger, P.S., Peterson, C.J., Graybosch, R.A., McVey, D.V., 1995. The 1BL/1RS Translocation: Agronomic Performance of F3-Derived Lines from a Winter Wheat Cross. Crop Sci. 35, 1051–1055. doi:10.2135/cropsci1995.0011183X003500040022x

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

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

Niciporovic, A.A., 1956. Photosynthesis and the Theory of High Crop Yield, in: 15th Timirjazev Lecture. Timirjazey Inst. Plant Pathology, USSR Academy of Sciences, Moscow.

Nyquist, W.E., 1962. Differential Fertilization in the Inheritance of Stem Rust Resistance in Hybrids Involving a Common Wheat Strain Derived from Triticum Timopheevi. Genetics 47, 1109–1124.

Phillips, S.B., Keahey, D.A., Warren, J.G., Mullins, G.L., 2004. Estimating Winter Wheat Tiller Density Using Spectral Reflectance Sensors for Early-Spring, Variable-Rate Nitrogen Applications. Agron. J. 96, 591. doi:10.2134/agronj2004.0591

Poland, J.A., Brown, P.J., Sorrells, M.E., Jannink, J.-L., 2012. Development of High-Density Genetic Maps for Barley and Wheat Using a Novel Two-Enzyme Genotyping-by-Sequencing Approach. PLoS ONE 7, e32253. doi:10.1371/journal.pone.0032253

Putnam, N.H., O'Connell, B.L., Stites, J.C., Rice, B.J., Blanchette, M., Calef, R., Troll, C.J., Fields, A., Hartley, P.D., Sugnet, C.W., Haussler, D., Rokhsar, D.S., Green, R.E., 2016. Chromosome-scale shotgun assembly using an in vitro method for long-range linkage. Genome Res. 26, 342–350. doi:10.1101/gr.193474.115

R Core Team, 2015. R: A language and environment for statistical computing.

Reif, J.C., Gowda, M., Maurer, H.P., Longin, C.F.H., Korzun, V., Ebmeyer, E., Bothe, R., Pietsch, C., Würschum, T., 2011. Association mapping for quality traits in soft winter wheat. Theor. Appl. Genet. 122, 961–970. doi:10.1007/s00122-010-1502-7

Reynolds, M.P., Pellegrineschi, A., Skovmand, B., 2005. Sink-limitation to yield and biomass: a summary of some investigations in spring wheat. Ann. Appl. Biol. 146, 39–49. doi:10.1111/j.1744-7348.2005.03100.x

Richards, R.A., 2000. Selectable traits to increase crop photosynthesis and yield of grain crops. J. Exp. Bot. 51, 447–458. doi:10.1093/jexbot/51.suppl_1.447

Roberts, I.N., Caputo, C., Criado, M.V., Funk, C., 2012. Senescence-associated proteases in plants. Physiol. Plant. 145, 130–139. doi:10.1111/j.1399-3054.2012.01574.x

Saghai-Maroof, M.A., Soliman, K.M., Jorgensen, R.A., Allard, R.W., 1984. Ribosomal DNA spacer-length polymorphisms in barley: mendelian inheritance, chromosomal location, and population dynamics. Proc. Natl. Acad. Sci. U. S. A. 81, 8014–8018.

Salvi, S., Tuberosa, R., 2015. The crop QTLome comes of age. Curr. Opin. Biotechnol., Food Biotechnology • Plant Biotechnology 32, 179–185. doi:10.1016/j.copbio.2015.01.001

Schmitz-Linneweber, C., Small, I., 2008. Pentatricopeptide repeat proteins: a socket set for organelle gene expression. Trends Plant Sci. 13, 663–670. doi:10.1016/j.tplants.2008.10.001

Segura, V., Vilhjálmsson, B.J., Platt, A., Korte, A., Seren, Ü., Long, Q., Nordborg, M., 2012. An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. Nat. Genet. 44, 825–830. doi:10.1038/ng.2314

Sharma, R.C., Crossa, J., Velu, G., Huerta-Espino, J., Vargas, M., Payne, T.S., Singh, R.P., 2012. Genetic Gains for Grain Yield in CIMMYT Spring Bread Wheat across International Environments. Crop Sci. 52, 1522–1533. doi:10.2135/cropsci2011.12.0634

Silva, S.A., Carvalho, F.I.F. de, Caetano, V. da R., Oliveira, A.C. de, Coimbra, J.L.M. de, Vasconcellos, N.J.S. de, Lorencetti, C., 2001. Genetic Basis of Stay-Green Trait in Bread Wheat. J. New Seeds 2, 55–68. doi:10.1300/J153v02n02_05

Singh, R.P., Huerta-Espino, J., Rajaram, S., Crossa, J., 1998. Agronomic Effects from Chromosome Translocations 7DL.7Ag and 1BL.1RS in Spring Wheat. Crop Sci. 38, 27–33. doi:10.2135/cropsci1998.0011183X003800010005x

Slafer, G., Calderini, D., Miralles, D., 1996. Yield components and compensation in wheat: opportunities for further increasing yield potential, in: Increasing Yield Potential in Wheat: Breaking the Barriers. pp. 101–133.

Slafer, G.A., Savin, R., 1994. Source—sink relationships and grain mass at different positions within the spike in wheat. Field Crops Res. 37, 39–49. doi:10.1016/0378-4290(94)90080-9

Somers, D.J., Isaac, P., Edwards, K., 2004. A high-density microsatellite consensus map for bread wheat (Triticum aestivum L.). Theor. Appl. Genet. 109, 1105–1114. doi:10.1007/s00122-004-1740-7

Spindel, J., Begum, H., Akdemir, D., Virk, P., Collard, B., Redoña, E., Atlin, G., Jannink, J.-L., McCouch, S.R., 2015. Genomic selection and association mapping in rice (Oryza sativa): effect of trait genetic architecture, training population composition, marker number and statistical model on accuracy of rice genomic selection in elite, tropical rice breeding lines. PLoS Genet. 11, e1004982. doi:10.1371/journal.pgen.1004982

Sukumaran, S., Dreisigacker, S., Lopes, M., Chavez, P., Reynolds, M.P., 2014. Genome-wide association study for grain yield and related traits in an elite spring wheat population grown in temperate irrigated environments. Theor. Appl. Genet. 128, 353–363. doi:10.1007/s00122-014-2435-3

Tadesse, W., Ogbonnaya, F.C., Jighly, A., Sanchez-Garcia, M., Sohail, Q., Rajaram, S., Baum, M., 2015. Genome-Wide Association Mapping of Yield and Grain Quality Traits in Winter Wheat Genotypes. PLoS ONE 10, e0141339. doi:10.1371/journal.pone.0141339

Takahashi, R., Bashir, K., Ishimaru, Y., Nishizawa, N.K., Nakanishi, H., 2012. The role of heavy-metal ATPases, HMAs, in zinc and cadmium transport in rice. Plant Signal. Behav. 7, 1605–1607. doi:10.4161/psb.22454

Tang, Y., Liu, X., Wang, J., Li, M., Wang, Q., Tian, F., Su, Z., Pan, Y., Liu, D., Lipka, A.E., Buckler, E.S., Zhang, Z., 2016. GAPIT Version 2: An Enhanced Integrated Tool for Genomic Association and Prediction. Plant Genome 9. doi:10.3835/plantgenome2015.11.0120

Terman, G.L., Ramig, R.E., Dreier, A.F., Olson, R.A., 1969. Yield-Protein Relationships in Wheat Grain, as Affected by Nitrogen and Water. Agron. J. 61, 755–759. doi:10.2134/agronj1969.00021962006100050031x

Tian, D., Traw, M.B., Chen, J.Q., Kreitman, M., Bergelson, J., 2003. Fitness costs of R-gene-mediated resistance in Arabidopsis thaliana. Nature 423, 74–77. doi:10.1038/nature01588

Tilman, D., Cassman, K.G., Matson, P.A., Naylor, R., Polasky, S., 2002. Agricultural sustainability and intensive production practices. Nature 418, 671–677. doi:10.1038/nature01014

Tucker, G., Price, A.L., Berger, B., 2014. Improving the Power of GWAS and Avoiding Confounding from Population Stratification with PC-Select. Genetics 197, 1045–1049. doi:10.1534/genetics.114.164285

United Nations Department of Economic and Social Affairs, Population Division, 2015. World Population Prospects: The 2015 Revision, Key Findings and Advance Tables.

USDA Foreign Agricultural Service, 2015. USDA Production, Supply and Distribution Online [WWW Document]. URL https://apps.fas.usda.gov/psdonline/ (accessed 3.30.16).

Valdar, W., Holmes, C.C., Mott, R., Flint, J., 2009. Mapping in Structured Populations by Resample Model Averaging. Genetics 182, 1263–1277. doi:10.1534/genetics.109.100727

Wallace, J.G., Zhang, X., Beyene, Y., Semagn, K., Olsen, M., Prasanna, B.M., Buckler, E.S., 2016. Genome-wide Association for Plant Height and Flowering Time across 15 Tropical Maize Populations under Managed Drought Stress and Well-Watered Conditions in Sub-Saharan Africa. Crop Sci. 56, 2365–2378. doi:10.2135/cropsci2015.10.0632

Wang, Q., Tian, F., Pan, Y., Buckler, E.S., Zhang, Z., 2014. A SUPER Powerful Method for Genome Wide Association Study. PLOS ONE 9, e107684. doi:10.1371/journal.pone.0107684

Wang, S., Wong, D., Forrest, K., Allen, A., Chao, S., Huang, B.E., Maccaferri, M., Salvi, S., Milner, S.G., Cattivelli, L., Mastrangelo, A.M., Whan, A., Stephen, S., Barker, G., Wieseke, R., Plieske, J., International Wheat Genome Sequencing Consortium, Lillemo, M., Mather, D., Appels, R., Dolferus, R., Brown-Guedira, G., Korol, A., Akhunova, A.R., Feuillet, C., Salse, J., Morgante, M., Pozniak, C., Luo, M.-C., Dvorak, J., Morell, M., Dubcovsky, J., Ganal, M., Tuberosa, R., Lawley, C., Mikoulitch, I., Cavanagh, C., Edwards, K.J., 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. doi:10.1111/pbi.12183

Weihofen, A., Binns, K., Lemberg, M.K., Ashman, K., Martoglio, B., 2002. Identification of signal peptide peptidase, a presenilin-type aspartic protease. Science 296, 2215–2218. doi:10.1126/science.1070925

Welter, D., MacArthur, J., Morales, J., Burdett, T., Hall, P., Junkins, H., Klemm, A., Flicek, P., Manolio, T., Hindorff, L., Parkinson, H., 2014. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. Nucleic Acids Res. 42, D1001–D1006. doi:10.1093/nar/gkt1229

Widmer, C., Lippert, C., Weissbrod, O., Fusi, N., Kadie, C., Davidson, R., Listgarten, J., Heckerman, D., 2014. Further Improvements to Linear Mixed Models for Genome-Wide Association Studies. Sci. Rep. 4, 6874. doi:10.1038/srep06874

Würschum, T., Langer, S.M., Longin, C.F.H., Korzun, V., Akhunov, E., Ebmeyer, E., Schachschneider, R., Schacht, J., Kazman, E., Reif, J.C., 2013. Population structure, genetic diversity and linkage disequilibrium in elite winter wheat assessed with SNP and SSR markers. Theor. Appl. Genet. 126, 1477–1486. doi:10.1007/s00122-013-2065-1

Yandell, M., Ence, D., 2012. A beginner's guide to eukaryotic genome annotation. Nat. Rev. Genet. 13, 329–342. doi:10.1038/nrg3174

Yang, J., Zaitlen, N.A., Goddard, M.E., Visscher, P.M., Price, A.L., 2014. Advantages and pitfalls in the application of mixed-model association methods. Nat. Genet. 46, 100–106. doi:10.1038/ng.2876

Yu, J.M., Pressoir, G., Briggs, W.H., Bi, I.V., Yamasaki, M., Doebley, J.F., McMullen, M.D., Gaut, B.S., Nielsen, D.M., Holland, J.B., Kresovich, S., Buckler, E.S., 2006. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat. Genet. 38, 203–208. doi:10.1038/ng1702

Zadoks, J.C., Chang, T.T., Konzak, C.F., 1974. A decimal code for the growth stages of cereals. Weed Res. 14, 415–421. doi:10.1111/j.1365-3180.1974.tb01084.x

Zamir, D., 2013. Where Have All the Crop Phenotypes Gone? PLOS Biol 11, e1001595. doi:10.1371/journal.pbio.1001595

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

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

Zanke, C.D., Ling, J., Plieske, J., Kollers, S., Ebmeyer, E., Korzun, V., Argillier, O., Stiewe, G., Hinze, M., Neumann, K., Ganal, M.W., Röder, M.S., 2014. Whole Genome Association Mapping of Plant Height in Winter Wheat (Triticum aestivum L.). PLOS ONE 9, e113287. doi:10.1371/journal.pone.0113287

Zhang, Z., Ersoz, E., Lai, C.-Q., Todhunter, R.J., Tiwari, H.K., Gore, M.A., Bradbury, P.J., Yu, J., Arnett, D.K., Ordovas, J.M., Buckler, E.S., 2010. Mixed linear model approach adapted for genome-wide association studies. Nat Genet 42, 355–360.

Zheng, X., Levine, D., Shen, J., Gogarten, S.M., Laurie, C., Weir, B.S., 2012. A high-performance computing toolset for relatedness and principal component analysis of SNP data. Bioinformatics 28, 3326–3328. doi:10.1093/bioinformatics/bts606

Table 1.1: Phenotypic traits assessed in Blacksburg and Warsaw, VA during the 2013-2014 and 2014-2015 growing seasons

Trait	Abbreviation	Units	Trait Ontology †	Ontology Description
Heading Date	HD	Julian days (Jan1)	TO:0000137	Days required for 50% of heads to emerge from boot
Flag Leaf Senescence	FLS	Julian days (Jan1)	TO:0000249	Days required for 50% of flag leaves to lose green color
Physiological Maturity	MAT	Julian days (Jan1)	TO:0000469	Days required for 50% of peduncles to lose green color
Days from Heading to Maturity	HP	days	-	Number of days from heading to maturity
Flag Leaf Stay Green	FLSG	days	TO:0000249	Days between heading and flag leaf senescence
Flag Leaf Angle at Boot Stage ‡	FLA	1-9 scale	TO:0000124	Flag leaf angle is scored based on the visual estimate of flag leaf angle at Zadok growth stage 50 (inflorescence emergence)
NDVI at Zadok's GS25	NDVI	-	CO_321:0000301	Normalized-difference vegetation index measured at spring green-up using a Trimble Greenseeker instrument
Canopy Temperature Depression at Anthesis ‡	CTDA	degrees Celsius	CO_321:0000006	Canopy temperature depression (CTD) is the difference between the ambient air temperature (Ta) and the canopy temperature (Tc): CTD=Ta-Tc
Canopy Temperature Depression at Grain Fill ‡	CTDF	degrees Celsius	CO_321:0000006	Same as above
Mature Plant Height	HT	cm	TO:0000207	Height of plant from soil surface to tip of spike excluding awns
Above-ground biomass at maturity	BIOM	g dwt m^{-1} row	CO_321:0000229	Weight of dry plant matter cut at soil level from a 1 m section of a center row of the plot
Harvest Index	HI	-	TO:0000128	Grain yield divided by above-ground biomass
Grain weight	GW	g dwt m^{-1} row	TO:0000589	Grain yield (g dry weight) from 1 m row cut at physiological maturity
Grains per square meter	GSQM	Grains m^{-2}	CO_321:0000017	Number of grains threshed from 1m row cutting, converted to square meters
Spikes per square meter	SSQM	Spikes m^{-2}	CO_321:0000166	Number of spikes (fertile culms) per unit area, sample or plant
Seeds per Head	SPH	count	TO:0002759	Number of grains within an inflorescence
Thousand Kernel Weight	TKW	grams	TO:0000382	Seed weight estimated by weighing 1000 seeds

Trait	Abbreviation	Units	Trait Ontology †	Ontology Description
Test Weight	TWT	g L^{-1}	TO:0000612	Weight per unit volume of grain at standard moisture level
Yield, 13.5% moisture	YLD	kg ha^{-1}	TO:0000396	Grain yield standardized to 13.5% moisture equivalence
NIR Whole grain starch	STARCH	%	-	-
NIR Wet Chemistry-validated whole grain protein	WCPROT	%	-	-
Lodging §	LOD	0-9 scale	CO_321:0001282	Degree of lodging; scale 0-9

† Matching trait ontologies in the Planteome database: http://browser.planteome.org/amigo/search/ontology

‡ No significant differences were observed between genotypes in data collected for the 2013-2014 season; data was subsequently not collected for the 2014-2015 season

§ Lodging was not extensive enough for reliable rating across environments

Table 1.2: Trait descriptive statistics, variance components, and entry-mean heritability for lines grown in Blacksburg, VA and Warsaw, VA for the 2013-2014 and 2014-2015 growing seasons

Trait †	Units	Descriptive Statistics				Variances and Heritability				
		min	mean	max	SD	σ^2 G	σ^2 E	σ^2 (G x E)	σ^2 ε	H^2
HD	Julian days (Jan1)	121	128	136	3.21	2.77	8.88	0.5	0.46	0.94
FLS	Julian days (Jan1)	148	156.7	169	4.55	2.42	20.77	0.56	1.48	0.88
MAT	Julian days (Jan1)	151	159.3	171	4.72	2.68	23.52	0.48	0.93	0.92
HP	days	23	31.27	39	2.84	1.1	5.93	0.84	1.23	0.75
FLSG	days	21	28.69	38	2.76	1.44	4.28	0.74	1.75	0.78
NDVI	-	0.26	0.54	0.75	0.08	0.0004	0.005	0.0002	0.001	0.67
HT	cm	59.69	85.43	119.4	9.29	21.27	73.03	3.47	7.9	0.92
BIOM	g dwt m^{-1} row	122.5	225.7	350	35.54	67.87	519.3	5.43E-12	777.5	0.41
HI	-	0.28	0.43	0.53	0.032	0.0005	6.50E-05	0.0001	0.0004	0.86
GW	g dwt m^{-1} row	47.68	96.64	157.9	16.53	33.19	108	2.38E-11	152.5	0.64
GSQM	Grains m^{-2}	8460	1.85E+04	3.13E+04	3277	3.17E+06	2.42E+06	3.21E+05	5.35E+06	0.81
SSQM	Spikes m^{-2}	459.3	853	1485	161.2	7005	1.05E+04	101.6	1.12E+04	0.83
SPH	count	8.54	21.94	33.29	3.04	4.92	1.29	0.75	2.64	0.90
TKW	grams	24.1	34.57	91.6	3.98	9.33	3.48	0.65	3.09	0.94
TWT	g L^{-1}	652.6	759	810.9	19.7	139.5	236.5	34.03	27.41	0.92
YLD	kg ha^{-1}	3579	6627	9053	1027	1.24E+05	9.75E+05	2.84E+04	1.49E+05	0.83
STARCH	%	46.88	52.51	56.49	1.41	0.29	1.41	0.15	0.37	0.78
WCPROT	%	9.67	12.34	16.04	1.01	0.21	0.51	0.09	0.35	0.76

σ^2 G genotypic variance; σ^2 E environmental variance; σ^2 (G x E) genotype x environment variance; σ^2 ε residual variance; H^2 entry-mean heritability.

† BIOM above-ground biomass; FLS flag leaf senescence; FLSG flag leaf stay green; GSQM grains per square meter; GW grain weight; HD heading date; HI harvest index; HP maturity date minus heading date; HT plant height; MAT physiological maturity date; NDVI normalized-difference vegetation index at Zadok's GS25; SPH seeds per head; SSQM spikes per square meter; STARCH whole-grain starch content; TKW thousand kernel weight; TWT test weight; WCPROT wet chemistry-validated whole-grain protein content; YLD grain yield;

Table 1.3: Allelic effects of SNP markers for agronomically important loci assessed via LGC® KASP™ SNP genotyping assays. Allelic effects are defined as the mean of lines homozygous for the alternate allele minus the mean of lines homozygous for the reference allele.

Trait †	1RS:1AL	1RS:1BL	Ppd-A1	Ppd-B1	Ppd-D1	Rht-B1	Rht-D1	Sr36	TaSus2-2B	Vrn-A1-Ex4	Vrn-A1-Ex7	Vrn-B1
HD	0.38*	0.06	-0.20	-0.41	-0.91*	0.15	-0.17	-0.03	-0.05	2.20**	0.88	-0.74
FLS	0.13**	0.57	0.087*	0.36*	-0.58	0.28	-0.01	0.58**	0.51	1.68**	1.18	0.27
MAT	0.40**	0.65*	0.19	-0.14*	-0.48	0.01	0.19	0.79**	0.79**	1.65**	1.18	0.23
HP	0.05	0.41**	0.26	0.20**	0.29	-0.12	0.28	0.57**	0.59**	-0.34	0.18	0.70**
FLSG	-0.15	0.41*	0.19	0.62**	0.21	0.10	0.15	0.48**	0.44**	-0.23	0.29	0.78**
NDVI	0.004	0.002	0.0001	0.003	0.0006	-0.001	-0.0009	-0.002	-0.0004	0.02**	-0.002	-0.004
HT	-3.41**	-1.05	0.47	-0.05	-4.06**	0.52	-2.00*	-4.87**	-4.79**	2.49	1.43	-0.24
BIOM	0.36	2.04*	-2.31**	1.05	2.31**	-2.41**	2.81**	-0.54	-0.54	1.99	-1.74	2.25
HI	0.0009	-0.0001	0.0005	0.001	0.005	0.002	0.004	0.009**	0.009**	-0.003	-0.02	0.005
GW	0.23	1.71*	-1.69*	1.03	2.58**	-1.53*	2.72**	1.08	0.99*	1.09	-4.10	2.55*
GSQM	-139	-417	197	-705**	75	501	-267	32.70	101	635	-2040*	-322
SSQM	21.20	-26.40	4.38	-18.60	-5.42	37.50**	-31.90*	-2.97	-0.11	61.40**	1.46	-55.10**
SPH	-0.78	0.30	0.11	-0.41	0.301	-0.41	0.53	0.18	0.20	-0.97	-3.05*	1.26*
TKW	0.42	2.14**	-1.36	2.49**	1.56*	-2.21**	2.49**	0.72	0.49	-0.80	2.46	2.54**
TWT	-0.62*	-3.38	-0.64	-1.55	-0.29	-4.15*	4.37**	-7.28**	-8.38**	-10.20**	3.52	2.36
YLD	-153*	90.50	-90.20	64.10	107*	-81.4	166**	41.20	27.40	262*	-7.85	131
STARCH	-0.21	-0.10	-0.03	-0.06	0.05	0.01	0.13	0.27**	0.29**	0.01	-0.16	0.18
WCPROT	0.14	0.11	0.15	0.06	-0.05	0.07	-0.14	0.04	0.01	-0.14	0.65**	-0.15*

* Allelic effects are significant at the 0.05 level; ** Allelic effects are significant at the 0.01 level

† HD heading date; FLS flag leaf senescence; MAT physiological maturity date; HP maturity date minus heading date; FLSG flag leaf stay green; NDVI normalized-difference vegetation index at Zadok's GS25; HT mature plant height; BIOM above-ground biomass; HI harvest index; GW grain weight; GSQM grains per square meter; SSQM spikes per square meter; SPH seeds per head; TKW thousand kernel weight; TWT test weight; YLD grain yield; STARCH whole-grain starch content; WCPROT wet chemistry-validated whole-grain protein content

Table 1.4: Significant marker-trait associations identified by the GCTA leave-one-chromosome-out method.

TRAIT †	n PCs	CHROM	SNP §	REF	ALT	MAF	P-VALUE	EFFECT	UNITS	QTL LOWER BOUND	QTL UPPER BOUND
TWT	4	1A	S1A_583587147	T	C	0.24	1.13E-06	-3.25	g L^{-1}	NA	NA
SSQM	4	3A	S3A_3273716	T	C	0.42	1.91E-06	-21.8	Spikes m^{-2}	S3A_3086373	S3A_3273716
GSQM	3	5B	S5B_396479359	T	C	0.46	5.12E-06	-452	Grains m^{-2}	NA	NA
FLS	4	6B	S6B_181128808	A	G	0.06	6.14E-06	-0.95	Julian days (Jan1)	NA	NA
FLSG	4	7B	S7B_64393207	G	A	0.17	4.83E-06	0.31	Days	NA	NA
FLS	4	7D	S7D_58449294	G	A	0.34	1.09E-06	-0.44	Julian days (Jan1)	S7D_55733261	S7D_59494513
MAT	4	7D	S7D_58589271	G	A	0.33	8.77E-07	-0.49	Julian days (Jan1)	S7D_55733261	S7D_59494513

n PCs number of principle components included in analysis; CHROM chromosome; REF reference allele; ALT alternate allele; EFFECT mean phenotype of lines containing alternate allele minus mean phenotype of lines containing reference allele

† TWT test weight; SSQM spikes per square meter; GSQM grains per square meter; FLS flag leaf senescence date; FLSG flag leaf stay green duration; MAT physiological maturity; ‡ number of principle components used to model population structure

§ SNP name includes physical position on chromosome; grey boxes indicate pleiotropic QTLs affecting multiple traits.

Significance thresholds for a 95% confidence level were determined empirically by performing permutation testing for 1,000 repetitions per trait. For QTLs in which adjacent blocks of SNPs in high LD exceeded the significance threshold, only the SNP with the lowest p-value is reported, with the first and last SNPs in the QTL reported for the lower and upper bounds respectively

Table 1.5: Significant* marker-trait associations identified by the FarmCPU algorithm.

TRAIT †	n PCs	CHROM	SNP §	REF	ALT	MAF	P-VALUE	EFFECT	UNITS	RMIP
TKW	4	1A	S1A_22935081	C	T	0.09	1.05E-07	0.93	grams	0.73
SSQM	4	1A	S1A_50513589	G	A	0.10	1.70E-08	27.3	Spikes m^{-2}	0.64
TWT	4	1A	S1A_583587147	C	T	0.24	6.45E-07	-2.64	g L^{-1}	0.41
STARCH	4	1B	S1B_39294256	G	A	0.23	9.02E-18	-0.20	%	0.64
MAT	4	1B	S1B_44010984	A	G	0.30	3.33E-09	0.40	Julian days (Jan1)	0.23
HD	4	1B	S1B_50850397	A	G	0.28	1.49E-07	0.34	Julian days (Jan1)	0.54
STARCH	4	1B	S1B_659857468	C	T	0.44	7.50E-08	-0.09	%	0.06
STARCH	4	1D	S1D_6674498	G	C	0.11	7.05E-08	-0.14	%	0.15
HT	3	2A	S2A_764941637	G	A	0.22	5.46E-07	-0.97	cm	0.09
GSQM	4	2B	S2B_35041282	T	G	0.08	3.89E-07	-693	Grains m^{-2}	0.06
HP	4	2B	S2B_66522933	A	C	0.32	3.69E-07	0.20	days	0.31
HD	4	2D	S2D_9872868	A	T	0.36	5.64E-07	-0.30	Julian days (Jan1)	0.28
HD	4	2D	S2D_35084672	G	A	0.12	1.39E-10	-0.55	Julian days (Jan1)	0.88
MAT	4	2D	S2D_35084672	G	A	0.12	5.91E-09	-0.52	Julian days (Jan1)	0.26
SSQM	4	3A	S3A_2493807	C	T	0.43	4.59E-07	14.0	Spikes m^{-2}	0.15
GSQM	4	3A	S3A_20232500	C	A	0.14	2.47E-07	-416	Grains m^{-2}	0.05
FLS	4	3A	S3A_569991635	T	C	0.08	5.21E-07	0.61	Julian days (Jan1)	0.16
GSQM	4	3A	S3A_691815364	G	A	0.44	5.31E-09	-350	Grains m^{-2}	0.14
TWT	4	3B	S3B_695966897	A	G	0.07	1.78E-09	-4.79	g L^{-1}	0.5
YLD	4	3D	S3D_511264206	C	T	0.08	1.17E-06	104	kg ha^{-1}	0.09
YLD	4	4A	S4A_726716318	G	A	0.35	7.43E-07	63.6	kg ha^{-1}	0.1
GSQM	4	4A	S4A_739598141	C	T	0.10	7.34E-10	685	Grains m^{-2}	0.27
HI	3	4B	S4B_626390000	A	C	0.06	4.56E-07	-0.01	-	0.16
STARCH	4	4B	S4B_626390000	A	C	0.06	4.71E-07	-0.23	%	0.18
STARCH	4	5A	S5A_9462259	A	C	0.19	9.65E-07	0.10	%	0.1
STARCH	4	5B	S5B_34721398	A	G	0.45	2.48E-08	0.09	%	0.31
HT	3	5B	S5B_261134879	G	A	0.31	2.70E-12	1.20	cm	0.5

TRAIT †	n PCs	CHROM	SNP §	REF	ALT	MAF	P-VALUE	EFFECT	UNITS	RMIP
HI	3	5B	S5B_394707451	G	A	0.46	1.24E-06	-0.004	-	0.23
GSQM	4	5B	S5B_396479359	C	T	0.46	1.78E-06	-319	Grains m^{-2}	0.21
HI	3	5B	S5B_644947034	T	A	0.07	4.65E-09	-0.01	-	0.15
STARCH	4	5D	S5D_365732020	A	C	0.17	2.61E-10	-0.14	%	0.26
HT	3	5D	S5D_451607895	G	C	0.05	2.27E-06	-2.25	cm	0.11
SSQM	4	5D	S5D_499069158	G	A	0.21	9.49E-11	27.83	Spikes m^{-2}	0.91
FLS	4	6A	S6A_63296169	A	G	0.32	1.59E-10	-0.45	Julian days (Jan1)	0.36
SSQM	4	6A	S6A_614373502	A	G	0.33	5.55E-08	-11.85	Spikes m^{-2}	0.14
TWT	4	6A	S6A_614660970	T	C	0.11	1.95E-06	-3.00	g L^{-1}	0.31
MAT	4	6B	S6B_32730233	G	C	0.10	1.31E-07	-0.49	Julian days (Jan1)	0.1
YLD	4	6B	S6B_656279771	C	T	0.14	7.60E-10	-107	kg ha^{-1}	0.18
HT	3	6B	S6B_706326554	A	G	0.06	2.08E-06	1.57	cm	0.15
HD	4	6D	S6D_127384672	G	A	0.09	1.43E-08	-0.77	Julian days (Jan1)	0.67
TKW	4	6D	S6D_468113959	A	T	0.06	2.73E-09	-1.33	grams	0.38
GSQM	4	7A	S7A_673387152	C	T	0.46	1.47E-06	302	Grains m^{-2}	0.2
TKW	4	7A	S7A_673436887	T	C	0.48	6.34E-07	-0.55	grams	0.54
HP	4	7B	S7B_41890395	A	C	0.26	3.77E-11	0.25	days	0.7
FLSG	4	7B	S7B_63999446	A	T	0.23	9.19E-07	0.21	days	0.35
FLS	4	7D	S7D_58449294	A	G	0.34	2.24E-10	-0.40	Julian days (Jan1)	0.44
HD	4	7D	S7D_58589271	A	G	0.33	7.52E-13	-0.58	Julian days (Jan1)	0.82
MAT	4	7D	S7D_58589271	A	G	0.33	8.22E-08	-0.36	Julian days (Jan1)	0.35
HD	4	7D	S7D_553110861	C	T	0.37	1.49E-07	-0.30	Julian days (Jan1)	0.21
FLS	4	KASP	KASP_TAPPDDD001	-	indel	0.33	7.65E-08	-0.35	Julian days (Jan1)	0.55
FLSG	4	KASP	PPD1_B1	-	indel	0.26	3.74E-08	0.26	days	0.5

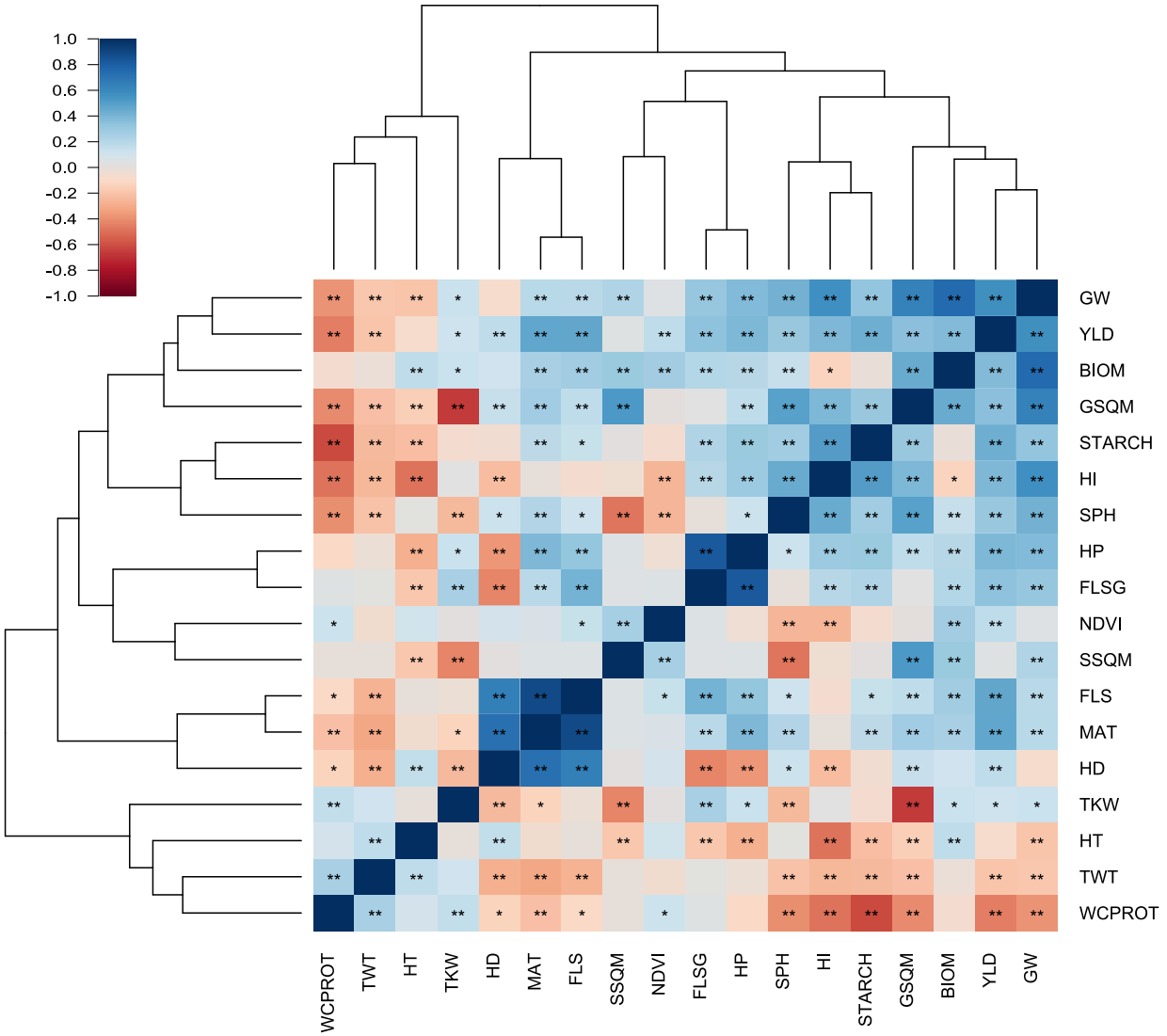
* Significance thresholds for a 95% confidence level were determined empirically by performing permutation testing for 1,000 repetitions per trait.

n PCs number of principle components included in analysis; CHROM chromosome; REF reference allele; ALT alternate allele; EFFECT mean phenotypic value of lines containing alternate allele minus mean phenotypic value of lines containing reference allele; RMIP resample model inclusion probability (proportion of bootstraps in which SNP exceeded significance threshold)

† TKW thousand kernel weight; SSQM spikes per square meter; TWT test weight; STARCH whole-grain start content; MAT physiological maturity date; HD heading date; HT mature plant height; GSQM grains per square meter; HP maturity date minus heading date; FLS flag leaf senescence date; YLD grain yield; HI harvest index; FLSG flag leaf stay green duration

§ SNP name includes physical position on chromosome; grey boxes indicate pleiotropic QTLs affecting multiple traits

Figure 1.1: Pearson's correlation coefficients among traits in the ABB panel based upon genotypic BLUPs calculated from data collected from Warsaw, VA and Blacksburg, VA in the 2013-2014 and 2014-2015 growing seasons



* Correlation is significant at the 0.05 level; ** Correlation is significant at the 0.01 level

GW grain weight; YLD grain yield; BIOM above-ground biomass; GSQM grains per square meter; STARCH whole-grain starch content; HI harvest index; SPH seeds per head; HP maturity date minus heading date; FLSG flag leaf stay green duration; NDVI normalized-difference vegetation index at Zadok's GS25; SSQM spikes per square meter; FLS flag leaf senescence date; MAT physiological maturity date; HD heading date; TKW thousand-kernel weight; HT mature plant height; TWT test weight; WCPROT wet chemistry-validated whole-grain protein content

Figure 1.2a: Portion of variance explained by each genotypic principle component for all lines included in both years of the study

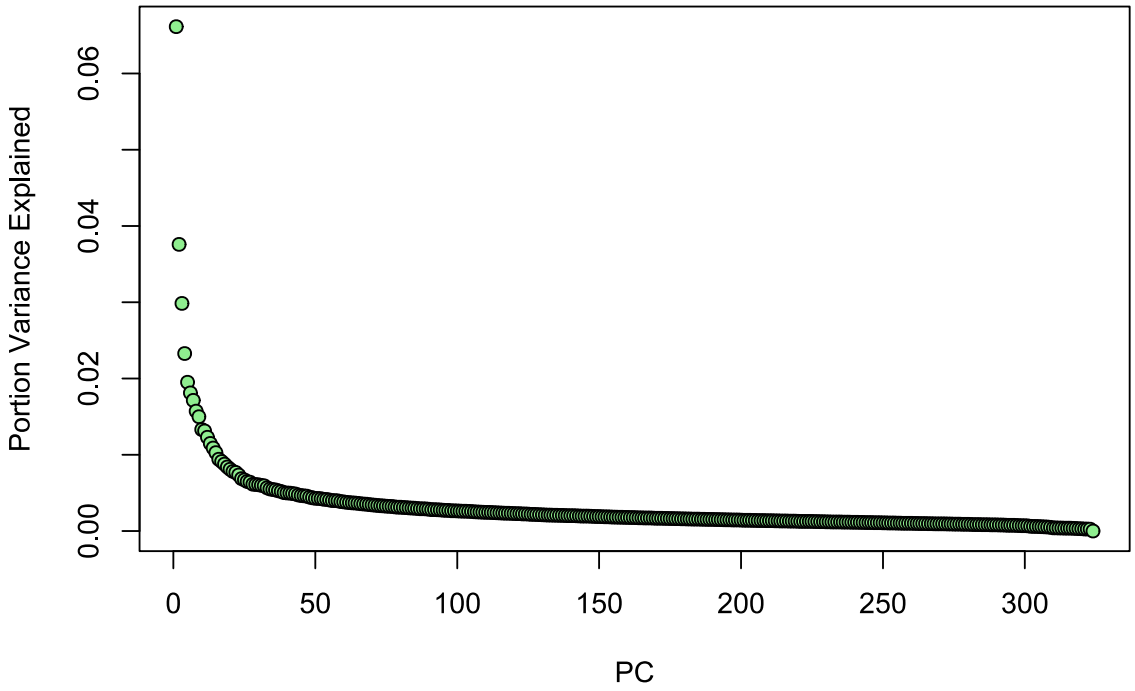


Figure 1.2b: Cumulative portion of variance explained by each genotypic principle component for all lines included in both years of the study

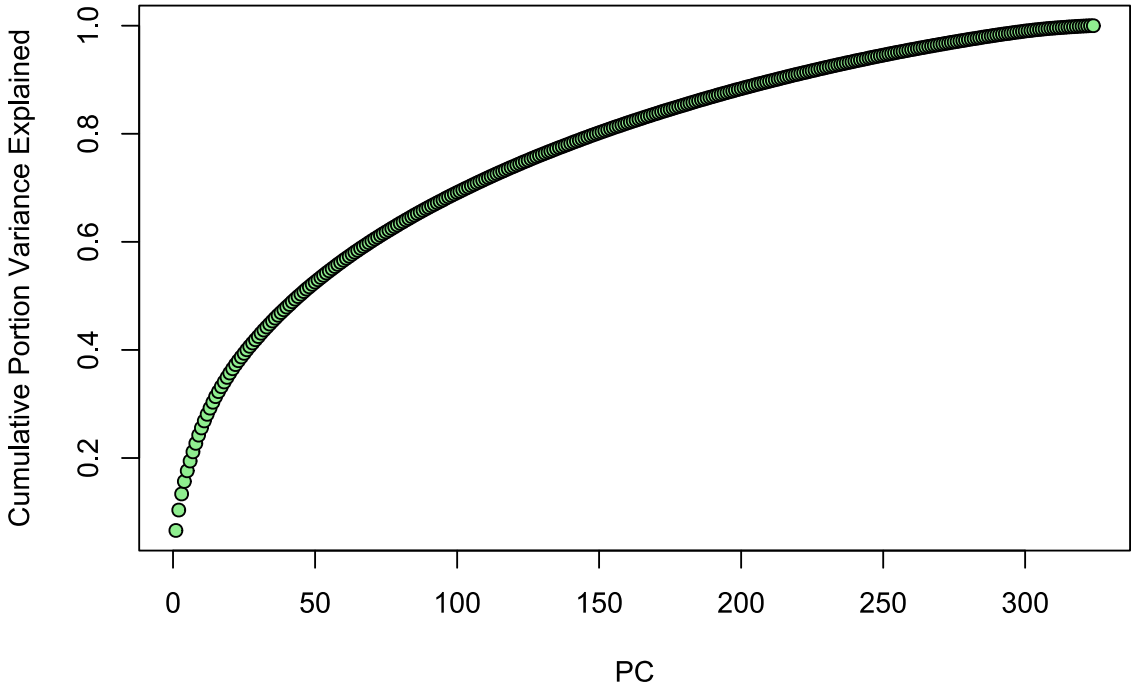


Figure 1.3: Density plots and biplots of the first four principal components of the genotypic data shaded by presence or absence of the *Sr36* stem rust resistance gene located on chromosome 2B

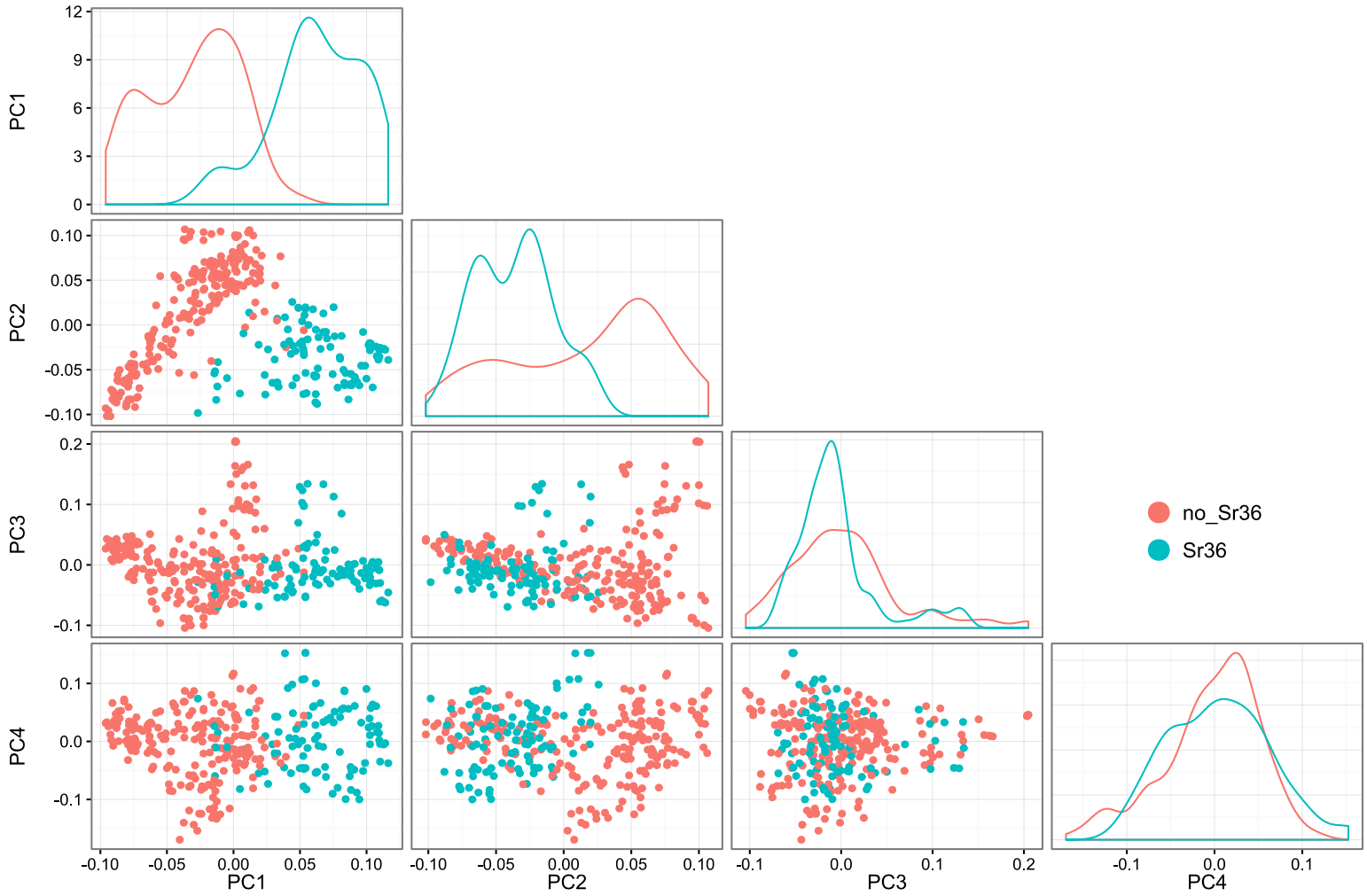


Figure 1.4: Linkage disequilibrium by sub-genome for the combined ABB genotypic data. Lines represent second-degree LOESS curves fit to 20,000 randomly-selected intra-chromosomal pairwise genotypic r^2 estimates pooled from chromosomes in each sub-genome. Horizontal line corresponds to the 98th percentile of pairwise r^2 estimates for non-linked (i.e. inter-chromosomal) SNPs.

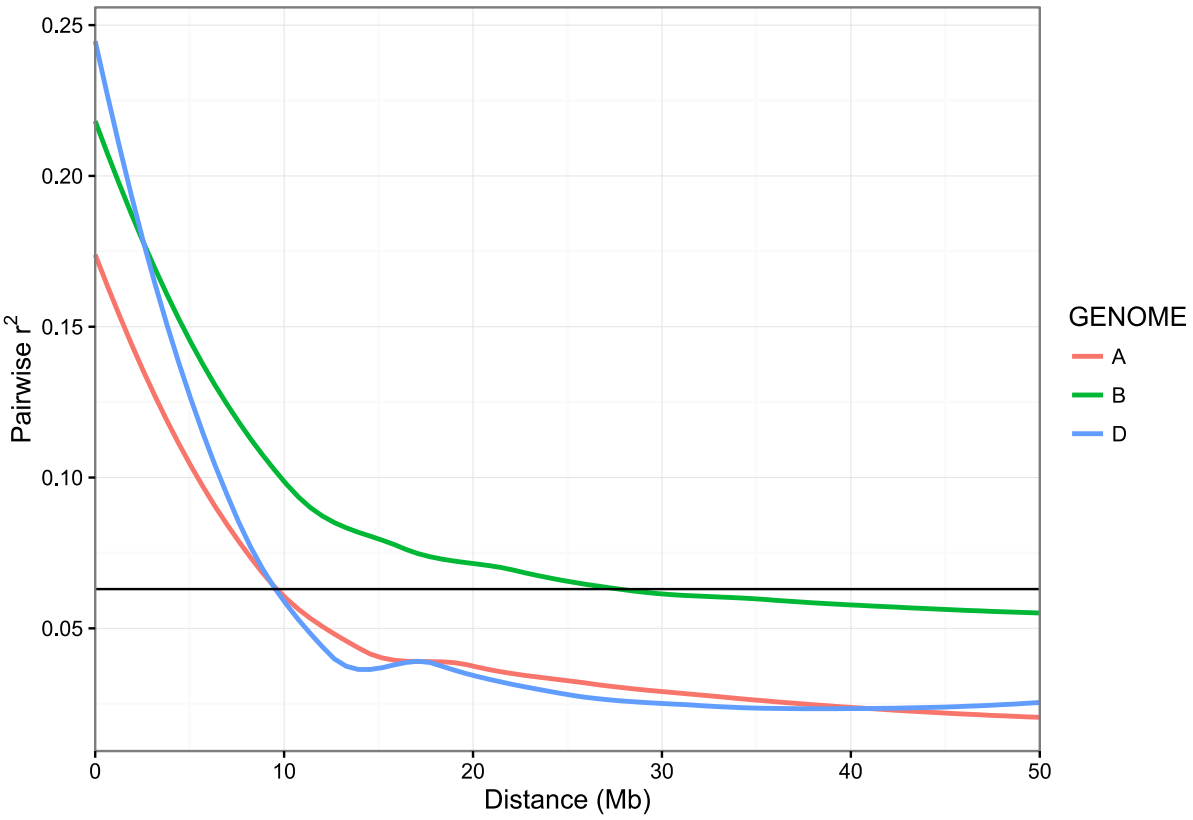


Figure 1.5: Linkage disequilibrium in each chromosome of the B genome for the combined ABB genotypic dataset. Lines represent second-degree LOESS curves fit to 20,000 randomly-selected intra-chromosomal pairwise genotypic r^2 estimates from each chromosome in the B genome. Horizontal line corresponds to the 98th percentile of pairwise r^2 estimates for non-linked (i.e. inter-chromosomal) SNPs.

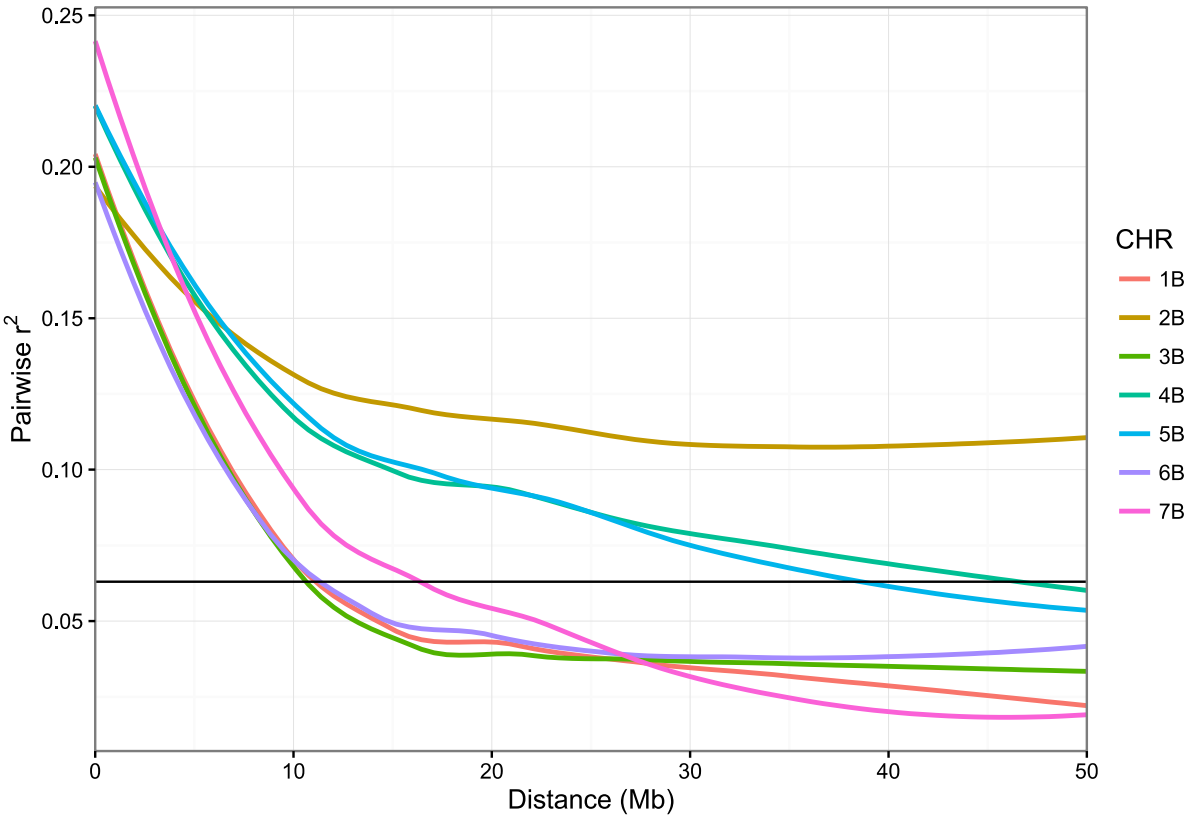


Figure 1.6: Genotypic linkage disequilibrium (r^2) measured between the KASP *Sr36* marker and all other SNPs present on chromosome 2B

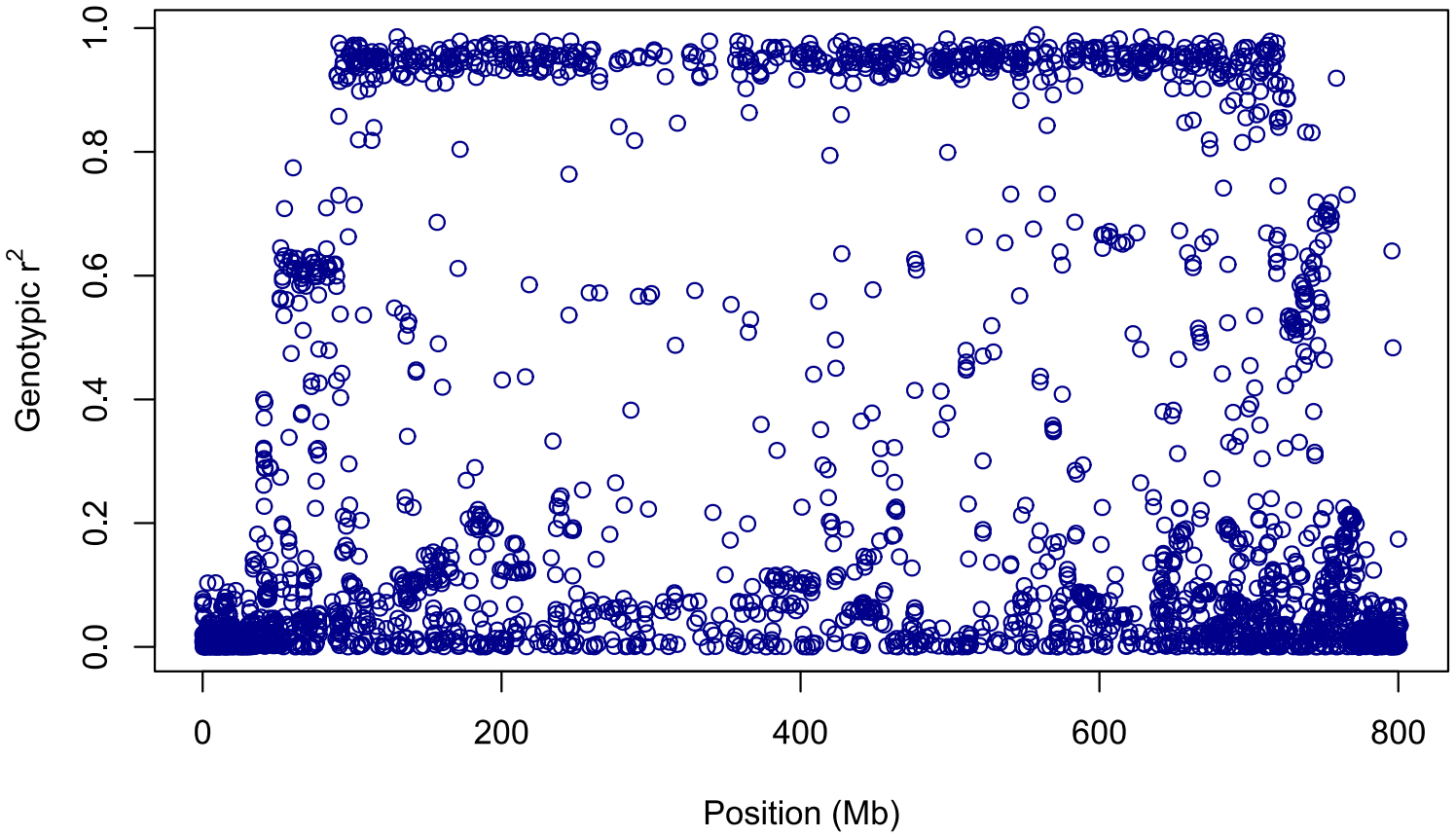
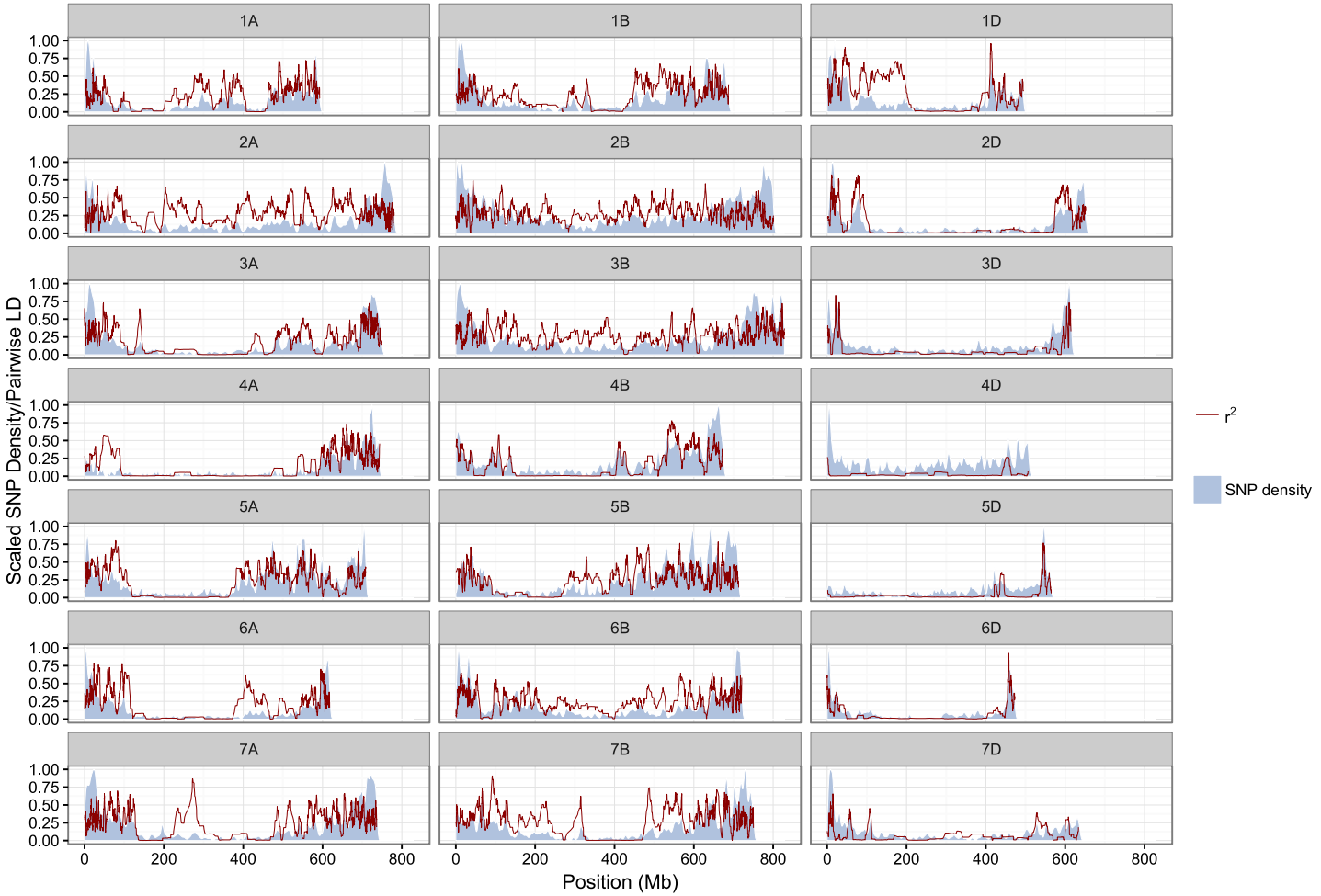


Figure 1.7: Linkage disequilibrium and SNP density for the combined ABB genotypic data. Red line indicates the moving average of genotypic LD (r^2) between adjacent SNPs using a 20-SNP sliding window. Grey density curve represents SNP physical density scaled to a maximum value of 1



CHAPTER II

Use of Multivariate Genomic Selection Models in Unbalanced Early-Generation Wheat Yield Trials

B. Ward, F. Kolb, D. Van Sanford, G. Brown-Guedira, P. Tyagi, C. Sneller, and C. Griffey

Abstract

Genomic selection (GS) is a recently-developed form of marker-assisted selection (MAS) in which many markers spread throughout the genome are used to estimate the genetic relationships between individuals and generate genomic-estimated breeding values (GEBVs). While GS has now been used in livestock breeding for several decades, it has only recently been adopted in plant breeding. Thus far, the majority of studies evaluating GS for plant breeding have utilized two-step approaches in which adjusted means across environments are first calculated and then used to generate GEBVs in a subsequent mixed linear model. In addition, GS has typically been performed on one or a few traits using balanced experimental designs. However, such studies ignore or minimize some of the unique complexities of plant breeding, including highly significant genotype-by-environment interaction (GEI), highly unbalanced designs across locations and years, and the need to perform selection on multiple correlated traits simultaneously. A goal of this study was to test GS for phenotypic prediction in scenarios that more accurately reflect early-generation yield testing by examining a total of 16 traits of varying heritability. Multivariate genomic best linear unbiased prediction (GBLUP) methods were used to build both multi-environment and multi-trait models. In general, the advantages of multi-environment models were most pronounced for moderate-heritability traits with homogeneous patterns of GEI among genotypes. In contrast, multi-environment models performed worse than a two-step model using the adjusted means across environments if excessive numbers of cross-over type GEI patterns were present in the training population. Multi-trait models using highly correlated traits generally far outperformed single-trait models in cases where GS was used to predict performance when each genotype had already been phenotyped for some, but not all traits.

Introduction

Genomic selection (GS) is a type of marker-assisted selection (MAS). However, traditional MAS utilizes a small number of markers, each in high linkage disequilibrium (LD) with quantitative trait loci (QTLs) exerting large effects upon the trait of interest, whereas GS utilizes a large number of markers spread throughout the genome, such that every QTL is assumed to be in high LD with at least one marker (Meuwissen et al., 2001). All GS models utilize a training population (TP), which is both genotyped and phenotyped, to produce predictions of breeding values in a validation population (VP), which is only genotyped. Thus there are several qualities of the TP, and its relationship to the VP, which are critical to achieving high predictive accuracy. These include the TP's nominal population size (N), its effective population size (N_e), the respective levels of linkage disequilibrium within the TP and the VP, and the degree of relatedness between the TP and VP (Bassi et al., 2016). Many of these qualities are interrelated; i.e. a large N_e is more generally associated with a low extent of LD within the population, and a lower degree of relatedness between individuals. In addition, larger N_e values require denser marker coverage to adequately saturate linkage groups. At the extreme, for a population of "unrelated" individuals, achieving a prediction accuracy of 0.9 would require ($10 \times N_e \times L$) SNP markers and a training population size of ($2 \times N_e \times L$), where L is the genome size in Morgans (Meuwissen, 2009). The central importance of N_e implies large differences in the resources required to effectively perform GS between species. For instance, in livestock breeding populations, N_e may be as low as 100, while in many human studies, N_e may be on the order of 10,000 (Hayes et al., 2009). The size of the human female genome has been calculated at 4,782 centiMorgans (Morton, 1991). Therefore, achieving a prediction accuracy of 0.9 for a trait in human females could require a training population size of close to 1 million, and the use of approximately 4.7 million markers.

Fortunately, crop species tend to require far fewer resources for effective prediction due to their generally low effective population sizes and highly extensive LD. For example, a study on Fusarium head blight (FHB) resistance in 251 soft winter wheat (*Triticum aestivum*) genotypes sourced from the eastern United States calculated an N_e value of 45 (Benson et al., 2012). For achieving prediction accuracies above 0.5 in wheat, Bassi et al. (2016) recommended a TP size of at least 50 individuals if entries of the VP are full-sibs of entries in the TP, 100 individuals for half-sibs, and 1,000 individuals for less related TPs and VPs.

A multitude of different models have been described for performing GS. Like many regression methods employing predictors sourced from genomic data, GS models must address the "big p, little n" problem, where the number of parameters (i.e. markers) typically far outweigh the number of observations (i.e. genotypes). In such settings, linear models constructed via ordinary least squares estimation will typically exhibit overfitting and consequently poor generalization across datasets. Regularization regression methods impose a penalty term for overfitting; these methods, which are often utilized in GS, include ridge regression (Tikhonov, 1963), the least absolute shrinkage and selection operator (Tibshirani, 1994), and the elastic net (Zou and Hastie, 2005).

Regularized regression has been incorporated into several mixed model methods for performing GS. These methods follow the assumption that marker effects are governed by Fisher's infinitesimal model (1918). That is, marker effects are assumed to represent a large number of genes uniformly spread throughout the genome, with each gene, and hence each marker, contributing equally to the observed phenotype. Ridge regression best linear unbiased prediction (RR-BLUP) calculates the summation of marker effects for each individual (Meuwissen et al., 2001; Whittaker et al., 2000). Genomic best linear unbiased prediction

(GBLUP) is a closely-related method in which a realized relationship matrix is calculated from marker data, and then used as the variance-covariance structure for the additive genetic effects in the mixed model (VanRaden, 2008; VanRaden et al., 2009). Habier et al. (2007) demonstrated that these two methods are equivalent for a fixed number of phenotypic observations as the number of markers approaches infinity. Both RR-BLUP and GBLUP may be contrasted with the traditional BLUP method used for solving mixed-model animal breeding equations (Henderson, 1975), in which the additive genetic variance-covariance structure is defined using the numerator relationship matrix, which is constructed from pedigree records. While BLUP methods for solving mixed models have found widespread use in animal breeding for decades, they have only more recently been adopted in plant breeding (Piepho et al., 2008).

In addition to the aforementioned frequentist methods, a number of Bayesian methods for performing GS have been described, starting with two models (Bayes A and Bayes B) introduced by Meuwissen et al. (2001). Since that time, there has been a rapid proliferation of Bayesian GS models (Gianola, 2013). These models all share the same linear regression of phenotypes on markers, but differ in the specific prior probability distributions that are employed for marker effects. Unlike RR-BLUP and GBLUP, in which all marker effects are assumed equal, the Bayesian alphabet methods as well as LASSO and elastic net incorporate variable selection, such that marker effects are allowed to vary. This should theoretically improve their accuracy when working with traits which are assumed to be controlled by fewer QTLs of large effect. All Bayesian methods share the characteristic of high computational burdens due to the Markov-chain Monte-Carlo estimation of posterior distributions that must be employed. In addition, despite the large numbers of GS models introduced, only marginal differences in predictive ability between models has been observed, owing to the species being tested or the genetic

architecture of the trait being examined (dos Santos et al., 2016; Gianola et al., 2014). In practice the choice of model depends not only upon its predictive ability, but also its ease of implementation and use of computational resources. This combination of factors led Heslot et al. (2012) to recommend using either mixed-model BLUP methods incorporating shrinkage estimators, Bayesian LASSO, or weighted Bayesian shrinkage regression after fitting a total of eleven models to eight different wheat, barley (*Hordeum vulgare*), maize (*Zea mays*), and *Arabidopsis thaliana* datasets. Wimmer et al. (2013) likewise compared RR-BLUP against two variable-selection methods (Bayes B and LASSO), and found that the latter models did not outperform RR-BLUP when applied to plant breeding datasets. This applied even to traits assumed to be controlled by few QTLs of major effect, such as flowering time in rice (*Oryza sativa*) and FRIGIDA gene expression in *A. thaliana*.

Multivariate Methods

Multivariate BLUP was originally developed to generate breeding value predictions utilizing data from multiple traits in livestock breeding (Henderson and Quaas, 1976). If one considers that multivariate techniques are equally applicable to measurements of either multiple traits in a single environment, or a single trait in multiple environments (Falconer and Mackay, 1996), then the utility of multivariate BLUP for evaluating both multi-environment and multi-trait data in plant breeding trials becomes apparent. While multivariate BLUP has been utilized extensively in animal breeding, it was only much more recently adopted in plant breeding (Piepho et al., 2008).

Early GS studies carried out in plants utilized univariate models adopted from animal breeding and hence lacked the means for incorporating information across multiple environments

(Bernardo and Yu, 2007), which limited their applicability. Therefore, many crop GS studies have relied on a two-stage approach, where adjusted means across environments are first calculated, and then subsequently used as the phenotypic response in a GBLUP model (Oakey et al., 2016). Several recent GS studies have begun integrating GEI to model phenotypic performance across environments. Burgueño et al. (2012) developed a multi-environment model incorporating both pedigree and genome-wide marker data. In addition, they introduced nomenclature for two cross-validation schemes to test models across environments. The first scheme (CV1), predicts genotype performance across all environments, and hence simulates a scenario in which completely new genotypes are introduced to a breeding program for testing. The second (CV2), predicts genotype performance in some, but not all, environments, and hence simulates a scenario in which a genotype has been tested in a portion of the total number of environments included in a trial.

Subsequent studies have further investigated potential gains in GS predictive accuracy enabled by the incorporation of GEI information. One study on the traits leaf width and leaf length in a maize nested association mapping panel found that the CV1 multi-environment cross-validation accuracy was significantly greater than single-environment accuracy, and that CV2 accuracy was greater than CV1 accuracy (Guo et al., 2013). Lado et al. (2016) evaluated GS for grain yield in wheat in a total of 35 environments. They found that incorporating GEI information increased predictive accuracy, especially when predicting line performance within mega-environments. Oakey et al. (2016) likewise found improved predictive accuracy in modeling plant height data in barley when incorporating GEI information. Zhang et al. (2015) studied GS across multiple environments using a biparental maize population, and found that models incorporating GEI outperformed those ignoring GEI, but that the degree of superiority

was dependent upon the complexity and heritability of the trait in question; complex traits such as grain yield benefitted the most from incorporation of GEI, while simpler traits such as anthesis date demonstrated more modest gains in prediction accuracy.

Lopez-Cruz et al. (2015) introduced a model that incorporates marker × environment interactions, and used it to generate predictions for grain yield in wheat across rain-fed and irrigated environments. Their model functions in a manner similar to previous models used to detect quantitative trait loci (QTL) by environment interactions (Moreau et al., 2004). This marker × environment model was further investigated for the traits grain yield, 1000-kernel weight, and heading date in a durum wheat (*Triticum turgidum*) panel grown in four environments (Crossa et al., 2016). The consensus reached by the papers mentioned above is that the incorporation of GEI into GS models can lead to a significant increase in prediction accuracies, though the magnitude of this increase may be affected by the heritability and genetic architecture of the trait being studied.

Multivariate GS models are not limited to modeling GEI; several studies have incorporated multiple traits into GS models. Multiple-trait selection has long been employed to enable selection of low-heritability traits via indirect selection of correlated high-heritability traits, and to prevent excessive divergent response to selection in multiple traits of high importance (Falconer and Mackay, 1996). Multiple-trait analysis was adopted in a GS context to similarly leverage the information between highly correlated traits (Calus and Veerkamp, 2011). Various models for multi-trait GS were subsequently examined using simulated datasets (Guo et al., 2014; Hayashi and Iwata, 2013; Jia and Jannink, 2012). Multi-trait models can be utilized to increase prediction accuracy when data on a particular trait is impossible to measure in some individuals within a population, as is the case with sex-linked traits, or when a trait is simply too

difficult or expensive to measure in all individuals within a population. In monoecious plants, data on a particular trait may technically be measured on all individuals within a population, but there are many examples of traits with measurements that are highly laborious (e.g. the measurement of below-ground biomass or root structure). Several studies have assessed multi-trait genomic selection in datasets consisting of real, non-simulated phenotypic data (dos Santos et al., 2016; Jia and Jannink, 2012; Rutkoski et al., 2012; Schulthess et al., 2016; Wang et al., 2016).

Ideally, one could incorporate data from multiple traits and multiple environments simultaneously, in a unified multi-trait, multi-environment model. At least one multi-trait, multi-environment model has been developed (Montesinos-López et al., 2016). However, this model's Bayesian methods coupled with potentially large numbers of environment/trait combinations make it very computationally demanding.

GS in preliminary yield testing

Preliminary yield tests (PYT), e.g. first year replicated tests conducted over two or more environments, are a common feature of all wheat breeding programs. At this stage, genotypes have undergone enough generations of inbreeding prior to line selection so as to be relatively stable, with limited segregation due to a small number of remaining heterozygous loci. PYT present a few common features detailed below, mostly owing to limited seed and resource availability:

1. A limited number of total testing environments
2. Few or possibly no replications within each environment

3. Unbalanced designs across environments (primarily across years) due to annual selections and advancement

The second point above implies a tradeoff between the number of locations trials are carried out over, and the degree of replication within each location. In the current study, performance of GS in predicting genotype performance was evaluated under the conditions listed above for a variety of quantitative traits with widely varying heritability and genetic architectures. Endelman et al. (2014) examined the question of optimal PYT designs when genome-wide marker data is available by evaluating both genomic prediction accuracy and response to selection for a range of experimental designs. However, their study focused only on grain yield, and utilized a univariate model for genomic prediction in which location was a fixed effect. Lado et al. (2016) studied the use of GS models incorporating GEI data in unbalanced datasets, although they used a relatively large number of environments, and only focused on grain yield. One of the objectives of the current study was to test the utility of multivariate GS methods when testing panels represent more ad-hoc assemblies of genotypes, rather than assemblies of large families of full or half-sibs. To this end, our multivariate models estimate genetic correlation between environments or between traits using the realized relationship matrix calculated from the marker data.

Materials and Methods

Germplasm Selection

A total of 185 soft red winter wheat genotypes were included in each year of the study, with 41 genotypes being tested across both years (i.e. across all environments included in the study), and the rest changing between years. Thus the study included a total of 329 genotypes.

Within each year, genotypes were sourced from breeding programs in Illinois (31), Kentucky (30), Missouri (2), and Virginia (122). A list of all genotypes included in the study is shown in **Table B.1**. Seven genotypes were removed during quality filtering of the genotypic data, leaving 322 genotypes among both years used for further analysis. Five checks were included in the study, including 'Bess', 'Branson', IL00-8250, 'Roane', and 'Shirley'. With the exception of checks and several older cultivars, the majority of genotypes were either F_4 or F_5 filial generation.

Experimental Design and Field Management

Experimental plots were planted in the 2013-14 and 2014-15 winter wheat growing seasons. A generalized randomized complete block design (GRCBD) was utilized across two locations at Kentland Farm near Blacksburg, VA (Guernsey/Hayter silt loams, 37.1965° N, 80.5718° W, 531 m elevation) and the Eastern Virginia Agricultural Research and Extension Center (EVAREC) in Warsaw, VA (Kempsville sandy loam, 37.9879º N, 76.7770º W, 40 m elevation). Two randomized replications were planted at each location.

For the 2013-14 test at Warsaw, and the 2014-15 tests at Blacksburg and Warsaw, each experimental unit consisted of a seven-row plot with a length of 2.74 m, width of 0.91 m, row-spacing of 15.2 cm, and a harvested area of 2.49 m^2. Plots planted in Blacksburg for the 2013-14 crop season were smaller, with a length of 1.98 m, width of 0.91 m, row-spacing of 15.2 cm, and a harvested area of 1.80 m^2. However, at both locations, plot areas were adjusted to 4.18 m^2 to account for inflated yield values caused by border effects. All plots were sown with 70 g of seed. Seed was treated with Raxil® MD fungicide (0.48% tebuconazole/0.64% metalaxyl; Bayer CropScience) at a rate of 2.95 mL a.i. per kg of seed, and Gaucho® 600 flowable insecticide

(48.7% imidacloprid; Bayer CropScience) at a rate of 0.7 mL a.i. per kg of seed. At each location, seed was planted to roughly coincide with the average date of first frost (see **Table A.1**).

At Blacksburg and Warsaw, several tiller counts representative of the test area as a whole were used to calculate ideal nitrogen application rates at Zadok's growth stage 25 (Zadoks et al., 1974) in the spring, and plant tissue tests were used to calculate ideal nitrogen application rates at GS30, per standard regional recommendations from the Virginia Cooperative Extension Service (Alley et al., 1993). All plots at Blacksburg and Warsaw were treated with Palisade® 2EC growth regulator (trinexapac-ethyl; Syngenta Crop Protection) to minimize lodging. In addition, plots in each environment were treated with Tilt® fungicide (propiconazole, Syngenta Crop Protection) throughout the growing season, Prosaro® fungicide (prothioconazole/tebuconazole, Bayer CropScience) near heading date, Harmony® Extra SG herbicide (thifensulfuron-methyl/tribenuron-methyl, DuPont), and Starane® Ultra broadleaf herbicide (fluroxypyr 1-methylheptyl ester , Dow AgroSciences) as needed. The exact dates and rates of chemical applications for each environment included in the study are listed in **Table A.1**.

Phenotyping

Table 2.1 lists the phenotypic traits that were assessed across all environments, with their abbreviations and units of measure. For the 2014-2015 growing season only, seedling emergence was estimated for plots in Blacksburg and Warsaw by averaging the count of seedlings at the two-leaf stage (GS12) from two 0.348 m samples taken from two inner rows. There were no significant differences in seedling numbers between genotypes where seed originated from VA or IL, while the number of seedlings per unit of row length was lower for KY genotypes due to seed

source. Normalized Difference Vegetative Index (NDVI) was measured for each plot at GS25 as described by Phillips et al. (2004) using a Greenseeker® Handheld crop sensor (Trimble® Agriculture, Sunnyvale, CA).

Heading date was recorded as the Julian date at which 50% of plant tillers within a plot had extruded heads from the boot. After plants had reached physiological maturity (GS90), a single 0.914 m cutting of all above-ground plant material was taken from one of the three inner rows of each plot and placed within a paper bag. All cuttings were stored in a sheltered environment for several days to allow for equilibration to ambient moisture levels. Each bag was weighed to derive an estimate of above-ground biomass m^{-1} of row. Subsequently, the number of heads per cutting were counted manually to derive an estimate of heads m^{-2}. Cuttings were then threshed on a plot combine (Wintersteiger NA Inc., Salt Lake City, UT) with settings optimized to recover as much threshed seed as possible. Threshed seed was weighed to derive an estimate of seed weight m^{-1} of row. Harvest index was calculated as the ratio of seed weight to total above-ground biomass. The total number of seeds threshed from each cutting were then counted on a Count-A-Pak optical seed counter (Seedburo® Equipment, Des Plaines, IL) to derive an estimate of grains m^{-2}. Thousand-kernel weight was then calculated as the net weight of the threshed seed sample divided by the number of seeds present * 1,000.

Plant height was averaged from two measurements within each plot, and was recorded as the distance from the soil surface to the tip of the heads (excluding any awns if present). Lodging was measured on a 0 to 9 scale (0 corresponding to no lodging, 9 corresponding to complete lodging). Plots were harvested at maturity using a Wintersteiger plot combine. Moisture content and test weight (grain volume weight) of harvested grain was measured using a GAC® 2500-AGRI grain analysis computer (Dickey-John® Corporation, Auburn, IL). Grain yield was

calculated at 13.5% moisture equivalence.

Grain ash, crude fiber, fat, starch, and protein were estimated via near-infrared (NIR) spectroscopy for subsamples from each plot using an XDS Rapid Content Analyzer (FOSS NIR Systems, Laurel, MD). Fifteen grain samples from each location were sent to Cumberland Valley Analytical Services (Hagerstown, MD) for wet-chemistry analysis of protein, starch and dry matter in order to generate calibration curves for the NIR data.

Genotyping

Genomic DNA was isolated from fresh green leaf tissue using a cetyltrimethylammonium bromide (CTAB) extraction protocol (Saghai-Maroof et al., 1984). Genotyping-by-sequencing was performed at USDA Agricultural Research Service (ARS) facilities using a *PstI-MseI* double digest of genomic DNA. The SNP calling was performed using TASSEL-GBS in TASSEL 5.2.24 (Bradbury et al., 2007; Glaubitz et al., 2014). The Burrows-Wheeler aligner (Li and Durbin, 2009) was used to align SNPs to the International Wheat Genome Sequencing Consortium's whole genome assembly v0.4.

SNP Quality Filtering and Imputation

Prior to imputation of missing genotypes, the genotypic datasets for the 2013-2014 and 2014-2015 material were jointly filtered to remove SNPs with missing data frequencies >20%, heterozygous call frequencies >15%, and minor allele frequency <5%. In addition, all unaligned SNPs were removed. After the initial filtering, missing data in the genotypic dataset was imputed using LinkImpute (Money et al., 2015). LinkImpute implements a nearest-neighbor algorithm using both the *k* nearest individuals and the *l* SNPs in highest LD with the specific missing SNP

genotype that must be imputed. LinkImpute was used with its default settings, which optimize the number of nearest individuals and SNPs via data masking simulations at 10,000 randomly selected genotypes. After imputation, the dataset was once again filtered to remove SNPs with minor allele frequencies <5%. The imputed genotypic dataset was finally filtered in PLINK 1.9 (Chang et al., 2015) to remove all but one SNP in clusters separated by <64bp, as this is the tag size used in the TASSEL-GBS SNP-calling pipeline, (i.e. all SNPs located on the same tag should have the same genotype prior to imputation). In addition to the positional filtering, PLINK was used to remove all but one SNP in groups of SNPs in high LD ($r^2 > 0.8$) using a 250-SNP sliding window, advancing by 10 SNPs with each step.

Phenotypic Modelling and Genomic Selection

A total of four GS models were fit to test various methods of predicting phenotypic performance by utilizing data across multiple environments or traits. The nomenclature for these models largely follows that utilized in Lado et al. (2016):

1. A two-step model utilizing adjusted means (i.e. genotypic best linear unbiased estimates [BLUEs]) across all environments. Genotypic BLUEs were first estimated using the multi-environment mixed model described below. This two-step approach is hereafter referred to as the adjusted means model, and is denoted $GBLUP_M$.
2. A stratified model wherein genomic selection was performed within each environment separately. This model is hereafter referred to as the stratified model, and is denoted $GBLUP_S$.
3. A multivariate model fitting a GEI term, utilizing within-environment means. This model is hereafter referred to as the GEI model, and is denoted $GBLUP_{GE}$.

4. A multivariate model fitting a multi-trait (G×T) interaction term, utilizing the across-environment genotypic BLUEs. This model is hereafter referred to as the multi-trait model, and is denoted $GBLUP_{GT}$

Each location/year combination was considered as a unique environment in order to model phenotypic response across environments. For both the adjusted means model and the multi-trait model, the following mixed effects model was fit using the *lme4* package (Bates et al., 2015) in R:

$$Y_{ijk} = \mu + G_i + E_j + R_k(E_j) + GE_{ij} + \varepsilon_{ijk} \quad (1)$$

Where the phenotypic response (Y_{ijk}) is a function of the overall mean (μ), the ith genotype (G_i), the kth replication (R_k) nested within the jth environment (E_j), the genotype-environment interaction (GE_{ij}) and the residual error (ε_{ijk}). For each trait, variance components for all effects were estimated, and entry-mean heritability (H^2) was calculated. In addition, genotypic best-linear unbiased estimators (BLUEs) were calculated for use as the phenotypic vector for the subsequent GS mixed model described below.

Both the stratified model and the GEI GS models used simple genotypic arithmetic means calculated within each environment. Adjusted means within environments were not utilized as the missing data rate for all traits was extremely low.

For all of the GS models listed above, R package 'rrBLUP' (Endelman, 2011) was used to perform genomic prediction utilizing the genomic best-linear unbiased predictor (GBLUP) method. Briefly, GBLUP solves the following mixed model for u:

$$y = \boldsymbol{X}b + \boldsymbol{Z}u + \varepsilon \quad (2)$$

Where y is a vector ($n \times 1$) of phenotypic observations, b is a vector ($p \times 1$) of fixed effects, u is a vector ($q \times 1$) of random effects, and ε is a vector ($n \times 1$) of the residual variances. $\boldsymbol{X}$ and $\boldsymbol{Z}$ are

incidence matrices that relate the elements of b and u to y, with dimensions ($n \times p$) and ($n \times q$), respectively. Note that in the case of no additional fixed effects being supplied, $\mathbf{X}b$ is equivalent to a ($n \times 1$) vector of genotypic means (μ), such that the equation above is equivalent to:

$$y = \mu + \boldsymbol{Z}u + \varepsilon \quad (3)$$

The variance of u is $\sim N(0, \boldsymbol{G}\sigma_u^2)$, where **G** is a ($n \times n$) matrix of genetic relationships. The additive relationship matrix (**A**) was used to model genetic relationships, and was calculated from the marker data using the method of Endelman and Jannink (2012). The variance of ε is $\sim N(0, \boldsymbol{I}\sigma_\varepsilon^2)$, where **I** is a ($n \times n$) identity matrix.

For the $GLUP_M$ model, y consisted of the genotypic BLUEs calculated across all environments using the model described in equation (1) above. For the $GLUP_S$ model, the mixed model was run once for each environment, and y consisted of the genotypic means within that environment. As the study employed an unbalanced design, this implied recalculating a separate **A** matrix for each environment, using only those genotypes appearing within each environment.

Multi-Environment Genomic Selection

The case of performing genomic selection for a single trait across multiple environments will be presented here; the following section will demonstrate how an equivalent method may be used for performing GS on multiple traits. For the GEI model, y in equation (3) above consisted of combinations of genotypes and environments, such that y becomes a vector of length $g \times e$ where g is the number of genotypes, and e is the number of environments.

For the GEI model, var(u) is defined as the Kronecker product between the **A** matrix and the genetic correlation matrix between environments (**ρ**), such that $u \sim N(0, (\boldsymbol{A} \otimes \boldsymbol{\rho})\sigma_u^2)$, where $\otimes$ represents the Kronecker product. The resulting square matrix has dimensions ([$g \times e$] × [$g \times$

e]). For the $GBLUP_{GE}$ model, genetic correlations between environments were calculated with the R package 'sommer' (Covarrubias-Pazaran, 2016) using a multivariate model as described in equation (3) above, again using the **A** matrix to model additive genetic relationships between genotypes. Due to the unbalanced nature of the data, this model was fit separately for each pair of environments, using all genotypes shared between them. Genetic correlation between each pair of environments was calculated using the product-moment method of Falconer and Mackay (1996):

$$r_A = \frac{cov(x,y)}{\sqrt{\sigma_x^2 \times \sigma_y^2}} \quad (4)$$

Where the genetic correlation (r_A) between environments x and y is a function of the genetic covariance between these two environments [cov(x,y)], and the genetic variances within each environment (σ_x^2 and σ_y^2). This pairwise use of environments created unstructured variance-covariance matrices that were often not positive semi-definite. In these cases, the method of Higham (2002), as implemented in the nearPD() function of R package 'Matrix' (Bates and Maechler, 2016) was used to find a close, approximate, and valid correlation matrix.

Var(ε) for the GEI model is $\sim N(0, (\boldsymbol{R_0} \otimes \boldsymbol{I})\sigma_\varepsilon^2)$, where $\boldsymbol{R_0}$ is a ($e \times e$) matrix of residuals across different environments, with diagonal elements as the residual variance within level *e*, and off-diagonal elements as the reciprocal correlations between different environments. **I** is an identity matrix of dimensions ($g \times g$). Thus the matrix modeling the residual variance has dimensions ([$g \times e$] × [$g \times e$]).

Multi-Trait Genomic Selection

To test the efficacy of the GLUP model when using data on multiple traits, we utilized the traits BIOM, GW, MAT, and YLD, as these traits were all highly intercorrelated, and

exhibited widely differing values for entry mean heritability, from 0.41 for BIOM to 0.92 for MAT. As noted above, the GS model used for multi-trait data was the same as that used for multi-environment data. However, in this case the vector of phenotypic observations, (y in equation [3] above), consisted of all genotype-trait combinations for the traits of interest. In addition in the context of the multi-trait model, var(u) is defined by the Kronecker product between the realized relationship matrix (**A**), and the correlation matrix between traits, which results in a square matrix of dimensions ([$g \times t$] × [$g \times t$]), where t is the number of traits. Var(ε) is defined in a similar manner as described in the section above, however in this case the matrix $\mathbf{R_0}$ is a ($t \times t$) matrix of residuals, and hence the dimensions of the resulting matrix are ([$g \times t$] × [$g \times t$]).

For the $GBLUP_{GT}$ model, the genotypic BLUEs calculated from equation (1) were used to estimate genetic correlations between traits as in equation (4), but considering x and y as measurements of separate traits rather than measurements of the same trait in separate environments. This process was once again performed using package 'sommer' and the **A** matrix to model genetic relationships between genotypes. Unlike the $GLUP_{GE}$ model, for the $GLUP_{GT}$ model, the genetic correlation matrix was estimated for all pairs of traits simultaneously, as the experiment was balanced with respect to trait measurements. Note also that for the $GLUP_{GT}$ model, phenotypic data was standardized within each trait to avoid extreme heteroscedasticity between traits due to different units.

Cross-Validation

Random subset (i.e. Monte Carlo) cross-validation was utilized to assess model prediction accuracy by correlating the GBLUP-generated GEBVs of genotypes in the testing set

with their measured phenotypes. For each model, the cross-validation process was repeated 500 times, randomly dividing the total phenotypic observations into training and validation sets, and results were averaged across the replicates. For the $GLUP_S$ model, this entailed 500 replications of cross-validation for each separate environment.

For both multivariate models ($GLUP_{GE}$ and $GLUP_{GT}$), the two separate cross-validation schemes introduced by Burgueño et al. (2012), CV1 and CV2, were employed. For CV1, genotypes were assigned to either the TP or the VP across all environments (for the $GLUP_{GE}$ model) or across all traits (for the $GLUP_{GT}$ model). For both the $GLUP_{GE}$ and $GLUP_{GT}$ models, nomenclature will be used to designate the model and cross-validation scheme combination, so that, for instance, $GLUP_{GE}$/CV1 will refer to the use of the GEI model with the CV1 cross-validation scheme, while $GLUP_{GT}$/CV2 will refer to the use of the multi-trait model with the CV2 cross-validation scheme.

For the CV2 cross-validation scheme, individual cells from the $g \times l$ vector of phenotypic observations were assigned to the training population, ensuring that for each genotype assigned to the training population, only the phenotypic data from a single trait or environment would be assigned to the training population. A TP/VP percent split of 80/20 was used for all Monte Carlo cross-validations. For the $GLUP_M$ and the $GLUP_{GE}$/CV1 models, this entailed assigning 258 genotypes to the TP, and the remaining 64 to the VP. For the $GLUP_{GE}$/CV2 model, this entailed assigning 581 of the 726 total genotype/environment combinations to the TP, while the remaining 145 observations were assigned to the VP. Finally, ~185 genotypes were present in each environment (with some slight variation due to genotypes that were removed from the analysis), and hence an 80/20 training/validation split in the $GLUP_S$

model entailed assigning ~148 genotypes within each environment to the TP, with the remaining ~37 genotypes being assigned to the VP.

The GEBV-phenotype correlations were calculated both within each level of the interacting variable and across all levels (e.g. within environment and across all environments for the $GBLUP_{GE}$ model). Note that given the unbalanced experimental design employed, the CV2 cross-validation scheme primarily simulated a scenario in which phenotypic data is available for a particular genotype across some locations within a year, but not across years (**Tables 2.3a and 2.3b**). Mean cross-validation GEBV-phenotype correlation was recorded both across all environments and within each environment for all models except $GLUP_S$, where only the within-environment correlations were calculated. For the $GLUP_M$ model, calculating within-environment GEBV-phenotype correlations entailed correlating the model-generated across-environment GEBVs with the genotypic means calculated within each environment.

Note that there are a total of four possible simulation scenarios that arise when GEI information is incorporated into a GS model (Malosetti et al., 2013). These are the aforementioned CV1 and CV2, in addition to a scenario in which values for untested environments are predicted using tested genotypes (Heslot et al., 2013), as well as the "hardest" simulation, in which GEBVs are generated for untested genotypes in untested environments (CV4). Due to the limited number of environments included in the study, the latter two simulations were not considered.

Training Population Size

For the $GBLUP_M$ and $GLUP_{GE}$ models, the proportion of observations assigned to the TP and VP were varied to test the effects of varying training population size, using TP/VP percent

splits of 80/20, 60/40, 40/60, and 20/80. For each TP/VP ratio, 500 replications of Monte Carlo cross-validation were run, using both the CV1 and CV2 schemes for the $GLUP_{GE}$ model. Once again, GEBV-phenotype correlations within and across environments were recorded. Note that as the TP/VP ratio decreases, the CV2 cross-validation scheme used for the $GLUP_{GE}$ model becomes more similar to the CV1 cross-validation scheme.

Results

General Line Performance and Correlation of Traits

The variance components and entry-mean heritability across all four environments included in the study are shown in **Table 2.2**. For reference, trait names and their abbreviations are shown in **Table 2.1**. Entry mean heritability ranged from 0.41 for BIOM to 0.94 for HD and TKW. Heritability for YLD was slightly higher than expected at 0.83. **Figure 2.1** depicts the pairwise phenotypic and genetic correlation coefficients between all traits included in the study. YLD resides within a cluster of highly interrelated traits, all positively correlated with each other. This includes the relatively high-heritability phenological traits HD and MAT, as well as the lower-heritability traits GW and BIOM. The trait WCPROT was notable for its general negative correlations with traits within this group, in particular YLD, which is a finding that has been well-noted in many studies in the past (Cox et al., 1985; Groos et al., 2003; Terman et al., 1969). The strongest negative correlation (both phenotypic and genetic) existed between the traits TKW and GSQM, reflecting a general tradeoff between grain size and grain number per unit area due to trait compensation. WCPROT and STARCH also demonstrated a strong negative correlation. The genetic correlations between traits calculated using the realized relationship matrix (**A**) generally closely mirrored the corresponding phenotypic correlations. The

magnitudes of the genetic correlations were typically equal to or greater than their corresponding phenotypic correlations, though there were exceptions, for instance with the relationship between GW and BIOM.

Genomic Prediction Accuracy Across Environments

TP/VP percent splits of 80/20 were used to compare the performance of the $GLUP_M$ model against that of the $GLUP_{GE}$ model across all four environments included in the study. Cross-validation for the $GLUP_{GE}$ model was performed using both the CV1 and CV2 schemes. Generating prediction accuracy estimates across all environments tests a model's ability to predict overall genotype performance within a breeding program's set of environments. **Table 2.4** lists each trait's mean across-environment GEBV-phenotype correlations and standard errors calculated with the $GBLUP_M$ model and the CV1 and CV2 cross-validation schemes of the $GLUP_{GE}$ model, as well as the percent differences between the $GLUP_M$ correlations and the corresponding $GLUP_{GE}$ correlations. **Figure 2.2** presents regressions of the mean GEBV-phenotype correlations produced by the $GBLUP_M$ model against those produced by each cross-validation scheme of the $GLUP_{GE}$ model. Predictive accuracies generated by the $GLUP_M$ model and both cross-validation strategies of the $GBLUP_{GE}$ model varied widely across traits. In general, a trait's heritability only weakly explained the accuracy of GEBVs generated by any of the three model/cross-validation combinations, with r^2 values for regressions between trait heritability and mean GEBV-phenotype correlations ranging between 0.35 and 0.45 for the $GLUP_M$, $GLUP_{GE}$/CV1, and $GELUP_{GE}$/CV2 models (data not shown). However, the lowest-heritability trait (BIOM; $H^2 = 0.41$) did consistently have the lowest GEBV-phenotype correlations across all models. One of the highest heritability traits (TKW; $H^2 = 0.94$) also

produced the highest GEBV-phenotype correlations in the $GLUP_M$ model, though this was not the case in the $GLUP_{GE}$ model, where the highest predictive accuracies were observed for STARCH when using either cross-validation scheme.

Predictive accuracies for several traits were notably different between the $GLUP_M$ and $GLUP_{GE}$ models. These included YLD (with the $GLUP_{GE}$ model outperforming the $GLUP_M$ model by 98% and 64% for the CV1 and CV2 cross-validation schemes, respectively), STARCH (with the $GLUP_{GE}$ model outperforming the $GLUP_M$ model by 51% and 59% for the CV1 and CV2 cross-validation schemes, respectively), FLSG (with the $GLUP_{GE}$ model underperforming the $GLUP_M$ model by 28% and 44% for the CV1 and CV2 cross-validation schemes, respectively), and SSQM (with the $GLUP_{GE}$ model underperforming the $GLUP_M$ model by 37% and 59% for the CV1 and CV2 cross-validation schemes, respectively). Several of the highest-heritability traits (HT and TKW) had roughly equivalent predictive accuracies in both the $GLUP_M$ model and both cross-validation schemes of the $GLUP_{GE}$ model. However, other high-heritability traits, including HD, MAT, and TWT, exhibited more variable performance across models.

Genomic Prediction Accuracy Within Environments

The $GLUP_S$ model consistently underperformed both the $GLUP_M$ model and the $GLUP_{GE}$ model within environments. However, it should be noted that this is partially due to the unbalanced design of the experiment, as the training populations available to the $GLUP_S$ model within a given environment were much smaller than those available to either the $GLUP_M$ or $GBLUP_{GE}$ models. Were a completely balanced design across environments used, it is not known how the $GLUP_S$ model would perform in comparison to the $GBLUP_{GE}$ model.

The results produced by the $GLUP_{GE}$ model exhibited the effects of Simpson's Paradox, as GEBV-phenotype correlations within environments did not necessarily reflect correlations across environments, and vice-versa. The $GLUP_{GE}$/CV2 model generally outperformed all other models within environments, including the $GBLUP_{GE}$/CV1 model. In contrast, the $GLUP_{GE}$/CV1 model generally performed equivalently to the $GLUP_{M}$ model, though there were some traits for which the $GLUP_{GE}$/CV1 model consistently performed better or worse than the $GLUP_{M}$ model. Interestingly, there were several traits which exhibited high across-environment GEBV-phenotype correlations in the $GLUP_{GE}$/CV1 model, while simultaneously exhibiting several very low correlations within one or more environments. This was the case for YLD, for which the $GLUP_{GE}$/CV1 model approximately doubled GEBV-phenotype correlations produced by the $GLUP_{M}$. However, mean correlations within the 2014 Blacksburg environment were roughly half those produced by the $GLUP_{M}$ model for the same environment. This is somewhat surprising, as this environment appeared to cluster more closely with the majority of environments measured for YLD in the GGE analysis, while the 2015 Warsaw environment exhibited more distant similarity to the other environments.

Multi-Trait Genomic Prediction

In the context of a multi-trait model, the CV1 cross-validation scheme simulates a scenario in which a target genotype's performance across all traits is predicted using across-trait data from the training population. In contrast, the CV2 cross-validation scheme simulates a scenario in which data on some traits but not others is collected for a particular genotype, and hence a genotype's data for one or more traits is used to predict its values for one or more traits for which data was not collected. In this study, across-trait GEBV values were 0.33 ± 0.0029 for

the $GLUP_{GT}$ model utilizing CV1 cross-validation, and 0.63 ± 0.0012 when utilizing the CV2 cross-validation. As shown in **Figure 2.4**, the within-trait GEBV-phenotype correlations calculated with the $GLUP_M$ model and the $GLUP_{GT}$/CV1 model tended to be quite similar. However, the within-trait correlations calculated using the $GLUP_{GT}$/CV2 model tended to be far higher. As opposed to the $GBLUP_{GE}$ model, the within-trait and across-trait GEBV-phenotype correlations tended to be closely related for the $GLUP_{GT}$ model, with across-trait correlations tending towards the mean of each of the individual within-trait correlations. Notably, the mean across-trait correlation calculated using CV1 cross-validation was higher than the within-trait correlations for the traits BIOM and GW (whether these were calculated with the $GLUP_M$ model, or the $GLUP_{GT}$/CV1 model). In addition, the across-trait correlation calculated using the $GLUP_{GT}$/CV2 was higher than the within-trait correlations for BIOM and MAT.

Effects of Training Population Size

To test the effects of training-population size on generated GEBV-phenotype correlations, the TP/VP ratios used for the $GLUP_M$ and $GLUP_{GE}$ models were varied from an 80/20 split up to a 20/80 split. **Table 2.6** presents the across-environment results of this experiment for the $GLUP_M$ model as well as the $GLUP_{GE}$ model using the CV1 and CV2 cross-validations, while **Figure 2.5** presents these results for selected traits. In general, decreases in predictive accuracy caused by decreases in training population size were highly consistent for the $GLUP_M$ model; i.e. traits tended to retain their relative accuracy rankings with varying training population size, with relatively little crossover between traits. In contrast, varying training population size produced more unpredictable changes in GEBV-phenotype correlations for the $GLUP_{GE}$ model utilizing CV1 cross-validation, and especially for the $GLUP_{GE}$ model utilizing CV2 cross-validation. Notably, the mean GEBV-phenotype correlation for the trait BIOM

calculated using the $GLUP_{GE}$ model and CV2 cross-validation decreased from 0.07 for an 80/20 TP:VP ratio, down to precisely 0 for a 20/80 ratio. In contrast to the across-environment prediction accuracies, the within-environment mean GEBV-phenotype correlations for all models and cross-validation schemes tended to vary in a predictable fashion as TP size was varied. Results for within-environment accuracies are presented in **Table E.1**, and the results for select traits are shown in **Figure 2.6**. Interestingly, for many traits, the within-environment correlation accuracy for the $GLUP_{GE}$/CV2 model seemed to plateau as the TP/VP ratio was increased from 60/40 to 80/20.

Discussion

Overall, the results of this study suggest various scenarios in preliminary wheat breeding trials for which multivariate GS models are and are not well suited. Several previous studies reported that multi-trait GS models could be used to increase predictive accuracy for low-heritability traits that are highly correlated with auxiliary, higher-heritability traits (Jia and Jannink, 2012; Schulthess et al., 2016; Wang et al., 2016). In the present study, the predictive accuracy of the $GLUP_{GT}$ model was far greater than that of the $GLUP_{M}$ model when using the CV2 cross-validation scheme, but not when using the CV1 cross-validation scheme (**Figure 2.4**). However, it should be noted that for the multi-trait model used herein, the CV2 cross-validation entails a scenario in which data on a particular trait is only collected in a portion of the total genotypes, but this data is consistently collected on these genotypes across all environments in which they are tested. Nevertheless, the $GLUP_{GT}$/CV2 model performed well despite the use of an otherwise highly unbalanced design. Testing of a still sparser model, in which data for a

particular trait is only collected for some genotypes in some of the environments in which they are grown, was not attempted in this study.

For either the CV1 or CV2 cross-validations, all traits could be used together to form predictions, making the multi-trait model somewhat akin to a selection index. In this case, accuracies were roughly equivalent to the means of the predictive accuracies of the individual traits involved (as calculated using the $GLUP_M$ model), meaning that the across-trait correlation accuracies tended to be higher than the within-trait accuracies for low-heritability traits, and lower than the within-trait accuracies for high-heritability traits. Overall, the data suggest that while the prediction of multiple trait values for newly introduced genotypes is not especially accurate, sparse data collection coupled with genomic prediction could be a viable option to decrease phenotyping costs in the field.

In contrast to the $GLUP_{GT}$ model, the across-environment and within-environment correlations generated by the $GLUP_{GE}$ model did not closely reflect one another. This was likely due to the lack of standardization between environments, as opposed to the $GLUP_{GT}$ model, which used standardized trait values. Across-environment GEBV-phenotype correlations were roughly equivalent for the CV1 and CV2 cross-validations, though occasionally worse for the CV2 cross-validation (**Table 2.4**). This is in contrast to some previous findings in which correlations were generally higher for CV2 (Burgueño et al., 2012; Crossa et al., 2016; Guo et al., 2014; Lopez-Cruz et al., 2015; Zhang et al., 2015). However, in this study the CV2 cross-validation scheme consistently yielded better within-environment predictions than the CV1 scheme. Thus the prediction of genotype performance within environments when a genotype has already been phenotyped in at least some environments is less challenging than generating predictions of genotype performance in the absence of any phenotypic information. The GEBV-

phenotype correlations generated by the $GLUP_{GE}$ model for some traits were consistently higher or lower than the corresponding correlations produced by the $GLUP_{M}$ model, whether the CV1 or CV2 schemes were used. The reason for this will be explained in detail below.

The variable TP size experiment indicated that further gains in accuracy for the $GLUP_{M}$ and $GLUP_{GE}$ models could be realized by further increases in TP size, as the GEBV-phenotype correlations for the majority of traits did not appear to plateau with an 80/20 TP:VP split (**Figure 2.5**). In general, the within-environment GEBV-phenotype correlations varied in a more predictable manner as the training population size was varied (**Figure 2.6**). As previously noted, the correlations generated by the $GLUP_{GE}$/CV2 model appear to plateau for many traits at a 60/40 training/validation split. This suggests that relatively sparse data can be used to generate within-environment phenotypic predictions if genotypes are tested in some environments but not others, though across-environment prediction accuracy would likely suffer as a result.

As previously mentioned, several traits exhibited widely differing predictive accuracies in the $GLUP_{M}$ model vs. the $GLUP_{GE}$ model. These included YLD and STARCH (for which the $GLUP_{GE}$ model predictive accuracy was far higher than that of the $GLUP_{M}$ model), and SSQM and FLSG (for which the opposite was true). A trait's heritability only explained a relatively small portion of this differential performance among traits. It appears that a trait's patterns of GEI are much more important in determining its suitability to the $GLUP_{GE}$ model than its heritability and, by extension, overall magnitude of GEI. Finlay and Wilkinson (1963) proposed the regression of genotype trait values against an environmental index as a method of quantifying trait stability, and this method is still widely utilized today to examine GEI patterns among genotypes. In this method, the phenotypic values of genotypes across environments are compared to the mean response of all genotypes included in the trial. **Figure 2.7a** shows such a

genotype-by-environment interaction plot for the trait YLD (a trait for which the $GLUP_{GE}$ model far outperformed the $GLUP_M$ model), using the 41 genotypes that were tested across all environments. **Figure 2.7b** shows the same plot generated for the trait SSQM, for which the $GLUP_{GE}$ model underperformed the $GLUP_M$ model. Both traits have identical heritability (0.83) and therefore similar overall magnitudes of GEI across the tested environments. Examining the YLD response across environments shown in **Figure 2.7a**, it is clear that nearly all of the plotted genotypes are closely tracking the mean response across environments; i.e. while there is clearly some degree of overall GEI present, GEI patterns among genotypes are highly consistent. Eberhart and Russel (1966) labelled this pattern of GEI as "dynamic stability". In contrast, the SSQM response across environments shown in **Figure 7b** demonstrates a case in which GEI patterns among genotypes are inconsistent, with a mixture of genotypes that are highly responsive to environmental influence (highly sloped lines), and those that are highly stable across environments (nearly flat lines, exhibiting "static stability" in Eberhart and Russell's nomenclature). The overall GEI pattern shown for the trait SSQM exhibits a number of "crossover" or "rank-change" type interactions between genotypes. As the $GLUP_{GE}$ model estimates correlation between environments using the full set of genotypes shared between each pair of environments, these highly variable GEI patterns can have very adverse effects on prediction accuracies.

This has several practical implications for the implementation of GS models in multi-environment breeding trials. First, GS models incorporating GEI effects appear to be most suited to moderate-heritability traits with consistent patterns of GEI. Checking the patterns of GEI exhibited by the training population is crucial for ensuring that the incorporation of GEI effects into a GS model will not lead to worse predictions compared to using a strategy that ignores GEI

(e.g. performing GS using the adjusted means across environments). For some higher-heritability traits, GS models incorporating GEI effects are likely to offer little additional benefit over the two-step model incorporating adjusted means (Zhang et al., 2015). However, this was not the case for all high heritability traits included in this study, as some did exhibit marginally higher predictive accuracy when using the $GLUP_{GE}$ model (e.g. HD and TWT).

This leaves the question of potential methods to utilize to ensure that predictions incorporating GEI data are in fact trustworthy in the face of highly heterogeneous crossover patterns. One simple though somewhat naïve strategy for generating more reliable estimates of GEI would be to simply increase the number of testing environments. However, this strategy is often not feasible due to resource limitations (as in the present study). In addition, such a "blind" approach may inadvertently incorporate highly dissimilar environments that exhibit crossover GEI during model training, thereby reducing model predictive ability.

An alternative strategy may be to simply utilize a stratified analysis, running a separate GS analysis within each environment. Lopez-Cruz et al. (2015) found that a stratified model often yielded prediction accuracies comparable to those of a model including GEI effects, especially for a CV1 cross-validation scheme. However, in experiments with highly unbalanced designs, as was the case with the present study, this strategy is likely to yield quite low predictive accuracies due to the necessarily smaller training populations that are formed within each environment. In addition, if many environments are included, this method would entail running a correspondingly large number of independent models, and the combining of these results would utilize some form of post-hoc GEI analysis, making this an inelegant solution.

A more nuanced strategy would be to select genotypes and/or environments to find subsets of each that exhibit more homogeneous GEI patterns. The present study utilized GGE

biplots to assess relationships among environments, though selection of environments was not carried out due to the overall low number of environments. Several linear-bilinear models have also been employed as a method of identifying and clustering genotypes or environments based upon GEI, including the GGE biplot method that was used in this study. Linear-bilinear models that treat genotype and environment as fixed include the additive main effects and multiplicative interaction (AMMI) model (Gauch, 1988; Gauch and Zobel, 1997), the shifted multiplicative model (Cornelius et al., 1992) or the sites regression model (Cornelius et al., 1993; Crossa and Cornelius, 1997). Factor analytic (FA) mixed models are linear-bilinear models that can treat genotype, environment, or both as random effects (Piepho, 1998, 1997). Burgueño et al. (2008) later demonstrated how to use FA models to identify clusters of environments and genotypes exhibiting negligible crossover-type GEI. The identification of mega-environments prior to GS modelling has been performed in multiple studies (Burgueño et al., 2012; Lado et al., 2016; Lopez-Cruz et al., 2015). Heslot et al. (2013) characterized environments in a more explicit manner by using GS models to identify and remove less-predictive environments from the set of training environments. Hoffstetter et al. (2016) performed a GS study in which they selected genotypes exhibiting low GEI for inclusion in the training population, but found that this practice had little effect on predictive accuracy for most traits, and a detrimental effect for YLD. However, it should be noted that this study utilized a two-step adjusted means model.

However, tailoring a training population to minimize crossover GEI is of little use if an uncharacterized environment or genotype is in reality not very closely related to the environments or genotypes used for training. More empirical studies are required to determine how accurately GEI can be predicted for genotypes or environments using genetic relationships or environmental covariables, respectively. Several recent studies have examined GS for the

prediction of trait stability values. Wang et al. (2015) generated both static and dynamic stability estimates for six traits in two rye (*Secale cereale*) populations using the methods of Eberhart and Russel (1966). They then performed genomic selection and found it to be relatively useful in predicting genotype stability, though this varied between traits, stability parameters, and populations. Huang et al. (2016) performed genomic selection on several different agronomic and quality-related traits utilizing four different models and dynamic stability estimates generated via both AMMI models and Eberhart and Russell regression. They found that GS produced useful predictions of trait stability, with accuracies ranging from 0.33 to 0.67. In addition, they found that the majority of traits studied did not have a high correlation with their corresponding stability parameters. While more work is needed to assess the potential ability of GS to predict trait stability parameters, one can envision a scenario in which a two-step method is utilized to first characterize genotypes based upon their stability, followed by the generation of genomic predictions for the associated trait in a multi-environment model using training genotypes that are predicted to exhibit similar patterns of GEI. However, such efforts would likely require large numbers of genotypes and environments to ensure that identified genotype or environment clusters contain adequate numbers of observations for training purposes.

Conclusion

This study found that multivariate genomic selection models could be useful for increasing prediction accuracies over univariate models in a variety of settings. However, in the case of multivariate models incorporating GEI effects, careful attention must be paid to the patterns of GEI exhibited by genotypes within the training population; the presence of significant crossover GEI patterns among genotypes in the training or validation populations can lead to

predictive accuracies that are worse than those generated by a simpler two-step univariate model utilizing across-environment adjusted means. In general, multi-environment GS models showed the greatest gains over univariate models when used on moderate-heritability traits that exhibit relatively homogeneous GEI patterns. Multi-trait models generally performed equivalently to univariate models when predicting the performance of genotypes that lacked data on any trait, but exhibited much higher predictive ability if data on some traits were available for each genotype. Methods of clustering genotypes and/or environments used for GS model training deserve closer inspection, as the use of genotypes and environments exhibiting highly homogeneous GEI patterns can lead to much higher predictive accuracy.

References

Alley, M.M., Brann, D.E., Stromberg, E.L., Hagood, E.S., Herbert, A., Jones, E.C., Griffith, W.K., 1993. Intensive Soft Red Winter Wheat Production: A Management Guide.

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

Bates, D., Mächler, M., Bolker, B., Walker, S., 2015. Fitting Linear Mixed-Effects Models Using lme4. J. Stat. Softw. 67. doi:10.18637/jss.v067.i01

Bates, D., Maechler, M., 2016. Package "Matrix": Sparse and Dense Matrix Classes and Methods.

Benson, J., Brown-Guedira, G., Paul Murphy, J., Sneller, C., 2012. Population Structure, Linkage Disequilibrium, and Genetic Diversity in Soft Winter Wheat Enriched for Fusarium Head Blight Resistance. Plant Genome 5, 71–80. doi:10.3835/plantgenome2011.11.0027

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

Bradbury, P.J., Zhang, Z., Kroon, D.E., Casstevens, T.M., Ramdoss, Y., Buckler, E.S., 2007. TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23, 2633–2635. doi:10.1093/bioinformatics/btm308

Burgueño, J., Crossa, J., Cornelius, P.L., Yang, R.-C., 2008. Using Factor Analytic Models for Joining Environments and Genotypes without Crossover Genotype × Environment Interaction. Crop Sci. 48, 1291–1305. doi:10.2135/cropsci2007.11.0632

Burgueño, J., de los Campos, G., Weigel, K., Crossa, J., 2012. Genomic Prediction of Breeding Values when Modeling Genotype × Environment Interaction using Pedigree and Dense Molecular Markers. Crop Sci. 52, 707–719. doi:10.2135/cropsci2011.06.0299

Calus, M.P., Veerkamp, R.F., 2011. Accuracy of multi-trait genomic selection using different methods. Genet. Sel. Evol. 43, 26. doi:10.1186/1297-9686-43-26

Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., Lee, J.J., 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience 4, 7. doi:10.1186/s13742-015-0047-8

Cornelius, P., Crossa, J., Seyedsadr, M., 1993. Tests and estimators of multiplicative models for variety trials, in: 5th Annual Conference on Applied Statistics in Agriculture. Presented at the 5th Annual Conference on Applied Statistics in Agriculture, Manhattan, KS.

Cornelius, P.L., Seyedsadr, M., Crossa, J., 1992. Using the shifted multiplicative model to search for "separability" in crop cultivar trials. Theor. Appl. Genet. 84, 161–172. doi:10.1007/BF00223996

Covarrubias-Pazaran, G., 2016. Genome-Assisted Prediction of Quantitative Traits Using the R Package sommer. PLOS ONE 11, e0156744. doi:10.1371/journal.pone.0156744

Cox, M.C., Qualset, C.O., Rains, D.W., 1985. Genetic Variation for Nitrogen Assimilation and Translocation in Wheat. I. Dry Matter and Nitrogen Accumulation. Crop Sci. 25, 430–435. doi:10.2135/cropsci1985.0011183X002500030002x

Crossa, J., Cornelius, P.L., 1997. Sites Regression and Shifted Multiplicative Model Clustering of Cultivar Trial Sites under Heterogeneity of Error Variances. Crop Sci. 37, 406–415. doi:10.2135/cropsci1997.0011183X003700020017x

Crossa, J., de los Campos, G., Maccaferri, M., Tuberosa, R., Burgueño, J., Pérez-Rodríguez, P., 2016. Extending the Marker × Environment Interaction Model for Genomic-Enabled Prediction and Genome-Wide Association Analysis in Durum Wheat. Crop Sci. 56, 2193–2209. doi:10.2135/cropsci2015.04.0260

dos Santos, J.P.R., Vasconcellos, R.C. de C., Pires, L.P.M., Balestre, M., Pinho, R.G.V., 2016. Inclusion of Dominance Effects in the Multivariate GBLUP Model. PLOS ONE 11, e0152045. doi:10.1371/journal.pone.0152045

Eberhart, S.A., Russell, W.A., 1966. Stability Parameters for Comparing Varieties. Crop Sci 6, 36–40. doi:10.2135/cropsci1966.0011183X000600010011x

Endelman, J.B., 2011. Ridge Regression and Other Kernels for Genomic Selection with R Package rrBLUP. Plant Gen 4, 250–255. doi:10.3835/plantgenome2011.08.0024

Endelman, J.B., Atlin, G.N., Beyene, Y., Semagn, K., Zhang, X., Sorrells, M.E., Jannink, J.-L., 2014. Optimal Design of Preliminary Yield Trials with Genome-Wide Markers. Crop Sci. 54, 48–59. doi:10.2135/cropsci2013.03.0154

Endelman, J.B., Jannink, J.-L., 2012. Shrinkage Estimation of the Realized Relationship Matrix. G3 GenesGenomesGenetics 2, 1405–1413. doi:10.1534/g3.112.004259

Falconer, D.S., Mackay, T.F.C., 1996. Introduction to quantitative genetics. Longman, Essex, England.

Finlay, K., Wilkinson, G., 1963. The analysis of adaptation in a plant-breeding programme. Crop Pasture Sci. 14, 742–754.

Fisher, R.A., 1918. The Correlation Between Relatives on the Supposition of Mendelian Inheritance. Trans. R. Soc. Edinb. 52, 399–433.

Gauch, H.G., 1988. Model Selection and Validation for Yield Trials with Interaction. Biometrics 44, 705–715. doi:10.2307/2531585

Gauch, H.G., Zobel, R.W., 1997. Identifying Mega-Environments and Targeting Genotypes. Crop Sci. 37, 311–326. doi:10.2135/cropsci1997.0011183X003700020002x

Gianola, D., 2013. Priors in Whole-Genome Regression: The Bayesian Alphabet Returns. Genetics 194, 573–596. doi:10.1534/genetics.113.151753

Gianola, D., Weigel, K.A., Krämer, N., Stella, A., Schön, C.-C., 2014. Enhancing Genome-Enabled Prediction by Bagging Genomic BLUP. PLOS ONE 9, e91693. doi:10.1371/journal.pone.0091693

Glaubitz, J.C., Casstevens, T.M., Lu, F., Harriman, J., Elshire, R.J., Sun, Q., Buckler, E.S., 2014. TASSEL-GBS: A High Capacity Genotyping by Sequencing Analysis Pipeline. PLOS ONE 9, e90346. doi:10.1371/journal.pone.0090346

Groos, C., Robert, N., Bervas, E., Charmet, G., 2003. Genetic analysis of grain protein-content, grain yield and thousand-kernel weight in bread wheat. Theor. Appl. Genet. 106, 1032–1040. doi:10.1007/s00122-002-1111-1

Guo, G., Zhao, F., Wang, Y., Zhang, Y., Du, L., Su, G., 2014. Comparison of single-trait and multiple-trait genomic prediction models. BMC Genet. 15, 30. doi:10.1186/1471-2156-15-30

Guo, Z., Tucker, D.M., Wang, D., Basten, C.J., Ersoz, E., Briggs, W.H., Lu, J., Li, M., Gay, G., 2013. Accuracy of Across-Environment Genome-Wide Prediction in Maize Nested Association Mapping Populations. G3 GenesGenomesGenetics 3, 263–272. doi:10.1534/g3.112.005066

Habier, D., Fernando, R.L., Dekkers, J.C.M., 2007. The Impact of Genetic Relationship Information on Genome-Assisted Breeding Values. Genetics 177, 2389–2397. doi:10.1534/genetics.107.081190

Hayashi, T., Iwata, H., 2013. A Bayesian method and its variational approximation for prediction of genomic breeding values in multiple traits. BMC Bioinformatics 14, 34. doi:10.1186/1471-2105-14-34

Hayes, B.J., Visscher, P.M., Goddard, M.E., 2009. Increased accuracy of artificial selection by using the realized relationship matrix. Genet. Res. 91, 47–60. doi:10.1017/S0016672308009981

Henderson, C.R., 1975. Best Linear Unbiased Estimation and Prediction under a Selection Model. Biometrics 31, 423–447. doi:10.2307/2529430

Henderson, C.R., Quaas, R.L., 1976. Multiple Trait Evaluation Using Relatives' Records. J. Anim. Sci. 43, 1188–1197. doi:10.2527/jas1976.4361188x

Heslot, N., Akdemir, D., Sorrells, M.E., Jannink, J.-L., 2013. Integrating environmental covariates and crop modeling into the genomic selection framework to predict genotype by environment interactions. Theor. Appl. Genet. 127, 463–480. doi:10.1007/s00122-013-2231-5

Heslot, N., Yang, H.-P., Sorrells, M.E., Jannink, J.-L., 2012. Genomic Selection in Plant Breeding: A Comparison of Models. Crop Sci 52, 146–160. doi:10.2135/cropsci2011.06.0297

Higham, N.J., 2002. Computing the nearest correlation matrix—a problem from finance. IMA J. Numer. Anal. 22, 329–343. doi:10.1093/imanum/22.3.329

Hoffstetter, A., Cabrera, A., Huang, M., Sneller, C., 2016. Optimizing Training Population Data and Validation of Genomic Selection for Economic Traits in Soft Winter Wheat. G3 Genes Genomes Genet. 6, 2919–2928. doi:10.1534/g3.116.032532

Huang, M., Cabrera, A., Hoffstetter, A., Griffey, C., Sanford, D.V., Costa, J., McKendry, A., Chao, S., Sneller, C., 2016. Genomic selection for wheat traits and trait stability. Theor. Appl. Genet. 129, 1697–1710. doi:10.1007/s00122-016-2733-z

Jia, Y., Jannink, J.-L., 2012. Multiple-trait genomic selection methods increase genetic value prediction accuracy. Genetics 192, 1513–1522. doi:10.1534/genetics.112.144246

Lado, B., Barrios, P.G., Quincke, M., Silva, P., Gutiérrez, L., 2016. Modeling Genotype × Environment Interaction for Genomic Selection with Unbalanced Data from a Wheat Breeding Program. Crop Sci. 56, 2165–2179. doi:10.2135/cropsci2015.04.0207

Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25, 1754–1760. doi:10.1093/bioinformatics/btp324

Lopez-Cruz, M., Crossa, J., Bonnett, D., Dreisigacker, S., Poland, J., Jannink, J.-L., Singh, R.P., Autrique, E., Campos, G. de los, 2015. Increased Prediction Accuracy in Wheat Breeding Trials Using a Marker × Environment Interaction Genomic Selection Model. G3 GenesGenomesGenetics 5, 569–582. doi:10.1534/g3.114.016097

Malosetti, M., Ribaut, J.-M., van Eeuwijk, F.A., 2013. The statistical analysis of multi-environment data: modeling genotype-by-environment interaction and its genetic basis. Front. Physiol. 4. doi:10.3389/fphys.2013.00044

Meuwissen, T.H., 2009. Accuracy of breeding values of "unrelated" individuals predicted by dense SNP genotyping. Genet. Sel. Evol. 41, 35. doi:10.1186/1297-9686-41-35

Meuwissen, T.H.E., Hayes, B.J., Goddard, M.E., 2001. Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. Genetics 157, 1819–1829.

Money, D., Gardner, K., Migicovsky, Z., Schwaninger, H., Zhong, G.-Y., Myles, S., 2015. LinkImpute: Fast and Accurate Genotype Imputation for Nonmodel Organisms. G3 GenesGenomesGenetics 5, 2383–2390. doi:10.1534/g3.115.021667

Montesinos-López, O.A., Montesinos-López, A., Crossa, J., Toledo, F., Pérez-Hernández, O., Eskridge, K.M., Rutkoski, J., 2016. A Genomic Bayesian Multi-trait and Multi-environment Model. G3 GenesGenomesGenetics g3.116.032359. doi:10.1534/g3.116.032359

Moreau, L., Charcosset, A., Gallais, A., 2004. Use of trial clustering to study QTL x environment effects for grain yield and related traits in maize. Theor. Appl. Genet. 110, 92–105. doi:10.1007/s00122-004-1781-y

Morton, N.E., 1991. Parameters of the human genome. Proc. Natl. Acad. Sci. U. S. A. 88, 7474–7476.

Oakey, H., Cullis, B., Thompson, R., Comadran, J., Halpin, C., Waugh, R., 2016. Genomic Selection in Multi-environment Crop Trials. G3 GenesGenomesGenetics g3.116.027524. doi:10.1534/g3.116.027524

Phillips, S.B., Keahey, D.A., Warren, J.G., Mullins, G.L., 2004. Estimating Winter Wheat Tiller Density Using Spectral Reflectance Sensors for Early-Spring, Variable-Rate Nitrogen Applications. Agron. J. 96, 591. doi:10.2134/agronj2004.0591

Piepho, H.-P., 1998. Empirical best linear unbiased prediction in cultivar trials using factor-analytic variance-covariance structures. Theor. Appl. Genet. 97, 195–201. doi:10.1007/s001220050885

Piepho, H.-P., 1997. Analyzing Genotype-Environment Data by Mixed Models with Multiplicative Terms. Biometrics 53, 761–766. doi:10.2307/2533976

Piepho, H.P., Möhring, J., Melchinger, A.E., Büchse, A., 2008. BLUP for phenotypic selection in plant breeding and variety testing. Euphytica 161, 209–228. doi:10.1007/s10681-007-9449-8

R Core Team, 2015. R: A language and environment for statistical computing.

Rutkoski, J., Benson, J., Jia, Y., Brown-Guedira, G., Jannink, J.-L., Sorrells, M., 2012. Evaluation of Genomic Prediction Methods for Fusarium Head Blight Resistance in Wheat. Plant Gen 5, 51–61. doi:10.3835/plantgenome2012.02.0001

Saghai-Maroof, M.A., Soliman, K.M., Jorgensen, R.A., Allard, R.W., 1984. Ribosomal DNA spacer-length polymorphisms in barley: mendelian inheritance, chromosomal location, and population dynamics. Proc. Natl. Acad. Sci. U. S. A. 81, 8014–8018.

Schulthess, A.W., Wang, Y., Miedaner, T., Wilde, P., Reif, J.C., Zhao, Y., 2016. Multiple-trait- and selection indices-genomic predictions for grain yield and protein content in rye for feeding purposes. Theor. Appl. Genet. 129, 273–287. doi:10.1007/s00122-015-2626-6

Terman, G.L., Ramig, R.E., Dreier, A.F., Olson, R.A., 1969. Yield-Protein Relationships in Wheat Grain, as Affected by Nitrogen and Water. Agron. J. 61, 755–759. doi:10.2134/agronj1969.00021962006100050031x

Tibshirani, R., 1994. Regression Shrinkage and Selection Via the Lasso. J. R. Stat. Soc. Ser. B 58, 267–288.

Tikhonov, A., 1963. Solution of Incorrectly Formulated Problems and the Regularization Method. Sov. Math. - Dokl. 5, 1035–1038.

VanRaden, P.M., 2008. Efficient methods to compute genomic predictions. J. Dairy Sci. 91, 4414–4423. doi:10.3168/jds.2007-0980

VanRaden, P.M., Van Tassell, C.P., Wiggans, G.R., Sonstegard, T.S., Schnabel, R.D., Taylor, J.F., Schenkel, F.S., 2009. Invited Review: Reliability of genomic predictions for North American Holstein bulls. J. Dairy Sci. 92, 16–24. doi:10.3168/jds.2008-1514

Wang, X., Li, L., Yang, Z., Zheng, X., Yu, S., Xu, C., Hu, Z., 2016. Predicting rice hybrid performance using univariate and multivariate GBLUP models based on North Carolina mating design II. Heredity 118, 302–310. doi:10.1038/hdy.2016.87

Wang, Y., Mette, M.F., Miedaner, T., Wilde, P., Reif, J.C., Zhao, Y., 2015. First insights into the genotype–phenotype map of phenotypic stability in rye. J. Exp. Bot. 66, 3275–3284. doi:10.1093/jxb/erv145

Whittaker, J.C., Thompson, R., Denham, M.C., 2000. Marker-assisted selection using ridge regression. Genet. Res. 75, 249–252. doi:10.1111/j.1469-1809.1999.ahg634_0351_17.x

Wimmer, V., Lehermeier, C., Albrecht, T., Auinger, H.-J., Wang, Y., Schön, C.-C., 2013. Genome-Wide Prediction of Traits with Different Genetic Architecture Through Efficient Variable Selection. Genetics 195, 573–587. doi:10.1534/genetics.113.150078

Wright, K., Laffont, J.-L., 2016. Package "gge": Genotype Plus Genotype-by-Environment Biplots.

Yan, W., 2001. GGEbiplot—A Windows Application for Graphical Analysis of Multienvironment Trial Data and Other Types of Two-Way Data. Agron. J. 93, 1111–1118. doi:10.2134/agronj2001.9351111x

Yan, W., Hunt, L.A., Sheng, Q., Szlavnics, Z., 2000. Cultivar Evaluation and Mega-Environment Investigation Based on the GGE Biplot. Crop Sci. 40, 597–605. doi:10.2135/cropsci2000.403597x

Yan, W., Kang, M.S., 2003. GGE biplot analysis a graphical tool for breeders, geneticists, and agronomists. CRC Press, Boca Raton, Fla.

Zadoks, J.C., Chang, T.T., Konzak, C.F., 1974. A decimal code for the growth stages of cereals. Weed Res. 14, 415–421. doi:10.1111/j.1365-3180.1974.tb01084.x

Zhang, X., Pérez-Rodríguez, P., Semagn, K., Beyene, Y., Babu, R., López-Cruz, M.A., San Vicente, F., Olsen, M., Buckler, E., Jannink, J.-L., Prasanna, B.M., Crossa, J., 2015. Genomic prediction in biparental tropical maize populations in water-stressed and well-watered environments using low-density and GBS SNPs. Heredity 114, 291–299. doi:10.1038/hdy.2014.99

Zou, H., Hastie, T., 2005. Regularization and variable selection via the elastic net. J. R. Stat. Soc. Ser. B Stat. Methodol. 67, 301–320. doi:10.1111/j.1467-9868.2005.00503.x

Table 2.1: Phenotypic traits assessed in Blacksburg and Warsaw, VA during the 2013-2014 and 2014-2015 growing seasons

Trait	Abbreviation	Units	Trait Ontology †	Ontology Description
Heading Date	HD	Julian days (Jan1)	TO:0000137	Days required for 50% of heads to emerge from boot
Physiological Maturity	MAT	Julian days (Jan1)	TO:0000469	Days required for 50% of peduncles to lose green color
Flag Leaf Stay Green	FLSG	days	TO:0000249	Days between heading and flag leaf senescence
NDVI at Zadok's GS25	NDVI	-	CO_321:0000301	Normalized-difference vegetation index measured at spring green-up using a Trimble Greenseeker instrument
Mature Plant Height	HT	cm	TO:0000207	Height of plant from soil surface to tip of spike excluding awns
Above-ground biomass at maturity	BIOM	g dwt m^{-1} row	CO_321:0000229	Weight of dry plant matter cut at soil level from a 1 m section of a center row of the plot
Grain weight	GW	g dwt m^{-1} row	TO:0000589	Grain yield (g dry weight) from 1 m row cut at physiological maturity
Harvest Index	HI	-	TO:0000128	Grain yield divided by above-ground biomass
Grains per square meter	GSQM	Grains m^{-2}	CO_321:0000017	Number of grains threshed from 1m row cutting, converted to square meters
Spikes per square meter	SSQM	Spikes m^{-2}	CO_321:0000166	Number of spikes (fertile culms) per unit area, sample or plant
Seeds per Head	SPH	count	TO:0002759	Number of grains within an inflorescence
Thousand Kernel Weight	TKW	grams	TO:0000382	Seed weight estimated by weighing 1000 seeds
Test Weight	TWT	g L^{-1}	TO:0000612	Weight per unit volume of grain at standard moisture level
Yield, 13.5% moisture	YLD	kg ha^{-1}	TO:0000396	Grain yield standardized to 13.5% moisture equivalence
NIR Whole grain starch	STARCH	%	-	-
NIR Wet Chemistry-validated whole grain protein	WCPROT	%	-	-

† Matching trait ontologies in the Planteome database: http://browser.planteome.org/amigo/search/ontology

Table 2.2: Trait descriptive statistics, variance components, and entry-mean heritability for lines grown in Blacksburg, VA and Warsaw, VA for the 2013-2014 and 2014-2015 growing seasons

		Descriptive Statistics				Variances and Heritability				
Trait †	**Units**	**min**	**mean**	**max**	**SD**	**σ^2 G**	**σ^2 E**	**σ^2 (G x E)**	**σ^2 ε**	**H^2**
HD	Julian days (Jan1)	121	128	136	3.21	2.77	8.88	0.5	0.46	0.94
MAT	Julian days (Jan1)	151	159.3	171	4.72	2.68	23.52	0.48	0.93	0.92
FLSG	days	21	28.69	38	2.76	1.44	4.28	0.74	1.75	0.78
NDVI	-	0.26	0.54	0.75	0.08	0.0004	0.005	0.0002	0.001	0.67
HT	cm	59.69	85.43	119.4	9.29	21.27	73.03	3.47	7.9	0.92
BIOM	g dwt m^{-1} row	122.5	225.7	350	35.54	67.87	519.3	5.43E-12	777.5	0.41
GW	g dwt m^{-1} row	47.68	96.64	157.9	16.53	33.19	108	2.38E-11	152.5	0.64
HI	-	0.28	0.43	0.53	0.032	0.0005	6.50E-05	0.0001	0.0004	0.86
GSQM	Grains m^{-2}	8460	1.85E+04	3.13E+04	3277	3.17E+06	2.42E+06	3.21E+05	5.35E+06	0.81
SSQM	Spikes m^{-2}	459.3	853	1485	161.2	7005	1.05E+04	101.6	1.12E+04	0.83
SPH	count	8.54	21.94	33.29	3.04	4.92	1.29	0.75	2.64	0.90
TKW	grams	24.1	34.57	91.6	3.98	9.33	3.48	0.65	3.09	0.94
TWT	g L^{-1}	652.6	759	810.9	19.7	139.5	236.5	34.03	27.41	0.92
YLD	kg ha^{-1}	3579	6627	9053	1027	1.24E+05	9.75E+05	2.84E+04	1.49E+05	0.83
STARCH	%	46.88	52.51	56.49	1.41	0.29	1.41	0.15	0.37	0.78
WCPROT	%	9.67	12.34	16.04	1.01	0.21	0.51	0.09	0.35	0.76

σ^2 G genotypic variance; σ^2 E environmental variance; σ^2 (G x E) genotype x environment variance; σ^2 ε residual variance; H^2 entry-mean heritability.

† HD heading date; MAT physiological maturity date; FLSG flag leaf stay green; NDVI normalized-difference vegetation index at Zadok's GS25; HT plant height; BIOM above-ground biomass; GW grain weight; HI harvest index; GSQM grains per square meter; SSQM spikes per square meter; SPH seeds per head; TKW thousand kernel weight; TWT test weight; YLD grain yield; STARCH whole-grain starch content; WCPROT wet chemistry-validated whole-grain protein content

Table 2.3a: Example of two cross-validation schemes (CV1 and CV2) used for GEI model. NA's denote trait/environment combinations that were not tested. Dots represent values set to missing. Note that in this example, genotypes 5 and 6 were tested across all environments. Numeric values are within-environment means.

	CV1				CV2			
genotype	Env1	Env2	Env3	Env4	Env1	Env2	Env3	Env4
1	6.9	1.1	NA	NA	6.9	.	NA	NA
2	4.4	6.9	NA	NA	.	6.9	NA	NA
3	.	.	NA	NA	-0.1	1.4	NA	NA
4	4.5	4.7	NA	NA	4.5	.	NA	NA
5	.	.	.	.	-1.1	-1.9	.	-0.8
6	6.4	3.1	2.0	0.9	6.4	3.1	0.3	-1.1
…	..	…	…	…	..	…	…	…
319	NA	NA	8.2	5.7	NA	NA	8.2	5.7
320	NA	NA	.	.	NA	NA	1.4	.
321	NA	NA	3.0	2.4	NA	NA	.	2.4
322	NA	NA	6.0	4.5	NA	NA	6.0	.
323	NA	NA	.	.	NA	NA	.	-1.1
324	NA	NA	6.3	3.2	NA	NA	6.3	3.2

Table 2.3b: Example of two cross-validation schemes (CV1 and CV2) used for multi-trait model. Dots represent values set to missing. Numeric values are standardized across-environment adjusted means.

	CV1				CV2			
genotype	Trait1	Trait2	Trait3	Trait4	Trait1	Trait2	Trait3	Trait4
1	-0.8	-0.5	0.8	1.5	.	0.4	-1.3	0.1
2	-1.2	-0.3	-1.7	-1.3	-0.2	-1.5	.	-1.0
3	.	.	.	.	.	-0.4	-0.6	1.3
4	-0.1	0.3	0.5	0.2	0.6	-0.6	-0.7	.
5	.	.	.	.	-0.7	.	1.7	2.0
6	0.4	0.7	1.2	-1.1	-0.4	1.5	.	-0.4
…	..	…	…	…	..	…	…	…
317	-0.5	0.8	0.5	0.8	2.3	-0.8	1.4	-0.7
318	-0.5	-1.2	1.0	-0.7	.	-1.0	0.9	-1.0
319	.	.	.	.	.	-0.5	-0.2	0.7
320	.	.	.	.	-2.2	0.2	1.8	1.5
321	-0.8	1.2	-0.4	-0.2	0.1	.	-1.8	-1.1
322	0.4	-0.3	0.2	-0.2	0.3	1.1	.	0.1

Table 2.4: Mean GEBV-phenotype correlations and their standard errors* across all environments included in the study for the adjusted means model and the GEI model.

	$GBLUP_M$ ‡		$GBLUP_{GE}$ CV1 §		$GBLUP_{GE}$ CV2 §		% Difference ¶	
Trait †	mean	SE	mean	SE	mean	SE	CV1	CV2
HD	0.53	0.0036	0.65	0.0019	0.60	0.0020	24.7	14.6
MAT	0.46	0.0038	0.49	0.0018	0.35	0.0025	5.7	-24.4
FLSG	0.55	0.0034	0.40	0.0024	0.31	0.0027	-27.6	-44.1
NDVI	0.40	0.0041	0.47	0.0026	0.34	0.0027	16.1	-15.8
HT	0.57	0.0041	0.60	0.0019	0.57	0.0019	5.3	-0.4
BIOM	0.27	0.0048	0.21	0.0025	0.07	0.0031	-22.5	-74.1
GW	0.31	0.0043	0.43	0.0026	0.36	0.0028	38.4	16.2
HI	0.43	0.0045	0.39	0.0035	0.58	0.0024	-8.1	35.8
GSQM	0.35	0.0044	0.44	0.0033	0.46	0.0022	27.0	33.5
SSQM	0.49	0.0036	0.31	0.0030	0.20	0.0031	-37.5	-58.8
SPH	0.36	0.0041	0.43	0.0037	0.69	0.0016	19.6	94.2
TKW	0.66	0.0031	0.61	0.0031	0.66	0.0025	-7.1	0.8
TWT	0.58	0.0034	0.69	0.0019	0.68	0.0015	18.9	17.2
YLD	0.35	0.0039	0.68	0.0020	0.57	0.0026	97.7	63.5
STARCH	0.46	0.0040	0.70	0.0022	0.74	0.0016	51.2	59.0
WCPROT	0.36	0.0041	0.47	0.0030	0.53	0.0022	32.3	48.2

* Means and standard errors were calculated across 500 replications of Monte-Carlo cross validation, with 80% of observations used for training, and the remaining 20% used for validation

† HD heading date; MAT physiological maturity date; FLSG flag leaf stay green; NDVI normalized-difference vegetation index at Zadok's GS25; HT plant height; BIOM above-ground biomass; GW grain weight; HI harvest index; GSQM grains per square meter; SSQM spikes per square meter; SPH seeds per head; TKW thousand kernel weight; TWT test weight; YLD grain yield; STARCH whole-grain starch content; WCPROT wet chemistry-validated whole-grain protein content

‡ Model using the adjusted means across environments

§ Models incorporating GEI effects and utilizing either the CV1 or CV2 cross-validation schemes

¶ Percent difference in mean GEBV-phenotype correlation between the adjusted means model and the GEI models utilizing the CV1 or CV2 cross-validation schemes

Table 2.5: Mean GEBV-phenotype correlations and their standard errors* within environments included in the study for the adjusted means model, the stratified model, and the GEI model utilizing the CV1 and CV2 cross-validation schemes.

		14Bb ‡		14War		15Bb		15War	
Model	Trait †	mean	SE	mean	SE	mean	SE	mean	SE
adjusted means	HD	0.31	0.0060	0.44	0.0057	0.63	0.0040	0.60	0.0041
	MAT	0.33	0.0062	0.34	0.0059	0.52	0.0050	0.50	0.0048
	FLSG	0.22	0.0069	0.51	0.0051	0.52	0.0049	0.48	0.0052
	NDVI	0.30	0.0062	0.37	0.0053	0.35	0.0059	0.26	0.0068
	HT	0.46	0.0057	0.40	0.0068	0.63	0.0049	0.55	0.0056
	BIOM	0.12	0.0062	0.16	0.0066	0.27	0.0068	0.27	0.0066
	GW	0.14	0.0062	0.14	0.0066	0.34	0.0062	0.31	0.0057
	HI	0.45	0.0068	0.40	0.0065	0.38	0.0057	0.20	0.0062
	GSQM	0.46	0.0056	0.29	0.0070	0.15	0.0064	0.35	0.0060
	SSQM	0.45	0.0050	0.43	0.0059	0.39	0.0055	0.47	0.0054
	SPH	0.35	0.0061	0.32	0.0059	0.35	0.0056	0.36	0.0060
	TKW	0.63	0.0049	0.69	0.0043	0.62	0.0045	0.65	0.0044
	TWT	0.55	0.0056	0.45	0.0064	0.57	0.0047	0.56	0.0045
	YLD	0.26	0.0064	0.23	0.0065	0.33	0.0056	0.41	0.0051
	STARCH	0.44	0.0056	0.42	0.0052	0.30	0.0062	0.32	0.0061
	WCPROT	0.22	0.0070	0.17	0.0063	0.29	0.0060	0.34	0.0061
stratified	HD	0.21	0.0036	0.33	0.0034	0.43	0.0031	0.40	0.0031
	MAT	0.15	0.0037	0.19	0.0036	0.35	0.0029	0.31	0.0032
	FLSG	0.14	0.0036	0.40	0.0031	0.39	0.0029	0.35	0.0032
	NDVI	0.18	0.0039	0.22	0.0037	0.21	0.0033	0.18	0.0035
	HT	0.37	0.0030	0.33	0.0037	0.51	0.0027	0.38	0.0038
	BIOM	0.03	0.0029	0.05	0.0040	0.17	0.0039	0.16	0.0036
	GW	0.11	0.0036	0.04	0.0036	0.31	0.0029	0.26	0.0035
	HI	0.40	0.0035	0.27	0.0037	0.42	0.0026	0.17	0.0034
	GSQM	0.26	0.0037	0.10	0.0034	0.03	0.0030	0.16	0.0037
	SSQM	0.16	0.0036	0.19	0.0039	0.15	0.0035	0.21	0.0035
	SPH	0.21	0.0036	0.11	0.0035	0.24	0.0033	0.15	0.0033
	TKW	0.36	0.0038	0.45	0.0035	0.43	0.0025	0.45	0.0031
	TWT	0.48	0.0033	0.40	0.0032	0.46	0.0024	0.40	0.0031
	YLD	0.23	0.0037	0.17	0.0037	0.28	0.0031	0.37	0.0026
	STARCH	0.31	0.0035	0.29	0.0034	0.42	0.0025	0.34	0.0031
	WCPROT	0.15	0.0039	0.05	0.0029	0.22	0.0036	0.27	0.0032
	HD	0.20	0.0064	0.42	0.0057	0.52	0.0045	0.57	0.0042
	MAT	0.13	0.0066	0.14	0.0065	0.39	0.0057	0.36	0.0063
	FLSG	0.10	0.0069	0.53	0.0052	0.53	0.0044	0.48	0.0053
	NDVI	0.25	0.0062	0.38	0.0063	0.11	0.0069	0.23	0.0065
	HT	0.43	0.0055	0.41	0.0061	0.52	0.0048	0.42	0.0058

GEI CV1	BIOM	0.09	0.0067	0.00	0.0070	0.25	0.0065	0.21	0.0062
	GW	0.18	0.0068	0.15	0.0067	0.33	0.0061	0.30	0.0052
	HI	0.42	0.0066	0.41	0.0057	0.49	0.0047	0.24	0.0061
	GSQM	0.47	0.0059	0.29	0.0068	0.17	0.0073	0.37	0.0068
	SSQM	0.38	0.0059	0.40	0.0058	0.33	0.0056	0.42	0.0057
	SPH	0.45	0.0059	0.35	0.0060	0.35	0.0058	0.39	0.0067
	TKW	0.63	0.0051	0.68	0.0045	0.61	0.0048	0.66	0.0046
	TWT	0.55	0.0059	0.55	0.0053	0.51	0.0054	0.56	0.0043
	YLD	-0.01	0.0069	0.33	0.0060	0.32	0.0062	0.39	0.0057
	STARCH	0.42	0.0064	0.42	0.0050	0.15	0.0069	0.37	0.0053
	WCPROT	-0.01	0.0068	0.00	0.0066	0.38	0.0060	0.23	0.0059
GEI CV2	HD	0.55	0.0045	0.58	0.0043	0.75	0.0032	0.83	0.0018
	MAT	0.32	0.0059	0.33	0.0055	0.52	0.0049	0.54	0.0049
	FLSG	0.24	0.0070	0.58	0.0048	0.60	0.0039	0.55	0.0047
	NDVI	0.31	0.0056	0.41	0.0057	0.17	0.0072	0.27	0.0062
	HT	0.50	0.0051	0.56	0.0052	0.71	0.0037	0.60	0.0044
	BIOM	0.22	0.0066	0.14	0.0068	0.28	0.0064	0.25	0.0057
	GW	0.30	0.0061	0.23	0.0063	0.42	0.0056	0.37	0.0052
	HI	0.70	0.0046	0.67	0.0050	0.64	0.0040	0.46	0.0051
	GSQM	0.68	0.0037	0.55	0.0050	0.35	0.0060	0.50	0.0057
	SSQM	0.58	0.0048	0.57	0.0047	0.47	0.0054	0.56	0.0044
	SPH	0.81	0.0036	0.77	0.0038	0.67	0.0033	0.71	0.0032
	TKW	0.89	0.0018	0.93	0.0011	0.78	0.0048	0.80	0.0039
	TWT	0.76	0.0033	0.78	0.0028	0.73	0.0031	0.77	0.0028
	YLD	0.19	0.0075	0.33	0.0065	0.44	0.0059	0.49	0.0049
	STARCH	0.59	0.0055	0.63	0.0044	0.27	0.0069	0.35	0.0063
	WCPROT	0.24	0.0064	0.30	0.0065	0.49	0.0051	0.32	0.0056

* Means and standard errors were calculated across 500 replications of Monte-Carlo cross validation, with 80% of observations used for training, and the remaining 20% used for validation

† HD heading date; MAT physiological maturity date; FLSG flag leaf stay green; NDVI normalized-difference vegetation index at Zadok's GS25; HT plant height; BIOM above-ground biomass; GW grain weight; HI harvest index; GSQM grains per square meter; SSQM spikes per square meter; SPH seeds per head; TKW thousand kernel weight; TWT test weight; YLD grain yield; STARCH whole-grain starch content; WCPROT wet chemistry-validated whole-grain protein content

‡ 14Bb Blacksburg, VA 2014; 14War Warsaw, VA 2014; 15Bb Blacksburg, VA 2015; 15War Warsaw, VA 2015

Table 2.6: Mean across-environment GEBV-phenotype correlations and their standard errors* across varying training/validation set proportions, from 80% training/20% validation to 20% training/80% validation for the adjusted means model and the GEI model utilizing the CV1 and CV2 cross-validation schemes.

		80/20 ‡		60/40		40/60		20/80	
model	Trait †	mean	SE	mean	SE	mean	SE	mean	SE
adjusted means	HD	0.53	0.0036	0.50	0.0024	0.46	0.0020	0.39	0.0024
	MAT	0.46	0.0038	0.42	0.0025	0.37	0.0020	0.30	0.0025
	FLSG	0.55	0.0034	0.53	0.0023	0.50	0.0017	0.45	0.0019
	NDVI	0.40	0.0041	0.38	0.0026	0.35	0.0021	0.28	0.0027
	HT	0.57	0.0041	0.54	0.0024	0.51	0.0018	0.44	0.0023
	BIOM	0.27	0.0048	0.26	0.0028	0.25	0.0022	0.22	0.0024
	GW	0.31	0.0043	0.30	0.0027	0.28	0.0021	0.25	0.0024
	HI	0.43	0.0045	0.42	0.0026	0.39	0.0020	0.33	0.0026
	GSQM	0.35	0.0044	0.30	0.0029	0.25	0.0023	0.16	0.0030
	SSQM	0.49	0.0036	0.45	0.0026	0.38	0.0022	0.28	0.0027
	SPH	0.36	0.0041	0.29	0.0027	0.21	0.0023	0.14	0.0026
	TKW	0.66	0.0031	0.62	0.0022	0.57	0.0018	0.48	0.0023
	TWT	0.58	0.0034	0.55	0.0022	0.51	0.0019	0.44	0.0021
	YLD	0.35	0.0039	0.33	0.0026	0.29	0.0021	0.23	0.0025
	STARCH	0.46	0.0040	0.44	0.0026	0.41	0.0021	0.34	0.0026
	WCPROT	0.36	0.0041	0.33	0.0026	0.30	0.0020	0.24	0.0025
GEI CV1	HD	0.65	0.0019	0.63	0.0012	0.59	0.0010	0.51	0.0017
	MAT	0.49	0.0018	0.48	0.0012	0.46	0.0009	0.41	0.0014
	FLSG	0.40	0.0024	0.38	0.0016	0.35	0.0012	0.29	0.0014
	NDVI	0.47	0.0026	0.42	0.0016	0.36	0.0013	0.29	0.0017
	HT	0.60	0.0019	0.59	0.0012	0.57	0.0010	0.51	0.0014
	BIOM	0.21	0.0025	0.19	0.0018	0.16	0.0017	0.12	0.0018
	GW	0.43	0.0026	0.41	0.0016	0.36	0.0013	0.28	0.0015
	HI	0.39	0.0035	0.38	0.0023	0.35	0.0016	0.29	0.0023
	GSQM	0.44	0.0033	0.38	0.0024	0.30	0.0024	0.19	0.0024
	SSQM	0.31	0.0030	0.28	0.0020	0.23	0.0017	0.18	0.0017
	SPH	0.43	0.0037	0.38	0.0025	0.31	0.0022	0.22	0.0023
	TKW	0.61	0.0031	0.57	0.0022	0.52	0.0019	0.43	0.0021
	TWT	0.69	0.0019	0.64	0.0013	0.59	0.0012	0.51	0.0015
	YLD	0.68	0.0020	0.66	0.0012	0.60	0.0013	0.42	0.0024
	STARCH	0.70	0.0022	0.67	0.0015	0.62	0.0016	0.49	0.0024
	WCPROT	0.47	0.0030	0.46	0.0018	0.43	0.0015	0.34	0.0022
	HD	0.60	0.0020	0.55	0.0012	0.53	0.0010	0.39	0.0023
	MAT	0.35	0.0025	0.29	0.0017	0.33	0.0013	0.23	0.0024
	FLSG	0.31	0.0027	0.25	0.0018	0.25	0.0015	0.19	0.0018
	NDVI	0.34	0.0027	0.25	0.0021	0.23	0.0018	0.13	0.0027

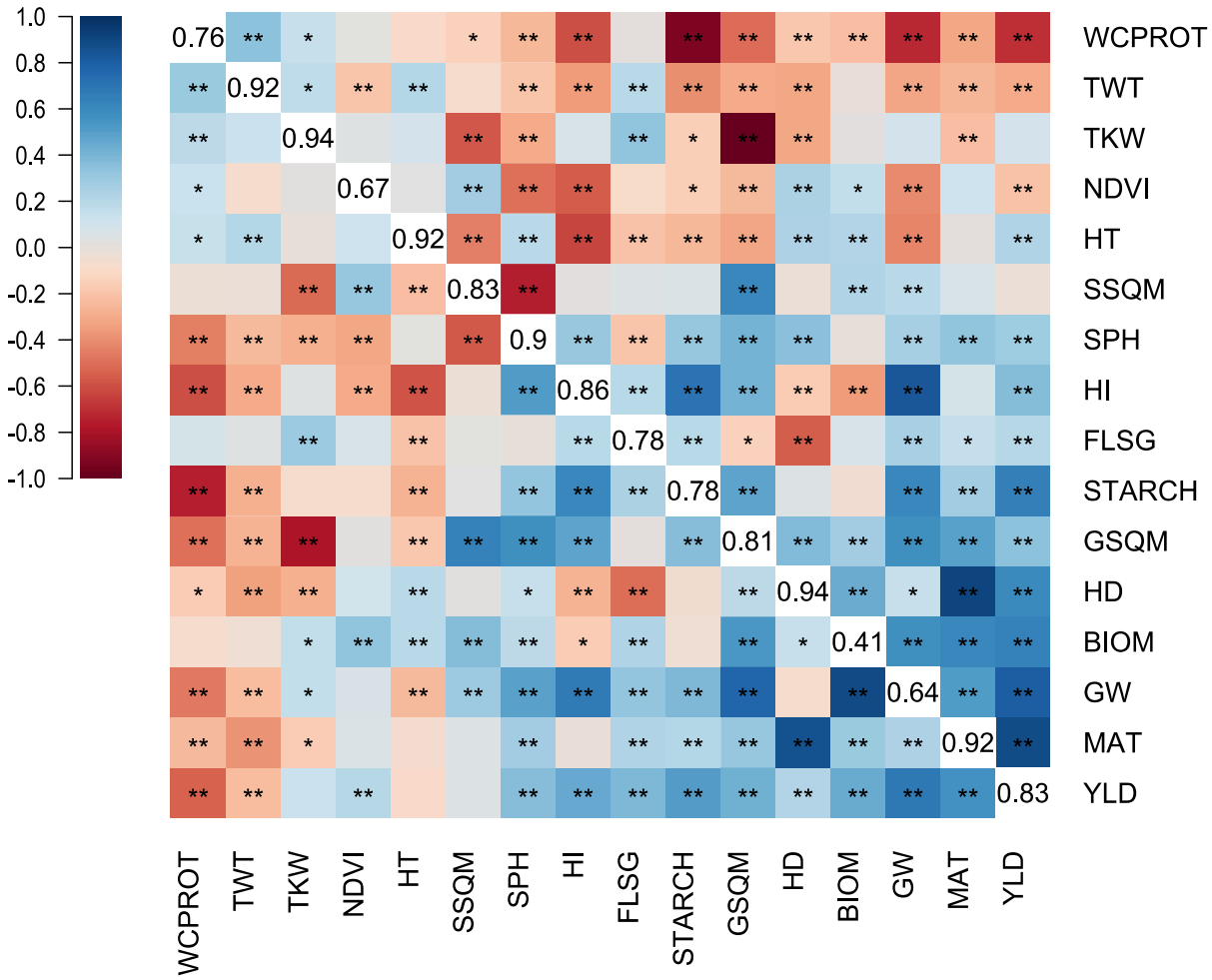
GEI CV2	HT	0.57	0.0019	0.54	0.0011	0.54	0.0010	0.46	0.0015
	BIOM	0.07	0.0031	0.01	0.0020	0.02	0.0018	0.00	0.0019
	GW	0.36	0.0028	0.29	0.0020	0.26	0.0016	0.18	0.0016
	HI	0.58	0.0024	0.57	0.0011	0.48	0.0012	0.36	0.0018
	GSQM	0.46	0.0022	0.40	0.0012	0.31	0.0014	0.19	0.0018
	SSQM	0.20	0.0031	0.17	0.0016	0.18	0.0014	0.19	0.0014
	SPH	0.69	0.0016	0.65	0.0008	0.52	0.0014	0.37	0.0016
	TKW	0.66	0.0025	0.63	0.0012	0.57	0.0010	0.47	0.0015
	TWT	0.68	0.0015	0.61	0.0010	0.56	0.0010	0.45	0.0019
	YLD	0.57	0.0026	0.42	0.0021	0.32	0.0020	0.14	0.0019
	STARCH	0.74	0.0016	0.68	0.0012	0.58	0.0016	0.32	0.0026
	WCPROT	0.53	0.0022	0.51	0.0012	0.46	0.0011	0.25	0.0026

* Means and standard errors were calculated across 500 replications of Monte-Carlo cross validation

† HD heading date; MAT physiological maturity date; FLSG flag leaf stay green; NDVI normalized-difference vegetation index at Zadok's GS25; HT plant height; BIOM above-ground biomass; GW grain weight; HI harvest index; GSQM grains per square meter; SSQM spikes per square meter; SPH seeds per head; TKW thousand kernel weight; TWT test weight; YLD grain yield; STARCH whole-grain starch content; WCPROT wet chemistry-validated whole-grain protein content

‡ Training/validation set ratio, expressed as % training / % validation

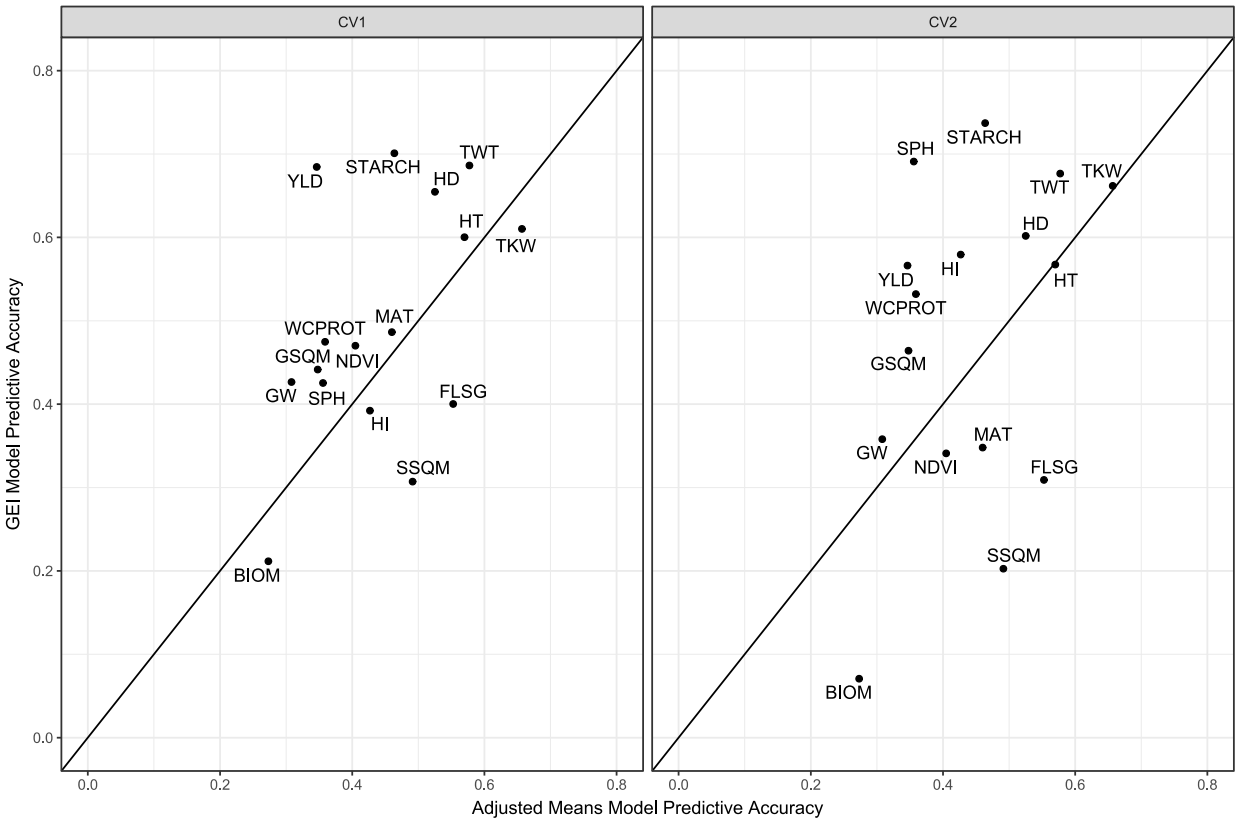
Figure 2.1: Correlations among traits calculated from the genotypic BLUEs. Phenotypic correlations are below the diagonal; genetic correlations are above. Numbers in the diagonal are each trait's entry-mean heritability calculated across all environments



WCPROT wet chemistry-validated whole-grain protein content; TWT test weight; TKW thousand kernel weight; NDVI normalized-difference vegetation index at Zadok's GS25; HT plant height; SSQM spikes per square meter; SPH seeds per head; HI harvest index; FLSG flag leaf stay green; STARCH whole-grain starch content; GSQM grains per square meter; HD heading date; BIOM above-ground biomass; GW grain weight; MAT physiological maturity date; YLD grain yield

* significant at the 0.05 level; ** significant at the 0.01 level

Figure 2.2: Regression of mean across-environment GEBV-phenotype correlations between the adjusted means model and the GEI model using the CV1 and CV2 cross-validation schemes using the 80%/20% training/validation split



HD heading date; MAT physiological maturity date; FLSG flag leaf stay green; NDVI normalized-difference vegetation index at Zadok's GS25; HT plant height; BIOM above-ground biomass; GW grain weight; HI harvest index; GSQM grains per square meter; SSQM spikes per square meter; SPH seeds per head; TKW thousand kernel weight; TWT test weight; YLD grain yield; STARCH whole-grain starch content; WCPROT wet chemistry-validated whole-grain protein content

Figure 2.3: Difference in accuracy prediction between the adjusted-means model, the stratified model, and the GEI model utilizing the CV1 and CV2 cross-validation schemes. Numeric values represent the percent difference between the listed model, environment, and trait combination and the corresponding environment, trait combination as evaluated by the model utilizing adjusted means.

		$GBLUP_S$ ‡				$GBLUP_{GE}$ CV1 §				$GBLUP_{GE}$ CV2 §			
Trait †	H^2	14Bb ¶	14War	15Bb	15War	14Bb	14War	15Bb	15War	14Bb	14War	15Bb	15War
HD	0.94	-31	-26	-31	-33	-35	-6	-17	-4	78	32	20	39
MAT	0.92	-54	-43	-33	-39	-60	-58	-26	-28	-3	-3	0	8
FLSG	0.78	-37	-21	-25	-29	-56	5	2	0	12	14	15	14
NDVI	0.67	-41	-40	-40	-29	-17	3	-68	-13	1	12	-53	4
HT	0.92	-18	-19	-19	-31	-6	1	-17	-24	9	39	12	9
BIOM	0.41	-77	-70	-35	-41	-25	-100	-4	-22	89	-16	5	-5
GW	0.64	-18	-70	-8	-17	30	2	-3	-2	123	62	22	19
HI	0.86	-11	-31	10	-14	-6	3	27	21	54	70	68	133
GSQM	0.81	-45	-67	-83	-55	2	1	9	5	47	89	129	42
SSQM	0.83	-65	-56	-61	-56	-15	-7	-15	-11	28	33	21	18
SPH	0.9	-39	-67	-33	-58	29	8	0	7	130	140	91	96
TKW	0.94	-44	-35	-32	-31	-1	-2	-2	1	41	35	26	22
TWT	0.92	-13	-12	-19	-27	-1	21	-9	1	37	73	30	38
YLD	0.83	-8	-26	-15	-10	-102	46	-4	-5	-24	48	33	17
STARCH	0.78	-29	-32	41	7	-5	0	-50	14	33	48	-11	9
WCPROT	0.76	-31	-73	-25	-21	-103	-100	30	-34	9	72	67	-7

Legend: 150, 125, 100, 75, 50, 25, 0, -25, -50, -75, -100

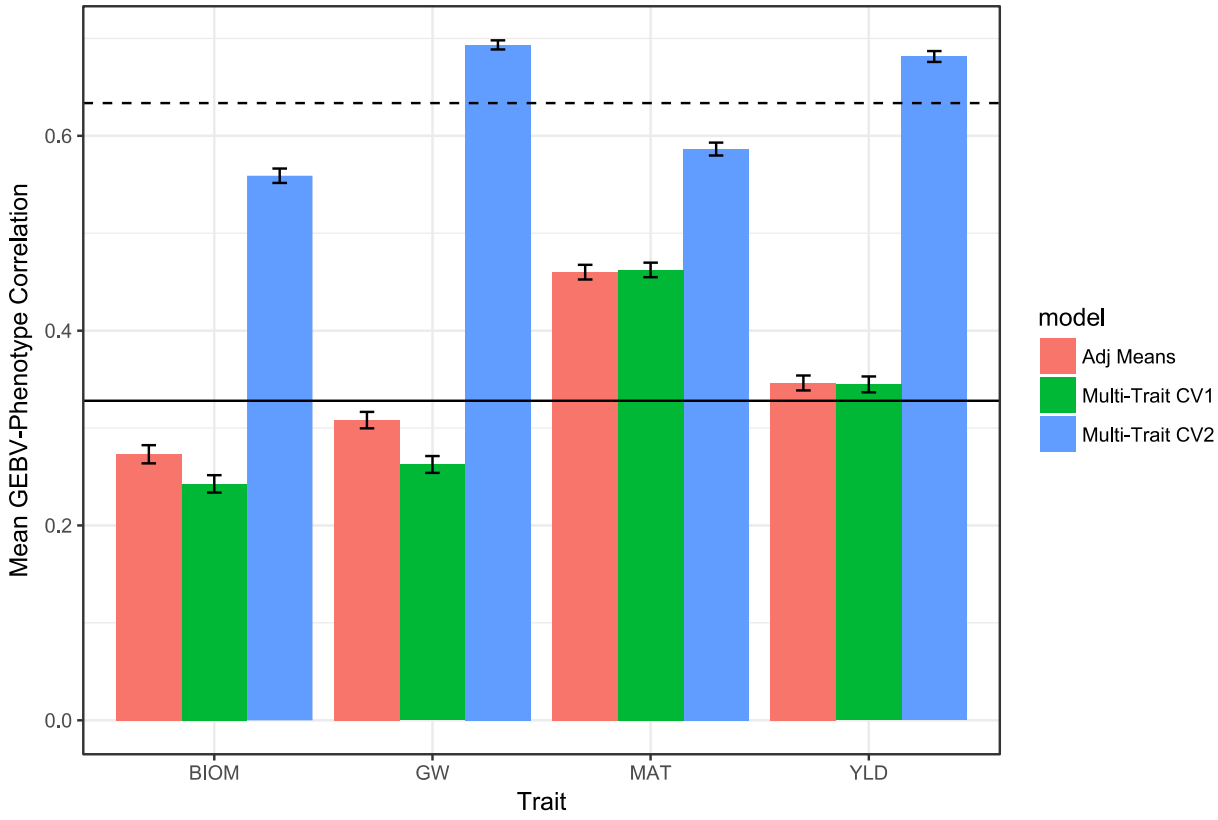
† HD heading date; MAT physiological maturity date; FLSG flag leaf stay green; NDVI normalized-difference vegetation index at Zadok's GS25; HT plant height; BIOM above-ground biomass; GW grain weight; HI harvest index; GSQM grains per square meter; SSQM spikes per square meter; SPH seeds per head; TKW thousand kernel weight; TWT test weight; YLD grain yield; STARCH whole-grain starch content; WCPROT wet chemistry-validated whole-grain protein content

‡ stratified model

§ GEI model using either CV1 or CV2 cross-validation schemes

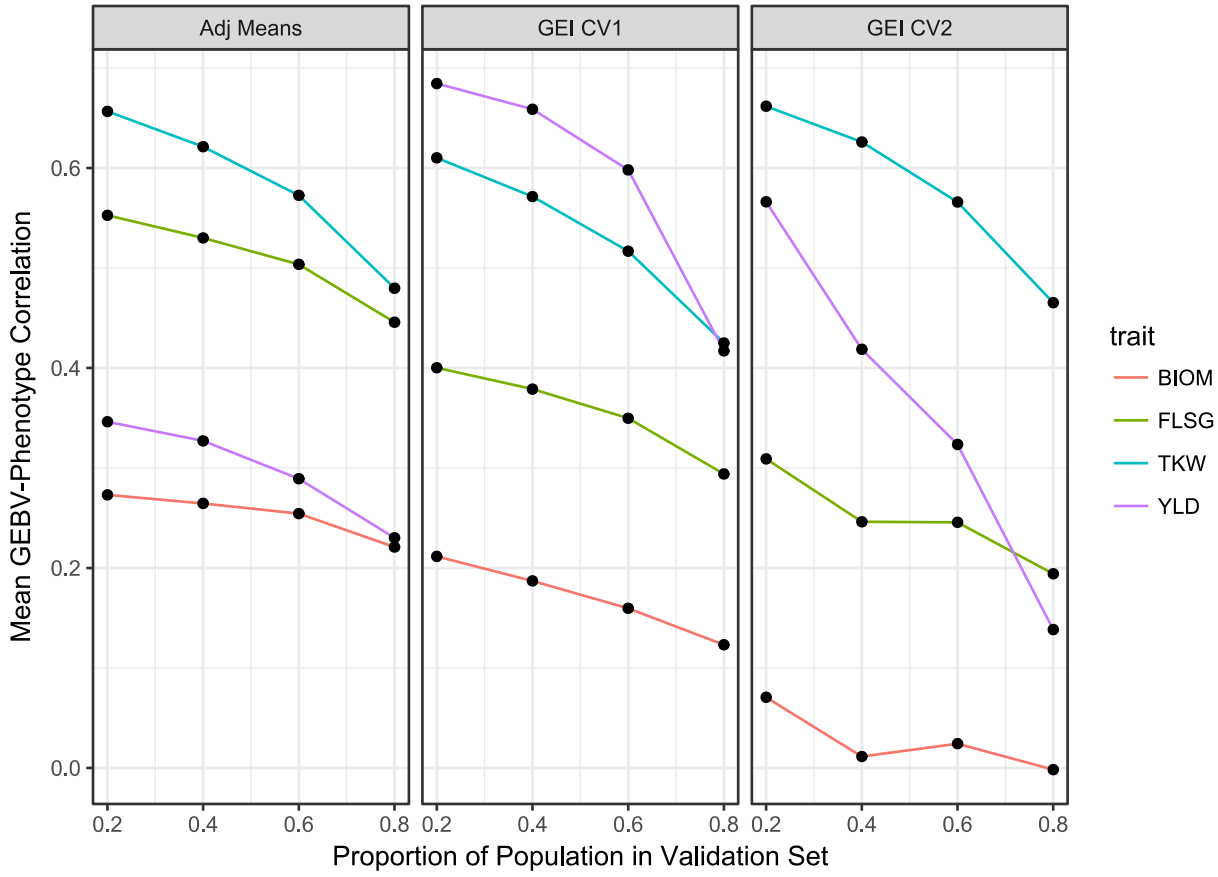
¶ 14Bb Blacksburg, VA 2014; 14War Warsaw, VA 2014; 15Bb Blacksburg, VA 2015; 15War Warsaw, VA 2015

Figure 2.4: Comparison of mean GEBV-phenotype correlations between the adjusted means model, and the multi-trait model using the CV1 and CV2 cross-validation schemes using an 80% / 20% training/validation split. Bars represent 95% confidence intervals. Solid and dashed horizontal lines represent the across-trait mean GEBV-phenotype correlations calculated using the $GLUP_{GE}$ model and the CV1 or CV2 cross-validation schemes, respectively.



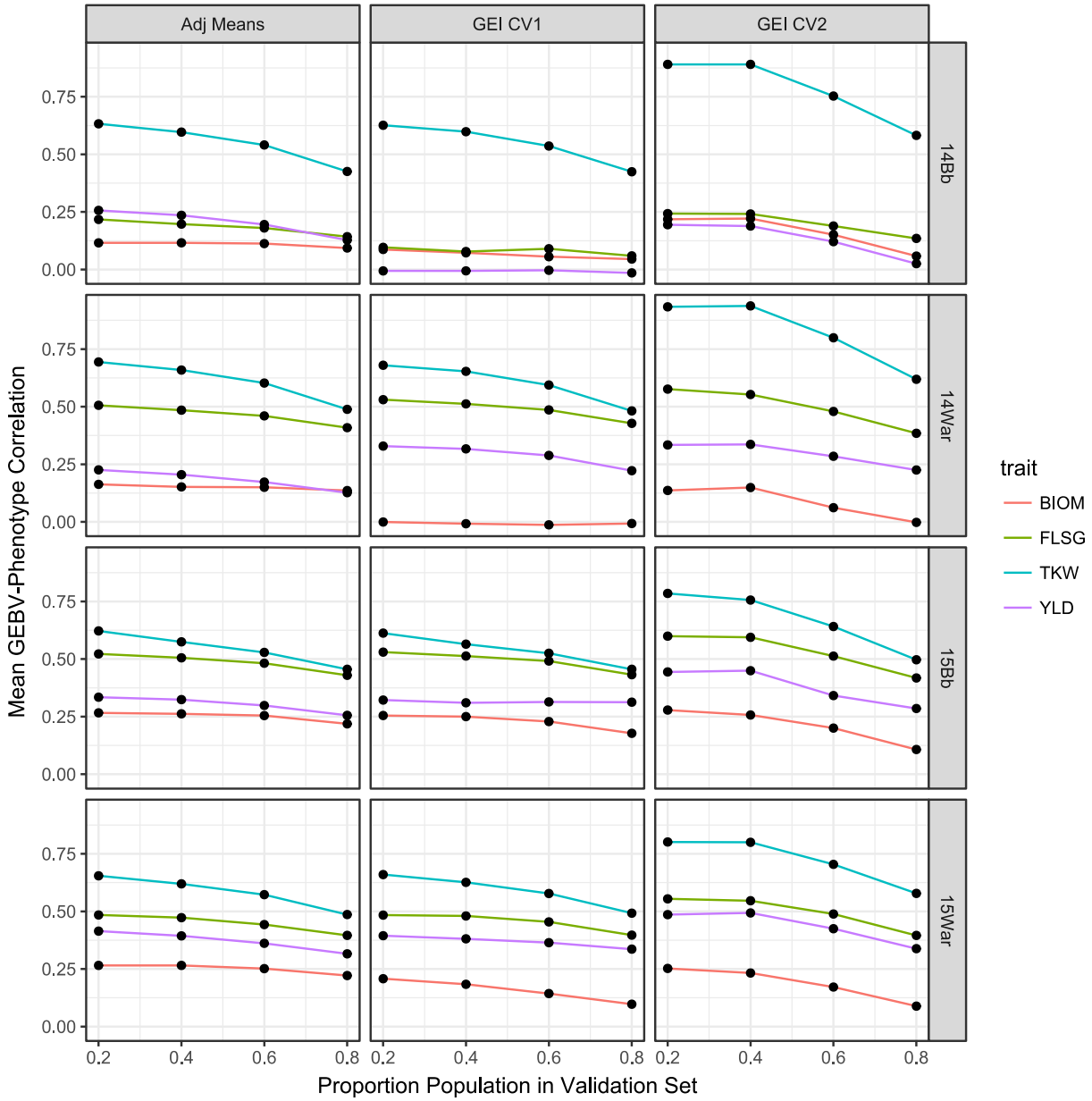
BIOM above-ground biomass; GW grain weight; MAT physiological maturity date; YLD grain yield

Figure 2.5: Mean across-environment GEBV-phenotype correlations across varying training/validation split proportions for selected traits calculated using the adjusted means model and the GEI model using the CV1 and CV2 cross-validation schemes



BIOM above-ground biomass; FLSG flag leaf stay green; YLD grain yield; TKW thousand kernel weight

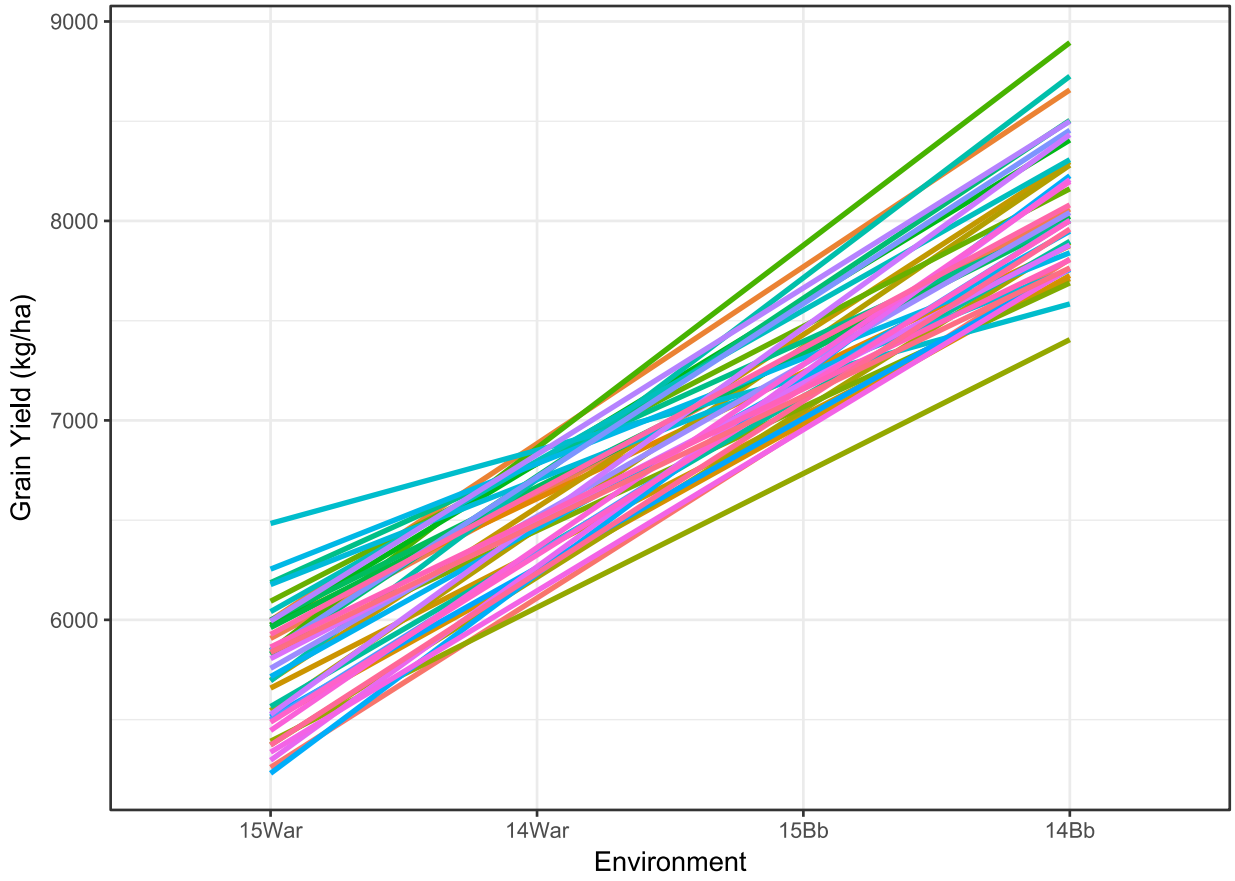
Figure 2.6: Mean within-environment GEBV-phenotype correlations across varying training/validation split proportions for selected traits calculated using the adjusted means model and the GEI model using the CV1 and CV2 cross-validation schemes



BIOM above-ground biomass; FLSG flag leaf stay green; YLD grain yield; TKW thousand kernel weight

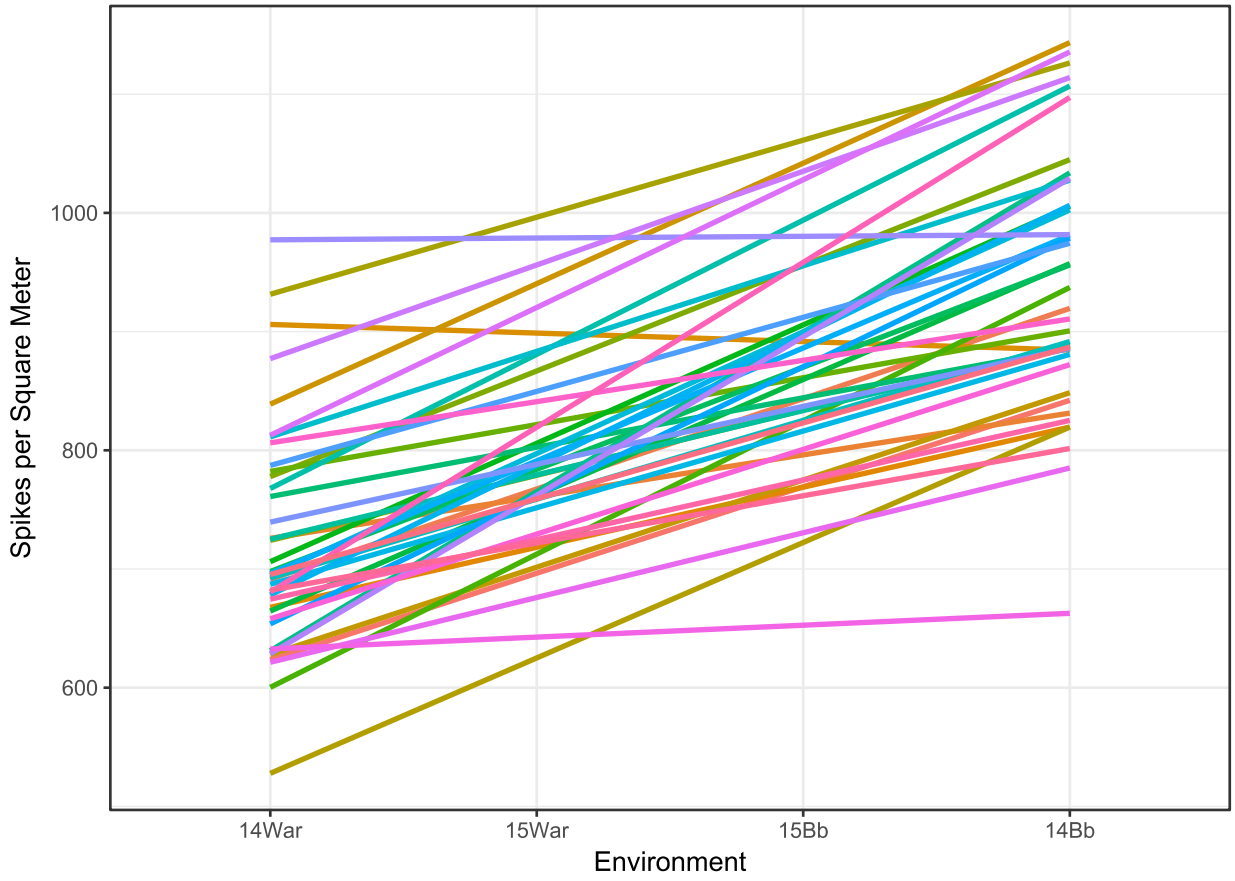
14Bb Blacksburg, VA 2014; 14War Warsaw, VA 2014; 15Bb Blacksburg, VA 2015; 15War Warsaw, VA 2015

Figure 2.7a: Genotype-by-environment interaction plot for the trait grain yield (YLD) across the four testing environments. Each line corresponds to one of the 41 genotypes tested in all environments.



15War Warsaw, VA 2015; 14War Warsaw, VA 2014; 15Bb Blacksburg, VA 2015; 14Bb Blacksburg, VA 2014

Figure 2.7b: Genotype-by-environment interaction plot for the trait spikes per square meter (SSQM) across the four testing environments. Each line corresponds to one of the 41 genotypes tested in all environments.



14War Warsaw, VA 2014; 15War Warsaw, VA 2015; 15Bb Blacksburg, VA 2015; 14Bb Blacksburg, VA 2014

CHAPTER III

Genome-Wide Association Studies in Two Panels of Elite Soft Winter Wheat Lines

B. Ward, M. Huang, D. Van Sanford, A. McKendry, G. Brown-Guedira, P. Tyagi, C. Sneller, and C. Griffey

Abstract

Myriad studies have identified quantitative trait loci (QTLs) affecting multiple traits in wheat (*Triticum aestivum*). However, few studies have verified the locations and effects of QTLs within additional mapping populations. This study utilized genome-wide association (GWA) analyses in two panels of soft winter wheat genotypes. The first, the elite panel (EP), consisted of 273 genotypes of elite breeding germplasm grown in a total of 14 environments. The second, the yield validation panel (YVP), consisted of 294 genotypes grown in a total of 12 environments. In both panels, the traits heading date, plant height, test weight, and grain yield were examined, in addition to stability estimates of test weight and grain yield. Mixed linear model GWA analysis was performed using both a single-locus model (GCTA) and a multi-locus model, which calculates the kinship matrix from a subset of markers (FarmCPU). The GCTA model identified a total of three significant marker-trait associations (MTAs) in the EP and one in the YVP. The FarmCPU model identified a total of 25 significant MTAs in the EP and six in the YVP. However, neither method identified any MTAs that were shared across both panels. Ultimately, the results suggest that more caution is warranted in regards to the identification of QTLs in winter wheat germplasm, and that more rigorous methods of QTL validation should be employed. Many QTLs of major effect may be fixed within elite winter wheat lines, indicating that further genetic gain would likely be more easily achieved via whole-genome prediction of breeding values.

Introduction

The proliferation of high-density DNA marker data for many species has led to the advent and subsequent popularization of linkage mapping procedures. These techniques aim to lift the “statistical fog” of quantitative genetics by identifying quantitative trait loci (QTLs),

which are the genomic regions that exert a high level of influence on the phenotypic variation observed within a population (Mauricio, 2001). The technique of linkage mapping has been used extensively to localize QTLs and estimate their phenotypic effects. Linkage mapping identifies QTLs within mapping populations, which are purposely-constructed to display widely differing phenotypes for a trait of interest. Mapping populations are most often constructed by crossing two parents with widely differing phenotypes for a trait of interest, and then selfing the progeny for several generations to form recombinant inbred lines (Liu, 1998). Linkage mapping studies have been successfully used to identify several stable QTLs of major effect that are widely deployed in plant breeding programs today; for instance the *Fhb1* QTL conferring partial resistance to Fusarium head blight (FHB; *Gibberella zea* [teleomorph]; *Fusarium graminearum* [anamorph]) in wheat (Waldron et al., 1999). However, linkage mapping with biparental populations has the disadvantages of low mapping resolution (due to all recombinations within the mapping population descending from a single meiosis), the presence of at most only two alleles at a given locus, and the relatively long time required to create and stabilize the population (Xu et al., 2016). Genetically pure lines can be developed via doubled haploid breeding methods, partially reducing the time required for population development. In addition, several mapping population designs have been developed by researchers to increase the resolution and/or allelic diversity of linkage mapping studies. These include advanced intercross lines (Darvasi and Soller, 1995), the multi-parent advanced generation inter-cross (MAGIC) population (Kover et al., 2009), and the *Arabidopsis* multiparent RIL (AMRIL) population (Huang et al., 2011). However, these population designs can only increase resolution and allelic diversity up to a point.

Obtaining a mechanistic understanding of the phenotypic effects of a QTL requires progressing from identification of the QTL to identification of candidate genes within the QTL, and ultimately to the identification of quantitative trait nucleotides (QTNs) – the physical polymorphisms underlying observed phenotypic variance (Mackay, 2001). Linkage analyses may occasionally resolve QTL positions to a great enough degree that only a handful of candidate genes are identified (Price, 2006). More frequently, initial mapping localizes a QTL to a large chromosomal segment containing dozens or hundreds of genes. In these cases additional fine-mapping is required to narrow the physical interval over which a QTL may reside. Such studies often have formidable resource requirements, as resolution is largely a function of the size of the mapping population. For example, *fw2.2*, a QTL affecting fruit weight in tomato, was the first QTL in plants for which the underlying causative genetic element was discovered. The effects of *fw2.2* are caused by a single gene (*ORFX*), which is expressed during flower development and controls carpel cell number; a finding which took many years of research and required the screening of thousands of progeny (Alpert et al., 1995; Alpert and Tanksley, 1996; Frary et al., 2000).

QTL Reliability

In addition to the logistical difficulties posed by linkage mapping, past research has found that estimates of the chromosomal location and phenotypic effects of QTLs can be highly influenced by several factors. QTL mapping studies are vulnerable to the same forms of main effect interactions that affect plant breeding trials in general. Mackay (2001) delineated three forms of QTL interaction, all of which may affect the reliability of QTL effect estimates:

1. Genotype × environment interaction (i.e. QTL-environment interaction)

2. Genotype × sex interaction
3. Genotype × genotype interaction (i.e. epistasis; QTL-genetic background interaction)

Note that in a plant genetics context, the second category, genotype × sex interaction, only applies to dioecious species. Mackay further notes that the power to detect epistatic interactions between QTLs is generally low in mapping populations, due to factors such as stringent multiple-testing corrections, the relative paucity of genotypes containing rarer alleles for multiple QTLs, even in large studies, and the segregation of alleles for multiple QTLs aside from those being studied.

The Beavis Effect is another form of bias that may affect the identification of QTLs and the estimation of their effects. Using a simulation study, Beavis found that as the size of a mapping population is decreased, the power to detect QTLs of smaller effect is diminished. Simultaneously, the estimates of the proportion of the genotypic variance explained from QTLs that are identified are inflated (1994, 1998). Beavis found that these effect estimates were highly overestimated when a biparental population consisting of 100 progeny was used for QTL mapping, slightly overestimated when the size of this population was increased to 500 progeny, and very nearly equivalent to the true magnitude when the population size was further increased to 1,000 progeny. The source of this bias is the multiple-testing correction procedure used to estimate a suitable significance threshold; since only those QTL exceeding the significance threshold are reported, QTL effect estimates are drawn from a truncated distribution (Xu, 2003).

Multiple studies have concluded that the estimated locations and effects of QTLs may vary substantially, even among studies utilizing the same population. Beavis (1994) summarized several QTL mapping studies that all examined the same genetic background: progeny of the biparental maize (*Zea mays* subsp. *mays*) cross B73 × Mo17. He found that even when using this

single population, the location and effects of reported plant height and grain yield QTL were inconsistent across experiments. It should be noted that this study suffered from several confounding factors, such as the use of different marker sets across the original experiments, and differential levels of inbreeding. However, Austin and Lee (1998) performed a study with maize in which the same population was tested across environments in a more consistent manner. They performed linkage mapping using 185 $F_{6:7}$ progeny of the biparental cross Mo17 × Hg99, evaluated for grain yield and four yield components in both moisture stressed and non-stressed conditions in the same location across two years. They found a total of 59 QTLs in at least one year, though only ten of these occurred across both years.

QTL effect inflation may also occur when effects are estimated in the same population that is used for mapping (Lande and Thompson, 1990). Melchinger et al (1998) performed a study in maize in which they mapped QTLs with restriction fragment length polymorphism (RFLP) markers using one set of genotypes, and subsequently estimated the effects of these QTLs in an independent validation set of genotypes. They also investigated the power to detect QTLs across samples of different sizes, and determined the consistency of QTLs across multiple testcrosses. They found that estimated QTL effects were significantly lower when calculated using the set of validation genotypes as opposed to the original set of genotypes in which the QTLs were mapped. In addition, fewer QTLs were identified when a smaller mapping population was used, confirming the predictions of Beavis' simulation models. Finally, >50% of identified QTLs were identified across multiple testcrosses for the majority of traits, with the exception of grain yield, for which no QTLs were shared across testcrosses. Given that this experiment was carried out using hybrid maize, the authors hypothesized that the lack of

common grain yield QTLs among the testcrosses was due to dominance effects (i.e. masking effects due to different alleles present within some tester lines).

This study was followed by one in which cross-validation was used to test the stability of QTLs across environments, samples of genotypes, or both by dividing testcrosses of 344 F_3 maize genotypes into various training and validation sets (Utz et al., 2000). They once again found an upward bias in estimated QTL effect sizes when using the training set of genotypes to both map QTLs and estimate their effects. In addition, they found that sampling genotypes generally had a greater effect on QTL effect estimates than sampling environments. The results led the authors to recommend the use of cross-validation methods to generate unbiased asymptotic estimates of QTL effects.

Finally, a larger study was carried out utilizing 976 F_5 maize testcross progenies across a total of 19 environments (Schön et al., 2004). The authors found that both the number of QTLs identified and the proportion of genotypic variance that they collectively explained increased as the number of genotypes and number of environments were increased. However, this effect was influenced more by the number of genotypes than the number of environments. In addition, the bias of QTL effect estimates generally decreased as the number of genotypes and environments increased. However, despite the large amounts of resources devoted to this project, the maximum collective amount of phenotypic variance explained by identified QTLs for any trait was 52.3%, indicating that the study exhibited a significant amount of "missing heritability," a phenomenon that has been well-noted in the context of QTL identification (Manolio et al., 2009).

Studies such as those mentioned above have highlighted some of the limitations of linkage mapping, and indeed, there have been relatively few contemporary examples of rigorously-validated, stable QTLs for highly quantitative traits being integrated into cultivars in

proportion to the number of QTL mapping studies that have been carried out. Bernardo (2008) estimated that over 10,000 marker-trait associations had been reported in 12 crop species by 2008, but examples of QTLs being successfully used in cultivar development remained rare.

Nevertheless, there are several success stories of QTLs that have been identified and subsequently exploited in crop species. One particularly noteworthy example, as previously mentioned, is the wheat QTL *Fhb1*, which confers partial resistance to fusarium head blight. No known qualitative resistance against FHB exists. *Fhb1* is located on chromosome 3B, and is derived from the Chinese cultivar Sumai 3. It was first identified in a population of recombinant inbred lines (RILs) derived from the cross Sumai 3 × 'Stoa' (Waldron et al., 1999). Shortly thereafter, the QTL was also detected in a RIL population derived from the cross ND2603 × 'Butte 86' (Anderson et al., 2001). The phenotypic effects of *Fhb1* were subsequently validated across multiple genetic backgrounds by forming multiple near-isogenic lines (NILs) differing only by the presence or absence of *Fhb1* (Pumphrey et al., 2007). Since the discovery of *Fhb1*, the gene(s) and QTNs underlying its FHB-resistance properties remained elusive. However, a recent study succeeded in positionally cloning *Fhb1*, and found that a pore-forming toxin-like (PFT) gene was the causative locus of the QTL (Rawat et al., 2016).

Genome-Wide Association Analysis

Genome-wide association (GWA) analyses are an alternative method of identifying marker-trait associations, which utilize assembled panels of individuals, rather than the purposely-constructed mapping populations used in linkage analysis. In GWA studies, the ability to identify QTLs is a result of the ancestral chromosomal recombinations present within the panel, rather than the recombinations stemming from one or several meiosis events within a

population, as in linkage mapping. In addition, since GWA panels are not derived from two or a few parents, allelic diversity at any given locus may be higher (Rafalski, 2010). GWA studies typically offer higher resolution than linkage mapping studies, due to the generally more extensive recombination and decay of linkage disequilibrium between more distantly-related individuals (Zhu et al., 2008). In addition, the ability to assemble panels from a wide range of germplasm with varying degrees of relatedness should in theory enable the identification of QTLs with effects that are less population-specific than those identified via linkage mapping (Jannink et al., 2010). The primary disadvantage of GWA analyses is that QTLs arising from rare alleles will typically not be detected, unlike in linkage mapping, where the construction of a mapping population will inflate the frequencies of rare alleles (Myles et al., 2009). Methods of directly assessing QTL-environment interaction have been implemented in linkage mapping (Boer et al., 2007; Malosetti et al., 2008, 2004; Moreau et al., 2004). However, fewer methods have been developed specifically for modeling QTL-environment interaction in the context of GWA studies, with some notable examples being a multi-trait mixed model (Korte et al., 2012), and the use of both linkage mapping and GWAS in the same study (Sterken et al., 2012).

In recent years, the reproducibility of scientific studies has become a major concern in multiple fields (Open Science Collaboration, 2015), and a source of ongoing debate (Ioannidis, 2014, 2005; Jager and Leek, 2014). In studies of human diseases, many thousands of GWA studies have been performed, but have come under scrutiny for yielding largely irreproducible results (Hirschhorn et al., 2002). This has been due in large part to the use of inadequate sample sizes, low coverage of the overall genetic variability within a population, and liberal significance thresholds (Zeggini and Ioannidis, 2009).

Very few assessments of GWA study reproducibility have been carried out in the context of plant breeding and genetics, though the accounts that have been written suggest similar problems as those encountered in human studies. One GWA study of southern leaf blight (*Cochliobolus heterostrophus* [teleomorph]; *Bipolaris maydis* [anamorph]) carried out in a maize nested association mapping (NAM) panel found 245 significant MTAs (Kump et al., 2011). The maize NAM panel consists of ~5,000 genotypes, and shares characteristics of both linkage mapping populations and GWA panels, as it consists of RILs descended from multiple founder lines crossed to a single common parent line (Yu et al., 2008). Thus, in studies utilizing a NAM panel, a GWA analysis is typically performed following a joint linkage mapping of the multiple RIL families. Kump et al. utilized a 1,106-marker map for performing the initial joint linkage mapping, and 1.6 million maize haplotype map (HapMap) v1 SNPs (Gore et al., 2009) for performing the subsequent GWA analysis. A later analysis of the same phenotypic data using an updated 7,386-marker map for joint linkage mapping and 28.5 million maize HapMap v2 SNPs (Chia et al., 2012) identified 192 significant MTAs, but found that only 6% of the combined set of significant SNPs from both studies co-localized within 10 Kbp windows (Bian et al., 2014). Subsequently, four additional GWAS models were compared to determine the effects of model input parameters (e.g. genetic map used for joint linkage mapping, number of SNPs used for GWA analysis) on the identification of marker-trait associations. The authors found that the GWA analyses were highly sensitive to the model inputs and for the best match between two GWA models only 26% of MTAs overlapped within 10Kb windows.

In the present study, GWA analyses were conducted on two separate panels grown in separate sets of environments to assess the traits heading date, test weight, plant height, and grain yield, as well as stability estimates for test weight and grain yield. The objectives of this study

were to (i) identify significant MTAs occurring within one or both mapping panels, and (ii) assess the stability of significant MTAs occurring in both panels. The same set of SNPs generated via a genotyping-by-sequencing (GBS) protocol was utilized in both panels, in order to reduce any bias due to differences in local and genome-wide marker densities.

Materials and Methods

Germplasm and Testing Environments

Two separate genotype panels were included in the study. The first was the elite panel (EP), which consisted of 273 elite soft winter wheat lines sourced from breeding programs at Purdue, Cornell, University of Kentucky, The Ohio State University, Virginia Tech, University of Maryland, University of Missouri, Michigan State University, University of Illinois, and the University of Arkansas. A list of genotypes included in the EP is presented in **Table F.1**. The second was the yield validation panel (YVP) that consisted of 294 elite soft winter wheat lines sourced from breeding programs at Purdue, University of Kentucky, The Ohio State University, Virginia Tech, University of Maryland, University of Missouri, and University of Illinois. The YVP consisted of multiple sets of two or more full sibling lines. A list of the genotypes included in the YVP is presented in **Table G.1**. Only two check lines were shared in common between the EP and the YVP and included the Purdue genotype 05247A1-7-3-120 and the Missouri genotype MO080864.

Phenotypic Data Collection and Analysis

The EP was grown in the 2011-2012 and 2012-2013 winter wheat growing seasons in a total of 14 environments located in Kentucky, Maryland, Missouri, Ohio, and Virginia. The

experimental sites located at Wooster, OH and Warsaw, VA utilized two nitrogen application rates – one moderate (100 kg ha^{-1}), and one low (67 kg ha^{-1}). All other locations only used the moderate application rate. An environment was defined as each unique combination of year, location, and nitrogen treatment rate.

The YVP was grown in the 2013-2014 and 2014-2015 winter wheat growing seasons in a total of 13 environments located in Kentucky, Missouri, Ohio, and Virginia. As with the EP, an environment was defined as a unique combination of year, location, and nitrogen application rate. For the YVP, the differential nitrogen applications (moderate and low) were only applied in Columbia, MO and Wooster, OH, with all other locations only receiving the moderate nitrogen application rate. All environments in which the EP and YVP were grown are listed in **Table H.1**.

Genotypes were planted in standard yield test plots which varied slightly by location depending on equipment used and the number and length of rows planted. In Virginia, the wheat lines of both tests were planted in individual plots with a length of 2.74 m, width of 0.91 m, row-spacing of 15.2 cm, and a harvested area of 2.49 m^2. The design utilized for each environment consisted of an incomplete block design, with repeated checks included in each block. For the EP, the check line 'Branson' was planted eight times within each block, with each block consisting of 64 plots (56 plots not including the check plots). Therefore, a single replication of the experimental design consisted of a total of five blocks. For the YVP, four separate check lines were randomly located once within each block; these were Branson, 'Shirley', 'Milton', and Pioneer '25R47'. Once again, each replication consisted of a total of five blocks, and each block consisted of a total of 60 plots (56 plots not including checks).

The traits heading date (HD), plant height (HGT), test weight (TW), and grain yield (YLD) were evaluated in both the EP and YVP. However, as shown in **Table H.1**, all traits were

not evaluated in all environments. Heading date was recorded as the Julian date at which 50% of plants within a plot had extruded heads from the boot (Zadoks et al., 1974). Plant height was recorded as the height, in centimeters, from the soil surface to the tip of the heads (excluding any awns if present). Once plants had fully ripened (Zadoks growth stage 92), plots were harvested using plot combines. Test weight (kg hl^{-1}) and grain moisture content (%) were either measured during harvest using the plot combine (if using a data-collecting combine), or else post-harvest using a GAC® 2500-AGRI grain analysis computer (Dickey-John® Corporation, Auburn, IL). Grain yield was calculated from the total harvested grain from each plot using units of kg ha^{-1} at 13.5% moisture equivalence.

Analysis of phenotypic data was carried out using a two-step approach. First, the repeated checks were used to calculate correction factors for each individual block/trait combination using the formula below:

$$Y_{adj} = Y_{raw} - (\mu_W - \mu_A) \quad (1)$$

A plot's phenotypic value adjusted for block effect (Y_{adj}), is a function of the unadjusted phenotypic value for that plot (Y_{raw}), minus the difference between the mean of replicated checks within a block (μ_W) and the mean of the replicated checks across all blocks (μ_A). For the YVP, which had four separate repeated check genotypes, all checks were pooled together to perform block corrections.

Since the EP and the YVP both contained many environments with single-replication designs, the following random effects model was fit using the *lme4* package (Bates et al., 2015) in the R statistical computing environment (R Core Team, 2015):

$$Y_{ij} = \mu + G_i + E_j + \varepsilon_{ij} \quad (2)$$

The block-adjusted phenotypic response (Y_{ij}) is a function of the overall mean (μ), the ith genotype (G_i), the jth environment (E_j), and the residual error (ε_{ij}). All effects were treated as random. Within-environment genotypic means were calculated for all environments containing two replications prior to running the model described in equation (2). For each trait, variance components for all effects were estimated (note that variance due to GEI cannot be directly assessed, and is contained within the residual variance). Entry mean heritability was calculated for each trait using the following equation:

$$H^2 = \frac{\sigma_g^2}{(\sigma_\varepsilon^2/E) + \sigma_g^2} \quad (3)$$

Entry mean heritability (H^2) is a function of genotypic variance (σ_g^2), the residual variance (σ_ε^2), and the number of environments (E). In addition, genotypic best-linear unbiased predictors (BLUPs) were calculated from the model described in equation (2) for use as the phenotypic input for the subsequent GWAS analyses.

Trait Stability Estimation

For the traits grain yield and test weight, stability estimates were generated using both Eberhart and Russel regression (Eberhart and Russell, 1966) and an additive main effect and multiplicative interaction effect (AMMI) model (Gauch, 1988), as described by Huang et al. (Huang et al., 2016). Briefly, ERR is a simple linear regression:

$$y_{ij} = \mu_i + \beta_i I_j + \varepsilon_{ij} \quad (4)$$

Where y_{ij}, the phenotypic response of the *i*th genotype in the *j*th environment, is a function of μ_i, the mean phenotype of the *i*th line across all environments, I_j, an environmental index defined as the mean performance of all genotypes within environment j, and the residual error ε_{ij}. The

regression coefficient, β_i, then functions as the stability estimate for the *i*th genotype. The AMMI model for the phenotypic response (*y*) of the *i*th genotype in the *j*th environment is:

$$y_{ij} = \mu + G_i + E_j + \sum_{n=1}^{N} \lambda_n \gamma_{in} \delta_{jn} + \varepsilon_{ij} \quad (5)$$

Where μ is the grand mean, G_i is the main effect of the *i*th genotype, E_j is the main effect of the *j*th environment, λ_n is the eigenvalue of the *n*th interaction principal component (IPC) axis, γ_{in} and δ_{jn} are the genotype and environment PCA scores, respectively, for IPC axis *n*, *N* is the total number of IPCs retained in the analysis, and ε_{ij} is the residual error. For the AMMI stability estimates, the first 10 and 3 IPCs were used for yield and test weight, respectively.

Genotyping

Genomic DNA was isolated from fresh green leaf tissue using a cetyltrimethylammonium bromide (CTAB) extraction protocol (Saghai-Maroof et al., 1984). Genotyping-by-sequencing was performed at USDA Agricultural Research Service (ARS) facilities using a *PstI-MseI* double digest of genomic DNA. The SNP calling was performed using TASSEL-GBS in TASSEL 5.2.24 (Bradbury et al., 2007; Glaubitz et al., 2014). The Burrows-Wheeler aligner (Li and Durbin, 2009) was used to align SNPs to the International Wheat Genome Sequencing Consortium's whole genome assembly v0.4. In addition to the GBS genotyping described above, several major gene loci highly associated with agronomic performance were genotyped using simple sequence repeat (SSR) markers and LGC® KASP™ SNP genotyping assays for the YVP only.

SNP Quality Filtering and Imputation

Prior to imputation of missing genotypes, the genotypic dataset for the EP and YVP material was jointly filtered to remove SNPs with missing data frequencies >50%, heterozygous call frequencies >15%, and minor allele frequency <5%. In addition, all unaligned SNPs were removed. After the initial filtering, missing data in the genotypic dataset was imputed using LinkImpute (Money et al., 2015). LinkImpute implements a nearest-neighbor algorithm using both the *k* nearest individuals and the *l* SNPs in highest LD with the specific missing SNP genotype that must be imputed. LinkImpute was used with its default settings, which optimize the number of nearest individuals and SNPs via data masking simulations at 10,000 randomly selected genotypes. After imputation, the dataset was once again filtered to remove SNPs with minor allele frequencies <5%. The imputed genotypic dataset was finally filtered in PLINK 1.9 (Chang et al., 2015) to remove all but one SNP in clusters separated by <64bp, as this is the tag size used in the TASSEL-GBS SNP-calling pipeline, (i.e. all SNPs located on the same tag should have the same genotype prior to imputation). In addition to the positional filtering, PLINK was used to remove all but one SNP in groups of SNPs in perfect LD ($r^2 > 0.99$) using a 200-SNP sliding window, advancing by 5 SNPs with each step.

Population Structure and Linkage Disequilibrium

Prior to performing GWAS, population structure for the EP and YVP material was examined via principle component analysis (PCA) of the filtered and imputed genotypic data using the SNPRelate package (Zheng et al., 2012) in R. The fixation index (F_{ST}) was calculated in SNPRelate using the method of Weir and Cockerham (1984) to determine the degree of genetic differentiation between the EP and YVP. Linkage disequilibrium (LD) was estimated on

a pairwise basis for all intra-chromosomal SNPs in the separate EP and YVP genotypic datasets, yielding 3,305,761 comparisons. The LD decay was plotted for the A, B, and D genomes separately by randomly selecting 20,000 pairwise comparisons from each genome. Then LD was plotted for each separate chromosome of the B genome using 20,000 randomly-selected pairwise comparisons from each chromosome. In addition, inter-chromosomal LD was calculated between chromosomes within each genome. For each genome, 1,000 SNPs were randomly selected to form inter-chromosomal pairs, yielding a total of 470,385 comparisons in the EP, and 471,019 comparisons in the YVP. For intra-chromosomal SNPs, r^2 values for pairwise LD comparisons were plotted against physical distance, and a second-degree locally-weighted scatterplot smoothing (LOESS) curve was fit to the data (Cleveland, 1979). For LD estimates from non-linked (i.e. inter-chromosomal) loci, the 98th percentile of the LD distribution was defined as the linkage-disequilibrium critical value. All r^2 values exceeding this value were assumed to have been caused by genetic linkage (Breseghello and Sorrells, 2006).

Genome-Wide Association Analysis

For each trait, genome-wide association analysis was performed using the Genome-Wide Complex Trait Analysis (GCTA) software (Yang et al., 2011), using a leave-one-chromosome-out (LOCO) method in which a separate genetic relationship matrix (GRM) is estimated from SNP data for each chromosome. Specifically, the LOCO approach entails excluding all SNPs located on the chromosome of the SNP undergoing testing when estimating the GRM. For each trait, permutation testing was performed to empirically determine a significance threshold by randomly shuffling phenotypic data and included principle components in unison, performing the

GCTA-LOCO analysis on the randomly-reordered data, recording the lowest observed p-value, and repeating this process 1,000 times.

In addition, GWA analysis was run on each trait using the Fixed and random model Circulating Probability Unification (FarmCPU) model (Liu et al., 2016) in R. Once again, permutation testing was performed for 1,000 iterations for each trait. In addition, to enhance confidence in QTLs with p-values exceeding the significance threshold in FarmCPU, a bootstrapping method utilized by Wallace, et al. (2016) was implemented, in which 10% of the phenotypic observations were replaced with missing data for a total of 100 runs of the model. Subsequently, for each trait the resample model inclusion probability (Valdar et al., 2009) was calculated for each SNP by determining the fraction of bootstraps in which its p-value exceeded the permutation-derived significance threshold. The value 0.05 was chosen as a lower threshold for the RMIP as it coincided with the point of inflection in the RMIP density curve (data not shown).

For each model, panel, and trait combination, the determination of the number of principle components to include to correct for population structure was based upon observation of the proportion of variance explained by each genotypic PC, as well as the inflation of p-values observed in preliminary GWA runs in which population structure was not corrected for. Ultimately, the first five PCs were included to correct for population structure in all GCTA models, as well as the FarmCPU models for direct trait measurements (i.e. not trait stability estimates). For the trait stability estimates analyzed with FarmCPU, the first two genotypic PCs were used.

Results

General Genotype Performance

Summary statistics of general performance of genotypes for all environments and for each panel are reported in **Table 3.1**. As previously stated in the materials and methods, a list of the individual environments included in the study is given in **Table H.1**. The environments and genotypes included in the YVP were characterized by a later mean heading date than that observed in the EP (136 Julian days vs. 125), as well as lower mean grain yields (4,838 kg ha^{-1} vs. 5,447). The across-environment standard deviations for each trait were highly consistent across both panels, with the exception of heading date, for which the standard deviation was over twice as high in the EP (14.05) vs. the YVP (6.34). The entry-mean heritability values calculated for each trait were generally highly consistent between both panels; the trait test weight exhibited the greatest difference in calculated heritability between the two panels (0.9 in the EP vs. 0.87 in the YVP), while all other traits exhibited differences in heritability values between the two panels of 0.01 or less. The trait with the lowest heritability was grain yield, with values of 0.79 and 0.8 in the EP and YVP respectively, while the trait with the highest heritability was plant height, with values of 0.96 and 0.95 in the EP and YVP respectively.

Patterns of GEI and similarity of environments were assessed for each trait using GGE biplots. The results for the EP are displayed in **Figures I.1 – I.4**, while those for the YVP are displayed in **Figures J.1 – J.4**. Note that genotype labels are suppressed in these figures for the sake of clarity. In the context of GGE biplots, a mega-environment is defined as a set of environments in which a single genotype exhibits the "best" performance. This is illustrated by the polygon demonstrating "who won where", where each vertex represents a genotype that performed best in a set of environments (Yan and Kang, 2003). Mega-environments are therefore delineated by the dashed lines radiating outward from the origin. Note that many of the identified

mega-environments may be hypothetical (i.e. not containing any of the environments included in the study).

A majority of traits demonstrated divergent GEI patterns across both panels. For instance, for the trait heading date, environments included in the EP mostly clustered within a single mega-environment, with only the 2012 Virginia and Maryland environments clustering into a second mega-environment. In contrast, for heading date in the YVP, environments were spread across three mega-environments, with both Virginia environments clustering in one mega-environment, the 2015 Wooster, OH environments clustering in a second, and all others falling within a third. The opposite trend was observed for the trait plant height, where the EP environments fell within three mega-environments, while all YVP environments (with the exception of the 2014 Kentucky environment) fell within a single mega-environment. The trait test weight exhibited the most consistent GEI patterns across the EP and YVP. In both panels, environments were split into three mega-environments, with the moderate and low nitrogen treatments of one year's Wooster, OH test forming a more distant cluster. In both panels, grain yield exhibited the most environmental diversity, with environments falling within four separate mega-environments.

Population Structure and Linkage Disequilibrium

Many of the population characteristics found in the allele-based breeding panel (Chapter 1) were also observed in both the EP and YVP. For instance, neither panel demonstrated very high levels of population substructure, with the first PC of the EP genotypic data explaining approximately 5.5% of variation, and the first PC of the YVP genotypic data explaining approximately 6% of variation (**Figure 3.1a**). Furthermore, no well-defined point of inflection

was observed in the genotypic PCA data (**Figures 3.1a and 3.1b**), suggesting significant admixture within each panel.

The two panels were also highly related to each other, as expected, with a F_{ST} value between the two of 0.009. The most significant portion of population structure within both panels appears to be explained by the presence or absence of the *Sr36* stripe-rust resistance gene, as determined by KASP assay. It should be noted that the YVP was genotyped with this assay, while the EP was not. However, based upon the clustering of genotypes in the principal component biplot shown in **Figure 3.2**, it appears likely that the EP (designated by "no data" points in the biplot) can likewise be divided into two separate groups based upon the presence or absence of *Sr36*. In contrast, neither the 1B:1R nor the 1A:1R alien translocations from rye (*Secale cereale*) appeared to explain a significant proportion of the observed population structure (data not shown).

The number of SNPs assigned to each chromosome varied widely by genome (**Figure 3.3**), with B-genome chromosomes generally having the greatest saturation. However, chromosomes 4A and 7A did have higher numbers of assigned SNPs than their B-genome homeologous counterparts. Chromosomes on the D genome invariably had the lowest numbers of SNPs for every homeologous chromosome set. The 98^{th} percentile of a sample of pairwise LD estimates for inter-chromosomal loci was used as a significance threshold for pairwise LD estimates for intra-chromosomal loci. All three genomes in both panels exhibited mean LD decay below this threshold at a distance of between 10Mb and 15Mb (**Figure 3.4a**). One exception to this was the D genome in the YVP, which exhibited a local "valley" of LD at ~12.5Mb, with a slight subsequent increase in mean LD at ~17Mb. This could be due to the exceptionally poor SNP saturation of several chromosomes in the D genome, in particular chromosome 4D, which

only contained 204 SNPs. In both the EP and YVP, chromosome 2B exhibited more extensive LD than other chromosomes within the B genome (**Figure 3.4b**), suggesting that this LD is due to the 2G:2B translocation.

Genome-Wide Association Studies

A greater number of significant MTAs were identified in the EP than the YVP when using either the GCTA or FarmCPU algorithms. Within each panel, the FarmCPU algorithm identified far more significant MTAs than the GCTA algorithm. In total, GCTA identified four MTAs exceeding the empirically-determined 0.05 alpha experiment-wide threshold (**Table 3.2**). This included three MTAs in the EP (two for plant height and one for test weight stability), and one MTA for ERR test weight stability in the YVP. FarmCPU identified a total of 31 significant MTAs (**Table 3.3**); 25 of these were in the EP, with the remaining 6 in the YVP. The FarmCPU-identified MTAs in the EP affected the traits heading date, plant height, and test weight; as well as the ERR stability estimates for test weight and yield, and the AMMI stability estimate for yield. The FarmCPU-identified MTAs in the YVP affected the traits heading date and plant height. Notably, no significant MTAs were identified for yield *per se*. Of the four MTAs identified by GCTA, two were also identified by FarmCPU. These were S5D_39764565 (for ERR test weight stability) and S6A_416179127 (for plant height); both of these MTAs occurred in the EP.

Overall there was little evidence to suggest the presence of any pleiotropic regions in the GCTA or FarmCPU results of either panel (**Tables 3.2 and 3.3**). In the FarmCPU EP results, two significant MTAs on chromosome 3A (S3A_15558351 S3A_699195908) were associated with the ERR stability estimate for test weight and plant height, respectively. Three significant MTAs

on chromosome 4B (S4B_25886179, S4B_289037575, and S4B_415342383) were associated with the traits plant height, ERR test weight stability, and ERR yield stability, respectively. Two MTAs located on chromosome 5B (S5B_42564147 and S5B_550563262) were associated with the ERR test weight stability estimate and heading date. Two MTAs on chromosome 5D (S5D_39764565 and S5D_543090768) were associated with the ERR test weight stability estimate, and the AMMI stability estimate for yield, respectively. Three MTAs on chromosome 6A (S6A_416179127, S6A_569599082, and S6A_99931123) were associated with plant height, and the ERR and AMMI yield stability estimates, respectively. Finally, three MTAs on chromosome 7B (S7B_641908294, S7B_653950790, and S7B_724034080) were associated with the ERR stability estimates for yield and test weight. In the YVP panel, two MTAs on chromosome 3B (S3B_23975004 and S3B_782374448) were associated with heading date and plant height, respectively. However, for all the groups of MTAs located on common chromosomes listed above, LD between individual SNPs tended to be quite low, with a maximum genotypic r^2 value of 0.16 between the SNPs S6A_99931123 and S6A_416179127 in the EP, suggesting that any significant MTAs located on common chromosomes likely do not reside within any common QTLs.

In addition, there was no evidence of any MTA being identified across both panels. In the FarmCPU analysis, MTAs for plant height were identified at two sites (S4B_25886179 and S4B_662927988) on chromosome 4B in the EP and YVP, respectively. However, the genotypic r^2 between these two SNPs in the joint EP/YVP genotypic data was low (0.002), suggesting that these two SNPs represent separate QTLs. The maximum genotypic r^2 value between two significant MTAs (0.44) occurred between the loci S6A_68995191 and S6A_416179127, which were associated with the trait heading date in the YVP and plant height in the EP. However, the

physical distance between these two SNPs is ~347Mb, making it highly unlikely that the linkage disequilibrium between these two loci is being caused by physical linkage.

Discussion

The two panels used in this study were highly genetically homogeneous with respect to each other, indicating that they could be assumed to have been sampled from a single population. Indeed, the division of genotypes between the two panels did not coincide with any obvious patterns of population structure present within the joint genotypic data (data not shown). Instead, it was the presence or absence of the *Sr36* stem rust resistance gene that best described the population structure of both panels (**Figure 3.2**). As noted in chapter 1, *Sr36* is located on the 2G:2B alien translocation, which originated from *Triticum timopheevi* (A^mA^mGG) (Brown-Guedira et al., 1996, p.; Nyquist, 1962). An LD analysis of the *Sr36* locus in the allele-based breeding panel in chapter 1 demonstrated very high levels of LD throughout most of chromosome 2B, indicating that the 2G:2B translocation has yet to undergo significant LD decay since its incorporation into common wheat breeding germplasm. While the *Sr36* locus was only interrogated in the YVP in the present study, the presence of higher degrees of long-range LD in chromosome 2B as opposed to the rest of the B genome (**Figure 3.4b**) indicates that the EP and YVP likely both exhibit a similar high-LD block present on 2B.

Given the close interrelatedness of the two testing panels, it is somewhat surprising that no common MTAs were identified between them for any of the assessed traits. In light of the very few significant MTAs identified in either panel by the conventional mixed model approach implemented in GCTA (**Table 3.2**), it is likely that the study was underpowered. The results of GWA analyses using elite breeding material carry the advantage of being more immediately

applicable to the objectives of an applied breeding program (Spindel et al., 2015), though one potential complication arising from such studies is an overall lack of genetic diversity. Given the relatively homogeneous nature of the tested germplasm, it is likely that many QTLs of major effect are already fixed within each panel. Ultimately, much larger, more powerful experimental designs may be required for further genetic dissection of traits in elite wheat germplasm, one example being the nested association mapping panel, which has the advantage of compensating for the effects of population structure by design (Yu et al., 2008). However, as previously noted, even this design has delivered inconsistent GWA results in studies of maize Southern Leaf Blight. Nevertheless, the FarmCPU method did identify many MTAs that GCTA failed to detect (**Table 3.3**). However, these QTLs were not reliably detected across both panels, suggesting that they are highly influenced by interaction effects, as will be discussed below. Despite a general inability to detect reliable MTAs, both panels did still demonstrate phenotypic variation for each trait (**Table 3.1**), suggesting the continued presence of exploitable genetic variation. However, this variation may arise largely due to the interaction of many loci of slight effect, as described by Fisher's infinitesimal model (Fisher, 1918). Therefore, the recent trend of using genome-wide marker data for predictive modelling of plant phenotypes rather than genetic dissection of traits (Jannink et al., 2010) would likely be well suited to germplasm panels such as those utilized herein.

Out of the traits that were analyzed in this study, heading date, as assessed using the GCTA model, came closest to identifying MTAs common to both the EP and YVP (**Figure 3.5a**). However, no SNPs exceeded the significance threshold for either panel. The SNP appearing closest to the top-right portion of the graph (i.e. the SNP with the highest combined -log(p) values for both the EP and YVP) is S7D_69293982. This pattern was not evident in the

FarmCPU data (**Figure 3.5b**). As FarmCPU implements a multi-locus model taking LD between SNPs into account, it indicates that a large proportion of SNPs appearing to show concordance between the EP and YVP GWA analyses in **Figure 3.5a** may in fact simply be in LD with one or a few QTLs of minor effect that are present in both panels, albeit without a strong enough association to surpass the multiple-testing significance thresholds for either panel. In the FarmCPU model, S7D_69293982 had a -log(p) value of 4.28, a value that is reasonably high, but nevertheless not qualifying as significant. In contrast, this SNP's -log(p) value in the EP (0.72) was unequivocally non-significant. A majority of traits analyzed using the GCTA algorithm exhibited bivariate -log(p) value distributions similar to that shown for the trait test weight in **Figure 3.5c**, with no SNPs demonstrating large -log(p) values in both panels.

As noted in the introduction, QTL effect estimates will be affected by both QTL-environment interaction and QTL-genetic background interaction, and in the case of the study detailed herein, these two interaction terms are confounded. Increasing the number of testing environments and using multiple mapping populations represents a simple, though "blind" approach to QTL validation, through which the confounding of QTL interaction effects may be lessened, but not resolved outright. As previously mentioned, experimental approaches have been developed specifically for evaluating QTL × environment interactions in linkage mapping studies (Boer et al., 2007; Malosetti et al., 2004; Moreau et al., 2004), as well as in GWA studies (Korte et al., 2012), though the latter was not used herein due to the large number of environments involved.

In addition, statistical methods exist for the detection of QTL × genetic background interaction; one straightforward approach being the implementation of a Monte Carlo sampling procedure in which a QTL mapping experiment is repeated many times, each time using only a

random portion of the mapping population (Beavis, 1994). The bootstrapping method used to calculate RMIP values for the FarmCPU MTAs in the present study is an example of such a Monte Carlo procedure. One of the two MTAs identified by both the GCTA and FarmCPU methods, S5D_39764565 (associated with the ERR estimate of test weight stability), did have the highest RMIP of any of the significant FarmCPU MTAs, at 0.66. However, the other MTA identified by both methods, S6A_416179127 (associated with plant height), had a RMIP of 0.28. Overall, there was no discernable relationship between a MTA's effect size (standardized in relation to the trait mean), p-value, or its RMIP value (data not shown). In the present study, a RMIP threshold of 0.05 was chosen based on a point of inflection in the density curve of the RMIP values of all FarmCPU-identified MTAs passing the chosen significance threshold (data not shown). However, in many cases it may be advisable to select a more stringent RMIP threshold to further limit identified MTAs to those that are more stable in relation to varying genetic backgrounds. For instance, Valdar et al. (2009) used a RMIP threshold of 0.25 to classify the most stable and robust MTAs.

Ultimately, while it is advantageous to detect patterns of QTL-environment and QTL-genetic background interaction, this information is of little practical use if it only identifies relatively unstable QTLs that cannot be deployed in applied breeding programs. The present study attempted to combine the identification of novel QTLs in two separate panels using GWA methods with a comparison of the results between panels. Searching for QTLs in multiple mapping populations or panels is a typical first step in validation, but is generally insufficient to test QTL effects for applied purposes due to confounding of the two forms of QTL/main effect interaction described above. In a GWA framework, a more elegant solution would be the use of a multi-trait mixed linear model (Korte et al., 2012), here considering the measurement of the

same trait in separate environments. However, the relatively large number of environments used herein could make this a burdensome task. An experimental design similar to that employed by Utz et al (2000) could be adapted for GWA studies, which would allow for bootstrapping of environments, genotypes or both in order to assess QTL stability across environments and genetic backgrounds. In addition, such a design would allow for the separation of QTL identification and effects estimation into different mapping populations. However, the time and resources required to establish a project of this scale were beyond the scope of the present study.

Meta-analyses of linkage mapping studies (Goffinet and Gerber, 2000) and GWA studies (Evangelou and Ioannidis, 2013; Zeggini and Ioannidis, 2009) offer modelling methods for standardizing and summarizing the results of multiple individual QTL identification experiments. However, the winter wheat research community has not had access to the same degree of standardized, curated data that has been available to human genetics researchers. In addition, the issue of "real world" validation of QTL effects remains. Pumphrey et al (2007) developed a method of creating multiple near-isogenic lines (NILs) to validate the effects of the *Fhb1* QTL. This method has the advantages of minimizing QTL-genetic background interaction for each NIL, and of allowing for the estimation of the effects of QTL introgression in a variety of genetic backgrounds used within an applied breeding program, making it somewhat more pragmatic. However, this technique requires a target QTL that is well characterized, and requires that this QTL can be reliably introgressed into various genetic backgrounds.

Conclusion

Previous studies have found that GWA analyses can be highly influenced by model inputs, and that QTLs identified via linkage mapping can be highly influenced by interaction

effects. The present study was not designed specifically to quantify the reliability of significant MTAs, but the lack of findings that were consistent across both panels included in the study suggests that substantial QTL-environment and QTL-genetic background effects are still present when utilizing assembled panels of genotypes for GWA analysis. In addition, it cannot be ruled out that future mapping of novel, reliable QTLs in winter wheat will simply require larger populations in order to resolve QTLs of smaller effect. Ultimately, in an applied plant breeding setting, GWA methods may be better suited to screening for loci of major effect, such as disease resistance genes, in order to enable subsequent introgression into breeding germplasm. In many cases, elite breeding lines may only be segregating for QTLs of more moderate effect, such that genome-wide marker data may be put to better use through genome-wide prediction of phenotypes.

References

Alpert, K.B., Grandillo, S., Tanksley, S.D., 1995. fw 2.2:a major QTL controlling fruit weight is common to both red- and green-fruited tomato species. Theor. Appl. Genet. 91, 994–1000. doi:10.1007/BF00223911

Alpert, K.B., Tanksley, S.D., 1996. High-resolution mapping and isolation of a yeast artificial chromosome contig containing fw2.2: A major fruit weight quantitative trait locus in tomato. Proc. Natl. Acad. Sci. 93, 15503–15507.

Anderson, J.A., Stack, R.W., Liu, S., Waldron, B.L., Fjeld, A.D., Coyne, C., Moreno-Sevilla, B., Fetch, J.M., Song, Q.J., Cregan, P.B., Frohberg, R.C., 2001. DNA markers for Fusarium head blight resistance QTLs in two wheat populations. Theor. Appl. Genet. 102, 1164–1168. doi:10.1007/s001220000509

Austin, D.F., Lee, M., 1998. Detection of Quantitative Trait Loci for Grain Yield and Yield Components in Maize across Generations in Stress and Nonstress Environments. Crop Sci. 38, 1296–1308. doi:10.2135/cropsci1998.0011183X003800050029x

Bates, D., Mächler, M., Bolker, B., Walker, S., 2015. Fitting Linear Mixed-Effects Models Using lme4. J. Stat. Softw. 67. doi:10.18637/jss.v067.i01

Beavis, W., 1994. The power and deceit of QTL experiments: lessons from comparative QTL studies, in: Proceedings of the Forty-Ninth Annual Corn and Sorghum Industry Research Conference. Chicago, IL, pp. 250–266.

Beavis, W.D., 1998. QTL analyses: power, precision, and accuracy. Mol. Dissection Complex Traits 1998, 145–162.

Bernardo, R., 2008. Molecular Markers and Selection for Complex Traits in Plants: Learning from the Last 20 Years. Crop Sci. 48, 1649–1664. doi:10.2135/cropsci2008.03.0131

Bian, Y., Yang, Q., Balint-Kurti, P.J., Wisser, R.J., Holland, J.B., 2014. Limits on the reproducibility of marker associations with southern leaf blight resistance in the maize nested association mapping population. BMC Genomics 15, 1068. doi:10.1186/1471-2164-15-1068

Boer, M.P., Wright, D., Feng, L., Podlich, D.W., Luo, L., Cooper, M., van Eeuwijk, F.A., 2007. A mixed-model quantitative trait loci (QTL) analysis for multiple-environment trial data using environmental covariables for QTL-by-environment interactions, with an example in maize. Genetics 177, 1801–1813. doi:10.1534/genetics.107.071068

Bradbury, P.J., Zhang, Z., Kroon, D.E., Casstevens, T.M., Ramdoss, Y., Buckler, E.S., 2007. TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23, 2633–2635. doi:10.1093/bioinformatics/btm308

Breseghello, F., Sorrells, M.E., 2006. Association Mapping of Kernel Size and Milling Quality in Wheat (Triticum aestivum L.) Cultivars. Genetics 172, 1165–1177. doi:10.1534/genetics.105.044586

Brown-Guedira, G.L., Badaeva, E.D., Gill, B.S., Cox, T.S., 1996. Chromosome substitutions of Triticum timopheevii in common wheat and some observations on the evolution of polyploid wheat species. Theor. Appl. Genet. 93, 1291–1298. doi:10.1007/BF00223462

Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., Lee, J.J., 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience 4, 7. doi:10.1186/s13742-015-0047-8

Chia, J.-M., Song, C., Bradbury, P.J., Costich, D., de Leon, N., Doebley, J., Elshire, R.J., Gaut, B., Geller, L., Glaubitz, J.C., Gore, M., Guill, K.E., Holland, J., Hufford, M.B., Lai, J., Li, M., Liu, X., Lu, Y., McCombie, R., Nelson, R., Poland, J., Prasanna, B.M., Pyhäjärvi, T., Rong, T., Sekhon, R.S., Sun, Q., Tenaillon, M.I., Tian, F., Wang, J., Xu, X., Zhang, Z., Kaeppler, S.M., Ross-Ibarra, J., McMullen, M.D., Buckler, E.S., Zhang, G., Xu, Y., Ware, D., 2012. Maize HapMap2 identifies extant variation from a genome in flux. Nat. Genet. N. Y. 44, 803–7. doi:10.1038/ng.2313

Cleveland, W.S., 1979. Robust Locally Weighted Regression and Smoothing Scatterplots. J. Am. Stat. Assoc. 74, 829–836. doi:10.1080/01621459.1979.10481038

Darvasi, A., Soller, M., 1995. Advanced intercross lines, an experimental population for fine genetic mapping. Genetics 141, 1199–1207.

Eberhart, S.A., Russell, W.A., 1966. Stability Parameters for Comparing Varieties. Crop Sci 6, 36–40. doi:10.2135/cropsci1966.0011183X000600010011x

Evangelou, E., Ioannidis, J.P.A., 2013. Meta-analysis methods for genome-wide association studies and beyond. Nat. Rev. Genet. Lond. 14, 379–89. doi:http://dx.doi.org.ezproxy.lib.vt.edu/10.1038/nrg3472

Fisher, R.A., 1918. The Correlation Between Relatives on the Supposition of Mendelian Inheritance. Trans. R. Soc. Edinb. 52, 399–433.

Frary, A., Nesbitt, T.C., Grandillo, S., Knaap, E., Cong, B., Liu, J., Meller, J., Elber, R., Alpert, K.B., Tanksley, S.D., 2000. fw2.2: a quantitative trait locus key to the evolution of tomato fruit size. Science 289, 85–88. doi:10.1126/science.289.5476.85

Gauch, H.G., 1988. Model Selection and Validation for Yield Trials with Interaction. Biometrics 44, 705–715. doi:10.2307/2531585

Glaubitz, J.C., Casstevens, T.M., Lu, F., Harriman, J., Elshire, R.J., Sun, Q., Buckler, E.S., 2014. TASSEL-GBS: A High Capacity Genotyping by Sequencing Analysis Pipeline. PLOS ONE 9, e90346. doi:10.1371/journal.pone.0090346

Goffinet, B., Gerber, S., 2000. Quantitative Trait Loci: A Meta-analysis. Genetics 155, 463–473.

Gore, M.A., Chia, J.-M., Elshire, R.J., Sun, Q., Ersoz, E.S., Hurwitz, B.L., Peiffer, J.A., McMullen, M.D., Grills, G.S., Ross-Ibarra, J., Ware, D.H., Buckler, E.S., 2009. A First-Generation Haplotype Map of Maize. Science 326, 1115–1117. doi:10.1126/science.1177837

Hirschhorn, J.N., Lohmueller, K., Byrne, E., Hirschhorn, K., 2002. A comprehensive review of genetic association studies. Genet. Med. Off. J. Am. Coll. Med. Genet. 4, 45–61. doi:10.1097/00125817-200203000-00002

Huang, M., Cabrera, A., Hoffstetter, A., Griffey, C., Sanford, D.V., Costa, J., McKendry, A., Chao, S., Sneller, C., 2016. Genomic selection for wheat traits and trait stability. Theor. Appl. Genet. 129, 1697–1710. doi:10.1007/s00122-016-2733-z

Huang, X., Paulo, M.-J., Boer, M., Effgen, S., Keizer, P., Koornneef, M., van Eeuwijk, F.A., 2011. Analysis of natural allelic variation in Arabidopsis using a multiparent recombinant inbred line population. Proc. Natl. Acad. Sci. U. S. A. 108, 4488–4493. doi:10.1073/pnas.1100465108

Ioannidis, J.P.A., 2014. Discussion: Why “An estimate of the science-wise false discovery rate and application to the top medical literature” is false. Biostat. Oxf. Engl. 15, 28-36-45. doi:10.1093/biostatistics/kxt036

Ioannidis, J.P.A., 2005. Why Most Published Research Findings Are False. PLoS Med. 2. doi:10.1371/journal.pmed.0020124

Jager, L.R., Leek, J.T., 2014. An estimate of the science-wise false discovery rate and application to the top medical literature. Biostat. Oxf. Engl. 15, 1–12. doi:10.1093/biostatistics/kxt007

Jannink, J.-L., Lorenz, A.J., Iwata, H., 2010. Genomic selection in plant breeding: from theory to practice. Brief. Funct. Genomics 9, 166–177. doi:10.1093/bfgp/elq001

Korte, A., Vilhjálmsson, B.J., Segura, V., Platt, A., Long, Q., Nordborg, M., 2012. A mixed-model approach for genome-wide association studies of correlated traits in structured populations. Nat. Genet. 44, 1066–1071. doi:10.1038/ng.2376

Kover, P.X., Valdar, W., Trakalo, J., Scarcelli, N., Ehrenreich, I.M., Purugganan, M.D., Durrant, C., Mott, R., 2009. A Multiparent Advanced Generation Inter-Cross to fine-map quantitative traits in Arabidopsis thaliana. PLoS Genet. 5, e1000551. doi:10.1371/journal.pgen.1000551

Kump, K.L., Bradbury, P.J., Wisser, R.J., Buckler, E.S., Belcher, A.R., Oropeza-Rosas, M.A., Zwonitzer, J.C., Kresovich, S., McMullen, M.D., Ware, D., Balint-Kurti, P.J., Holland, J.B., 2011. Genome-wide association study of quantitative resistance to southern leaf blight in the maize nested association mapping population. Nat. Genet. 43, 163–168. doi:10.1038/ng.747

Lande, R., Thompson, R., 1990. Efficiency of marker-assisted selection in the improvement of quantitative traits. Genetics 124, 743–756.

Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25, 1754–1760. doi:10.1093/bioinformatics/btp324

Liu, B.-H., 1998. Statistical genomics : linkage, mapping, and QTL analysis. CRC Press, Boca Raton.

Liu, X., Huang, M., Fan, B., Buckler, E.S., Zhang, Z., 2016. Iterative Usage of Fixed and Random Effect Models for Powerful and Efficient Genome-Wide Association Studies. PLOS Genet 12, e1005767. doi:10.1371/journal.pgen.1005767

Mackay, T.F.C., 2001. The Genetic Architecture of Quantitative Traits. Annu. Rev. Genet. 35, 303–339. doi:10.1146/annurev.genet.35.102401.090633

Malosetti, M., Ribaut, J.M., Vargas, M., Crossa, J., Eeuwijk, F.A. van, 2008. A multi-trait multi-environment QTL mixed model with an application to drought and nitrogen stress trials in maize (Zea mays L.). Euphytica 161, 241–257. doi:10.1007/s10681-007-9594-0

Malosetti, M., Voltas, J., Romagosa, I., Ullrich, S.E., Eeuwijk, F.A. van, 2004. Mixed models including environmental covariables for studying QTL by environment interaction. Euphytica 137, 139–145. doi:10.1023/B:EUPH.0000040511.46388.ef

Manolio, T.A., Collins, F.S., Cox, N.J., Goldstein, D.B., Hindorff, L.A., Hunter, D.J., McCarthy, M.I., Ramos, E.M., Cardon, L.R., Chakravarti, A., Cho, J.H., Guttmacher, A.E., Kong, A., Kruglyak, L., Mardis, E., Rotimi, C.N., Slatkin, M., Valle, D., Whittemore, A.S., Boehnke, M., Clark, A.G., Eichler, E.E., Gibson, G., Haines, J.L., Mackay, T.F.C., McCarroll, S.A., Visscher, P.M., 2009. Finding the missing heritability of complex diseases. Nature 461, 747–753. doi:10.1038/nature08494

Mauricio, R., 2001. Mapping quantitative trait loci in plants: uses and caveats for evolutionary biology. Nat. Rev. Genet. 2, 370–381. doi:10.1038/35072085

Melchinger, A.E., Utz, H.F., Schön, C.C., 1998. Quantitative Trait Locus (QTL) Mapping Using Different Testers and Independent Population Samples in Maize Reveals Low Power of QTL Detection and Large Bias in Estimates of QTL Effects. Genetics 149, 383–403.

Money, D., Gardner, K., Migicovsky, Z., Schwaninger, H., Zhong, G.-Y., Myles, S., 2015. LinkImpute: Fast and Accurate Genotype Imputation for Nonmodel Organisms. G3 GenesGenomesGenetics 5, 2383–2390. doi:10.1534/g3.115.021667

Moreau, L., Charcosset, A., Gallais, A., 2004. Use of trial clustering to study QTL x environment effects for grain yield and related traits in maize. Theor. Appl. Genet. 110, 92–105. doi:10.1007/s00122-004-1781-y

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

Nyquist, W.E., 1962. Differential Fertilization in the Inheritance of Stem Rust Resistance in Hybrids Involving a Common Wheat Strain Derived from Triticum Timopheevi. Genetics 47, 1109–1124.

Open Science Collaboration, 2015. Estimating the reproducibility of psychological science. Science 349, aac4716. doi:10.1126/science.aac4716

Price, A.H., 2006. Believe it or not, QTLs are accurate! Trends Plant Sci. 11, 213–216. doi:10.1016/j.tplants.2006.03.006

Pumphrey, M.O., Bernardo, R., Anderson, J.A., 2007. Validating the QTL for Fusarium Head Blight Resistance in Near-Isogenic Wheat Lines Developed from Breeding Populations. Crop Sci. 47, 200–206. doi:10.2135/cropsci2006.03.0206

R Core Team, 2015. R: A language and environment for statistical computing.

Rafalski, J.A., 2010. Association genetics in crop improvement. Curr. Opin. Plant Biol. 13, 174–180. doi:10.1016/j.pbi.2009.12.004

Rawat, N., Pumphrey, M.O., Liu, S., Zhang, X., Tiwari, V.K., Ando, K., Trick, H.N., Bockus, W.W., Akhunov, E., Anderson, J.A., Gill, B.S., 2016. Wheat Fhb1 encodes a chimeric

lectin with agglutinin domains and a pore-forming toxin-like domain conferring resistance to Fusarium head blight. Nat. Genet. 48, 1576–1580. doi:10.1038/ng.3706

Saghai-Maroof, M.A., Soliman, K.M., Jorgensen, R.A., Allard, R.W., 1984. Ribosomal DNA spacer-length polymorphisms in barley: mendelian inheritance, chromosomal location, and population dynamics. Proc. Natl. Acad. Sci. U. S. A. 81, 8014–8018.

Schön, C.C., Utz, H.F., Groh, S., Truberg, B., Openshaw, S., Melchinger, A.E., 2004. Quantitative Trait Locus Mapping Based on Resampling in a Vast Maize Testcross Experiment and Its Relevance to Quantitative Genetics for Complex Traits. Genetics 167, 485–498. doi:10.1534/genetics.167.1.485

Spindel, J., Begum, H., Akdemir, D., Virk, P., Collard, B., Redoña, E., Atlin, G., Jannink, J.-L., McCouch, S.R., 2015. Genomic selection and association mapping in rice (Oryza sativa): effect of trait genetic architecture, training population composition, marker number and statistical model on accuracy of rice genomic selection in elite, tropical rice breeding lines. PLoS Genet. 11, e1004982. doi:10.1371/journal.pgen.1004982

Sterken, R., Kiekens, R., Boruc, J., Zhang, F., Vercauteren, A., Vercauteren, I., Smet, L.D., Dhondt, S., Inzé, D., Veylder, L.D., Russinova, E., Vuylsteke, M., 2012. Combined linkage and association mapping reveals CYCD5;1 as a quantitative trait gene for endoreduplication in Arabidopsis. Proc. Natl. Acad. Sci. 109, 4678–4683. doi:10.1073/pnas.1120811109

Utz, H.F., Melchinger, A.E., Schön, C.C., 2000. Bias and Sampling Error of the Estimated Proportion of Genotypic Variance Explained by Quantitative Trait Loci Determined From Experimental Data in Maize Using Cross Validation and Validation With Independent Samples. Genetics 154, 1839–1849.

Valdar, W., Holmes, C.C., Mott, R., Flint, J., 2009. Mapping in Structured Populations by Resample Model Averaging. Genetics 182, 1263–1277. doi:10.1534/genetics.109.100727

Waldron, B.L., Moreno-Sevilla, B., Anderson, J.A., Stack, R.W., Frohberg, R.C., 1999. RFLP Mapping of QTL for Fusarium Head Blight Resistance in Wheat. Crop Sci. 39, 805–811. doi:10.2135/cropsci1999.0011183X003900030032x

Wallace, J.G., Zhang, X., Beyene, Y., Semagn, K., Olsen, M., Prasanna, B.M., Buckler, E.S., 2016. Genome-wide Association for Plant Height and Flowering Time across 15 Tropical Maize Populations under Managed Drought Stress and Well-Watered Conditions in Sub-Saharan Africa. Crop Sci. 56, 2365–2378. doi:10.2135/cropsci2015.10.0632

Weir, B.S., Cockerham, C.C., 1984. Estimating F-Statistics for the Analysis of Population Structure. Evolution 38, 1358–1370. doi:10.2307/2408641

Xu, S., 2003. Theoretical Basis of the Beavis Effect. Genetics 165, 2259–2268.

Xu, Y., Li, P., Yang, Z., Xu, C., 2016. Genetic mapping of quantitative trait loci in crops. Crop J. 5, 175–184. doi:10.1016/j.cj.2016.06.003

Yan, W., Kang, M.S., 2003. GGE biplot analysis a graphical tool for breeders, geneticists, and agronomists. CRC Press, Boca Raton, Fla.

Yu, J., Holland, J.B., McMullen, M.D., Buckler, E.S., 2008. Genetic Design and Statistical Power of Nested Association Mapping in Maize. Genetics 178, 539–551. doi:10.1534/genetics.107.074245

Zadoks, J.C., Chang, T.T., Konzak, C.F., 1974. A decimal code for the growth stages of cereals. Weed Res. 14, 415–421. doi:10.1111/j.1365-3180.1974.tb01084.x

Zeggini, E., Ioannidis, J.P.A., 2009. Meta-analysis in genome-wide association studies. Pharmacogenomics 10, 191–201. doi:10.2217/14622416.10.2.191

Zheng, X., Levine, D., Shen, J., Gogarten, S.M., Laurie, C., Weir, B.S., 2012. A high-performance computing toolset for relatedness and principal component analysis of SNP data. Bioinformatics 28, 3326–3328. doi:10.1093/bioinformatics/bts606

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

Table 3.1: Trait descriptive statistics, variance components, and entry mean heritability for genotypes in the elite panel and yield validation panel

			Descriptive Statistics§						Variances and Heritability§			
Panel †	**Trait ‡**	**Units**	**n Envs**	**min**	**mean**	**max**	**SD**	**CV**	**σ^2 G**	**σ^2 E**	**σ^2 ε**	**H^2**
EP	HD	Julian days (Jan1)	14	90	125	150	14.05	0.11	5.65	203	3.95	0.95
	HGT	cm	12	54.23	89.49	155.1	11.62	0.13	42.23	81.59	21.9	0.96
	TW	g L^{-1}	12	503.6	756	880.4	32.75	0.04	171	720.7	218.4	0.90
	YLD	kg ha^{-1}	12	1770	5447	9272	1197	0.22	99874	1075033	321971	0.79
	AMMI_TW	-	12	2.785	6.787	28.09	2.759	0.41	-	-	-	-
	AMMI_YLD	-	12	0.266	10.64	33.12	5.823	0.55	-	-	-	-
	ERR_TW	-	12	0.525	1.019	1.742	0.198	0.19	-	-	-	-
	ERR_YLD	-	12	0.364	0.978	1.544	0.179	0.18	-	-	-	-
YVP	HD	Julian days (Jan1)	9	120	136	151	6.34	0.05	2.43	49.4	1.76	0.94
	HGT	cm	11	56.39	86.1	128.3	8.96	0.10	31.2	30.45	19.39	0.95
	TW	g L^{-1}	8	572.2	727	940.7	38.41	0.05	183.6	1069	319.6	0.87
	YLD	kg ha^{-1}	12	1877	4838	7943	969	0.20	70554	688788	214884	0.80
	AMMI_TW	-	8	0.413	3.120	9.265	1.623	0.52	-	-	-	-
	AMMI_YLD	-	12	17.38	43.43	82.67	11.03	0.25	-	-	-	-
	ERR_TW	-	8	0.195	0.993	2.083	0.272	0.27	-	-	-	-
	ERR_YLD	-	12	0.487	0.999	1.574	0.202	0.20	-	-	-	-

† EP elite panel; YVP yield validation panel

‡ HD heading date; HGT plant height; TW test weight; YLD grain yield; AMMI_TW AMMI estimate of test weight stability; AMMI_YLD AMMI estimate of grain yield stability; ERR_TW Eberhart Russel regression estimate of test weight stability; ERR_YLD Eberhart Russel regression estimate of grain yield stability

§ SD standard deviation; CV coefficient of variation; σ^2 G genotypic variance; σ^2 E environmental variance; σ^2 ε residual variance; H^2 entry-mean heritability

Table 3.2: Significant marker-trait associations identified by the GCTA-LOCO algorithm†

Panel‡	Trait§	n PCs	Chrom	SNP¶	Ref	Alt	MAF	P-Value	Effect	Units
EP	HGT	5	4D	S4D_25419579	C	T	0.079926	7.44E-06	3.50881	cm
EP	HGT	5	6A	S6A_416179127 #	A	G	0.421933	7.30E-06	-1.57832	cm
EP	ERR_TW	5	5D	S5D_39764565 #	C	G	0.052632	5.64E-06	0.145293	-
YVP	ERR_TW	5	7A	S7A_709743594	A	T	0.05137	3.86E-06	0.2306	-

† Significance thresholds for a 95% confidence level were determined empirically by performing permutation testing for 1,000 repetitions per trait.

Column label abbreviations: n PCs, number of principle components included in analysis; Chrom chromosome; SNP single nucleotide polymorphism; Ref, reference allele; Alt, alternate allele; MAF, minor allele frequency; Effect, mean value of lines containing alternate allele minus mean of lines containing reference allele

‡ EP elite panel; YVP yield validation panel

§ HGT plant height; TW test weight; "AMMI" and "ERR" indicate AMMI model and Eberhart and Russel Regression estimates of trait stability, respectively

¶ SNP name includes physical position on chromosome

SNP also identified as significant using FarmCPU

Table 3.3: Significant marker-trait associations identified by the FarmCPU algorithm[†]

Panel[‡]	Trait[§]	n PCs	Chrom	SNP[¶]	MAF	P-Value	Effect	Units	RMIP
EP	HD	5	2A	S2A_49189737	0.421933086	6.90E-07	0.426915757	Julian days (Jan1)	0.07
EP	HD	5	2D	S2D_73207210	0.43866171	3.12E-06	-0.40499538	Julian days (Jan1)	0.22
YNVP	HD	5	3B	S3B_23975004	0.110169492	2.51E-06	-0.376602915	Julian days (Jan1)	0.26
YNVP	HD	5	5A	S5A_579626304	0.227118644	2.73E-08	-0.381922012	Julian days (Jan1)	0.09
EP	HD	5	5B	S5B_550563262	0.42936803	5.57E-09	0.537947102	Julian days (Jan1)	0.37
YNVP	HD	5	6A	S6A_68995191	0.484745763	1.39E-12	0.410135591	Julian days (Jan1)	0.44
EP	HGT	5	3A	S3A_699195908	0.074349442	9.20E-07	2.393005081	cm	0.23
YNVP	HGT	5	3B	S3B_782374448	0.447457627	5.66E-07	0.597132427	cm	0.22
EP	HGT	5	4A	S4A_617431757	0.32527881	2.65E-10	-1.632926237	cm	0.45
EP	HGT	5	4B	S4B_25886179	0.06133829	2.00E-06	2.260203461	cm	0.53
YNVP	HGT	5	4B	S4B_662927988	0.169491525	1.82E-07	1.475843966	cm	0.63
EP	HGT	5	6A	S6A_416179127 [#]	0.421933086	4.75E-09	-1.482115269	cm	0.28
EP	HGT	5	6B	S6B_162549774	0.197026022	7.91E-10	-1.819714331	cm	0.44
YNVP	HGT	5	7A	S7A_727298781	0.13220339	1.50E-07	-1.49728356	cm	0.42
EP	TW	5	1B	S1B_667448097	0.130111524	6.84E-12	-4.738346122	$g\ L^{-1}$	0.32
EP	TW	5	2B	S2B_3264617	0.288104089	7.08E-08	2.473327564	$g\ L^{-1}$	0.4
EP	TW	5	2B	S2B_33519842	0.07063197	9.27E-07	-3.723769347	$g\ L^{-1}$	0.13
EP	TW	5	7A	S7A_694870296	0.072490706	3.96E-07	-5.331670075	$g\ L^{-1}$	0.23
EP	ERR_TW	2	3A	S3A_15558351	0.443609023	1.10E-07	-0.040385479	-	0.06
EP	ERR_TW	2	4B	S4B_289037575	0.278195489	4.28E-07	0.047275099	-	0.34
EP	ERR_TW	2	5B	S5B_42564147	0.088345865	1.09E-06	0.087106602	-	0.13
EP	ERR_TW	2	5D	S5D_39764565 [#]	0.052631579	1.21E-10	0.145277609	-	0.66
EP	ERR_TW	2	7B	S7B_653950790	0.434210526	4.09E-10	0.05487766	-	0.15
EP	AMMI_YLD	2	3D	S3D_565086710	0.172932331	1.01E-07	1.73282742	-	0.49
EP	AMMI_YLD	2	5D	S5D_543090768	0.244360902	1.78E-08	1.612227799	-	0.41
EP	AMMI_YLD	2	6A	S6A_99931123	0.172932331	1.81E-06	1.388438128	-	0.12
EP	ERR_YLD	2	4B	S4B_415342383	0.163533835	7.20E-09	0.063210363	-	0.22
EP	ERR_YLD	2	6A	S6A_569599082	0.12406015	1.97E-07	-0.065043739	-	0.17

Panel‡	Trait§	n PCs	Chrom	SNP¶	MAF	P-Value	Effect	Units	RMIP
EP	ERR_YLD	2	7B	S7B_641908294	0.182330827	1.99E-06	0.04324683	-	0.06
EP	ERR_YLD	2	7B	S7B_724034080	0.411654135	6.64E-07	0.036787807	-	0.09
EP	ERR_YLD	2	7D	S7D_582607444	0.080827068	2.31E-06	0.061460157	-	0.1

† Significance thresholds for a 95% confidence level were determined empirically by performing permutation testing for 1,000 repetitions per trait.

Column label abbreviations: n PCs, number of principle components included in analysis; Chrom chromosome; SNP single nucleotide polymorphism; Ref, reference allele; Alt, alternate allele; MAF, minor allele frequency; Effect, mean value of lines containing alternate allele minus mean of lines containing reference allele

‡ EP elite panel; YVP yield validation panel

§ HGT plant height; TW test weight; “AMMI” and “ERR” indicate AMMI model and Eberhart and Russel Regression estimates of trait stability, respectively

¶ SNP name includes physical position on chromosome

SNP also identified as significant using FarmCPU

Figure 3.1a: Portion of variance explained by each genotypic principal component for genotypes in the elite panel (EP) and yield validation panel (YVP)

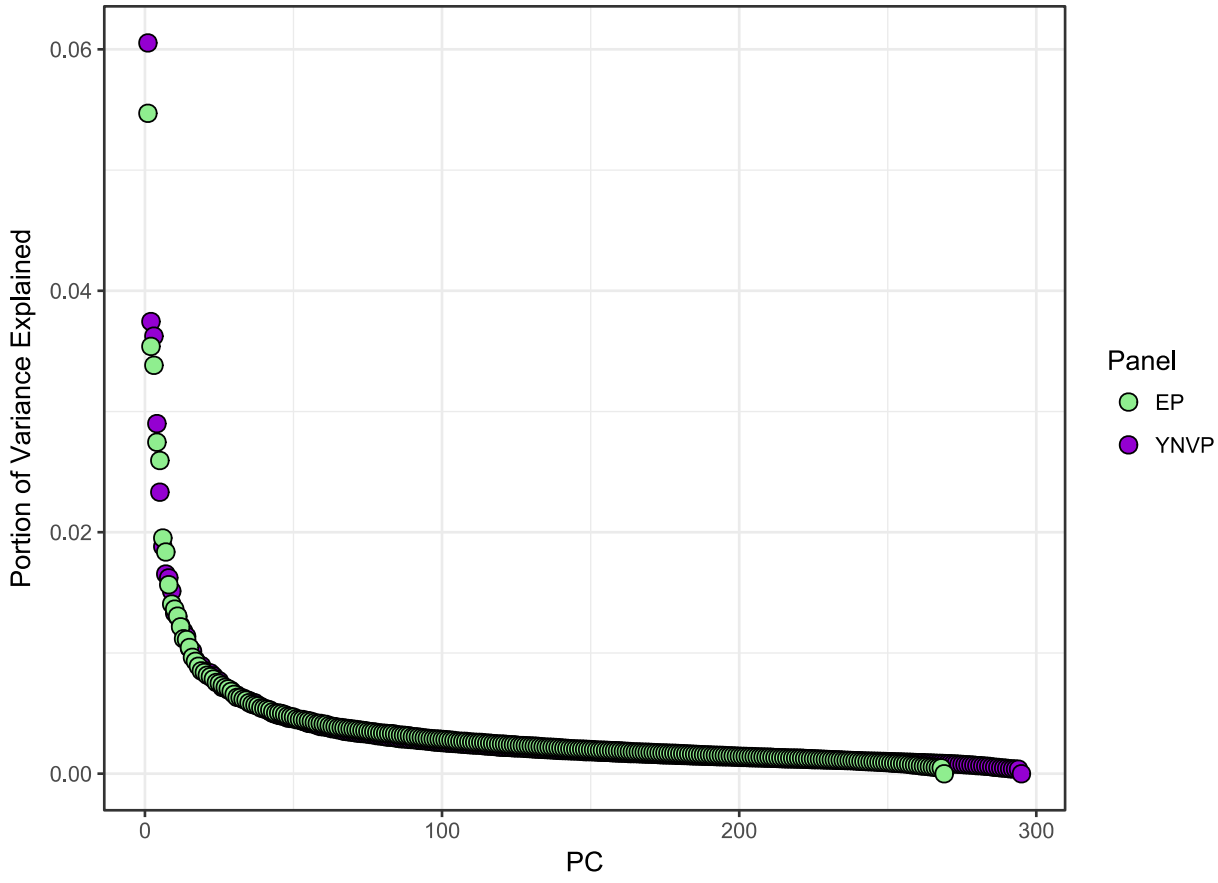


Figure 3.1b: Cumulative Portion of variance explained by each genotypic principal component for genotypes in the elite panel (EP) and yield validation panel (YVP)

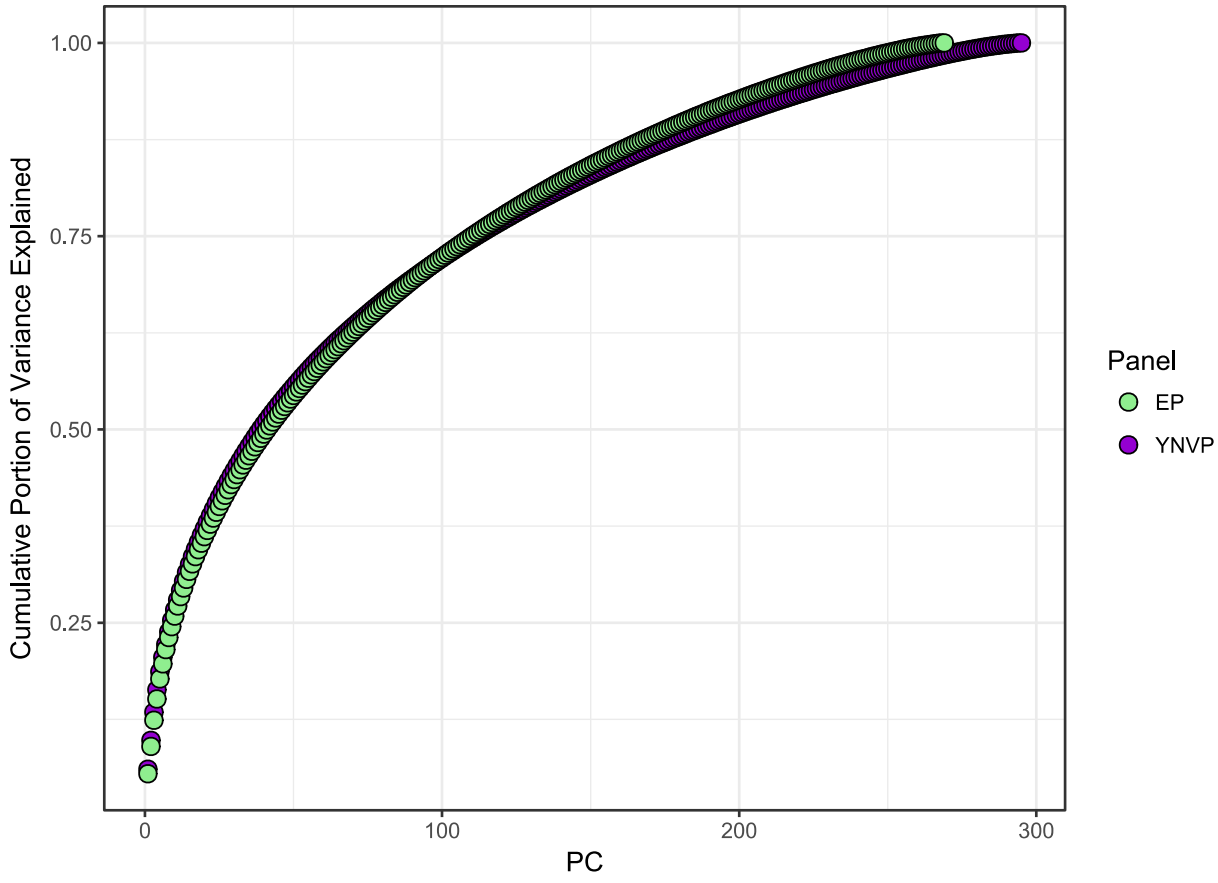


Figure 3.2: Biplot of the first two genotypic principal components of the combined elite panel and yield validation panel data, with coloring by presence or absence of the *Sr36* stem rust resistance gene located on chromosome 2B. The majority of genotypes with *Sr36* data are in the yield validation panel, while those without data are in the elite panel.

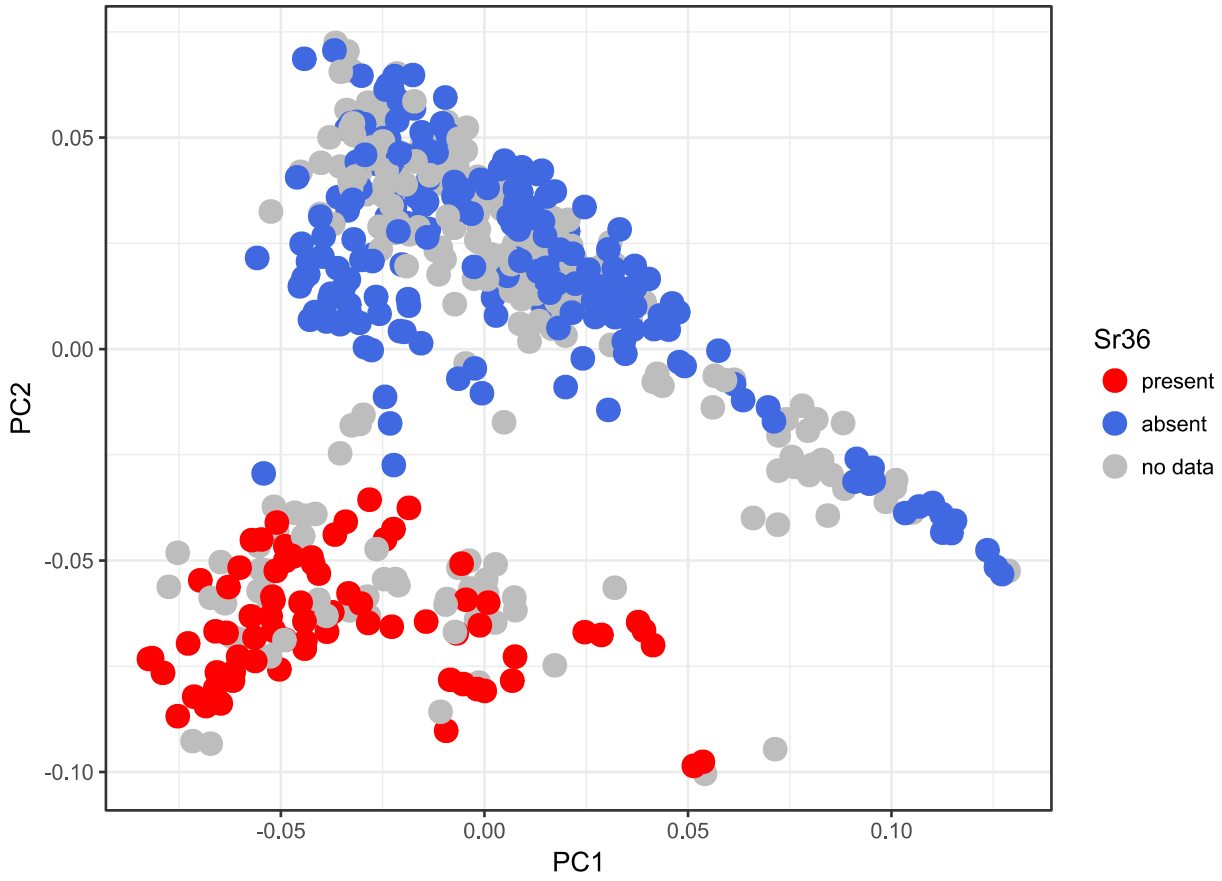


Figure 3.3: Number of SNPs per chromosome in the processed, filtered GBS genotypic dataset. This figure applies to both the elite panel and the yield validation panel, as genotypic data was generated for these two panels jointly.

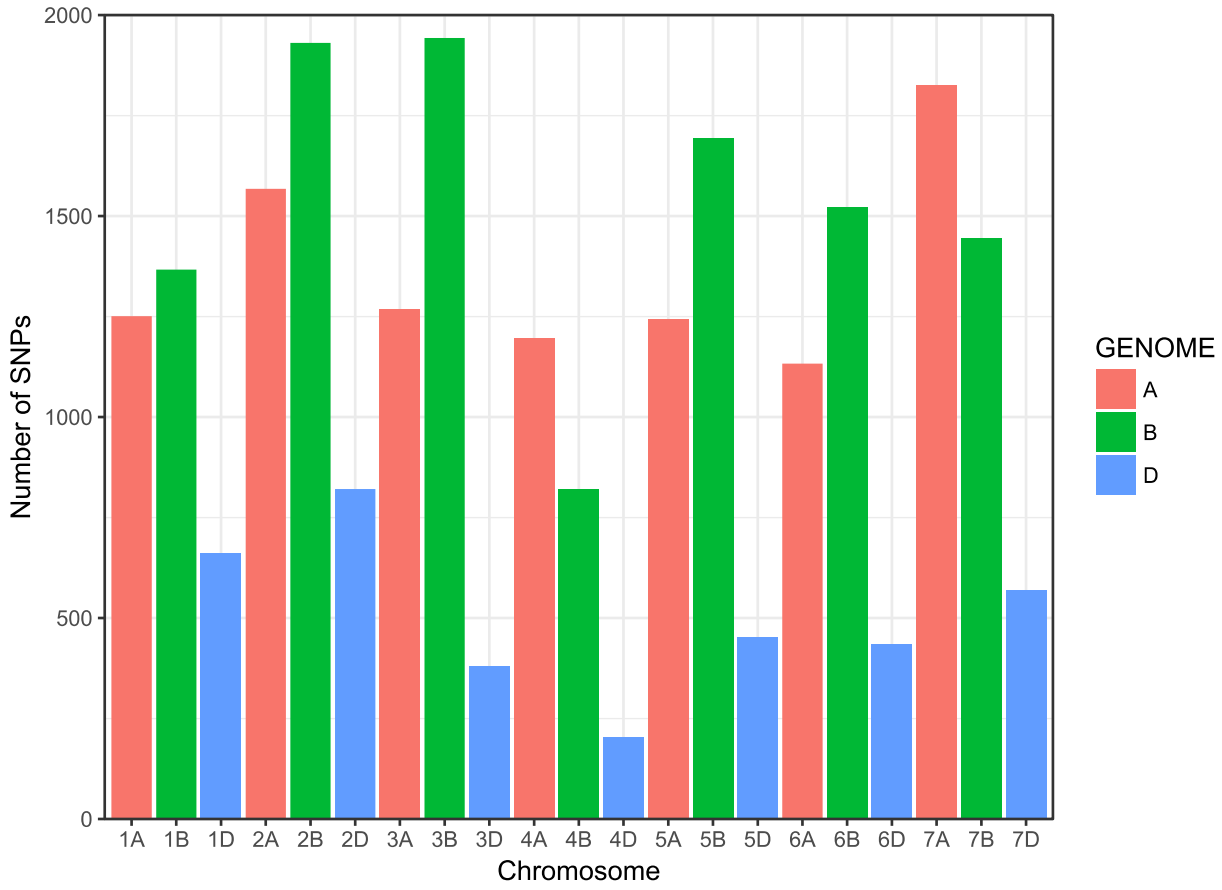


Figure 3.4a: Linkage disequilibrium by genome for the elite panel (EP) and yield validation panel (YVP). Lines represent second-degree LOESS curves fit to 20,000 randomly-selected intra-chromosomal pairwise genotypic r^2 estimates pooled from chromosomes in each sub-genome. Horizontal line corresponds to the 98th percentile of pairwise r^2 estimates for non-linked (i.e. inter-chromosomal) SNPs averaged between the two panels.

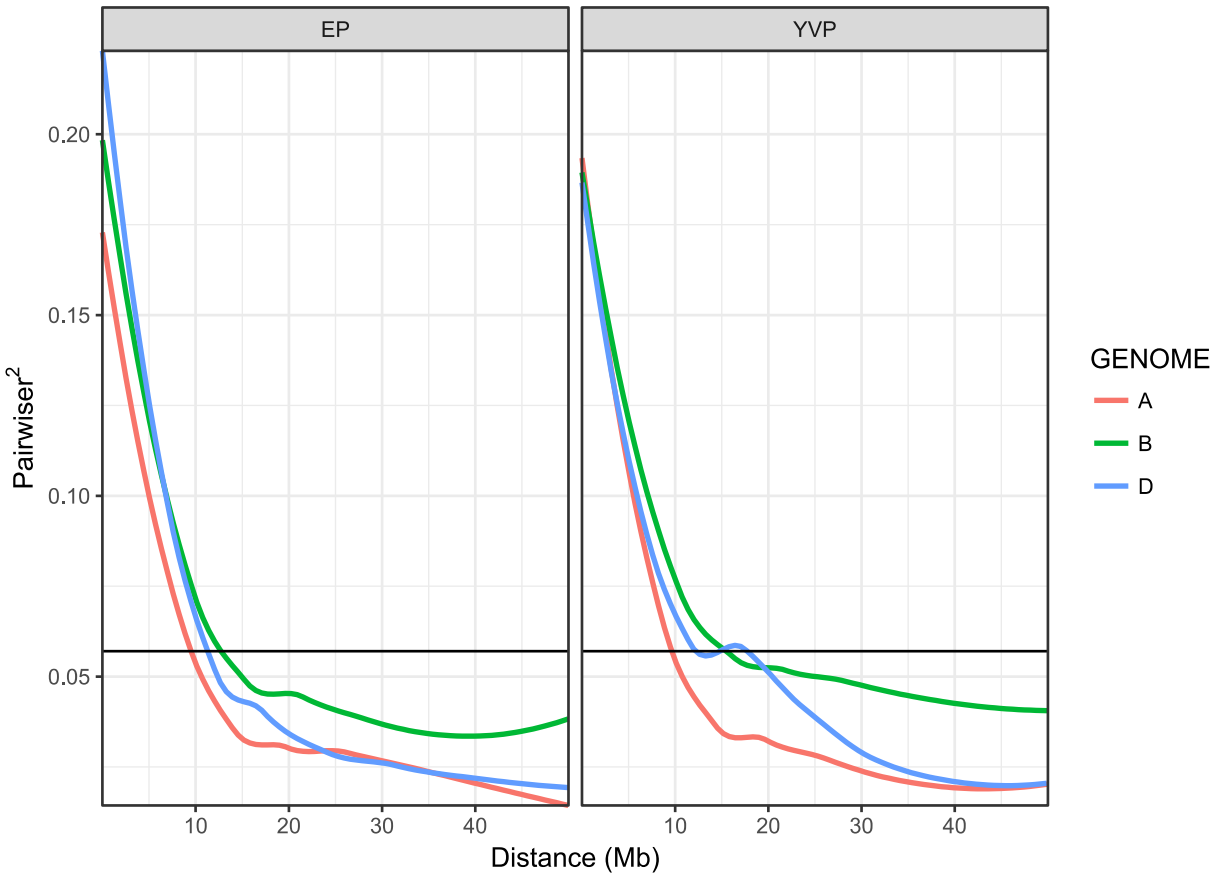


Figure 3.4b: Linkage disequilibrium in each chromosome of the B genome for the elite panel (EP) and yield validation panel (YVP). Lines represent second-degree LOESS curves fit to 20,000 randomly-selected intra-chromosomal pairwise genotypic r^2 estimates from each chromosome in the B genome. Horizontal line corresponds to the 98th percentile of pairwise r^2 estimates for non-linked (i.e. inter-chromosomal) SNPs averaged between the two panels.

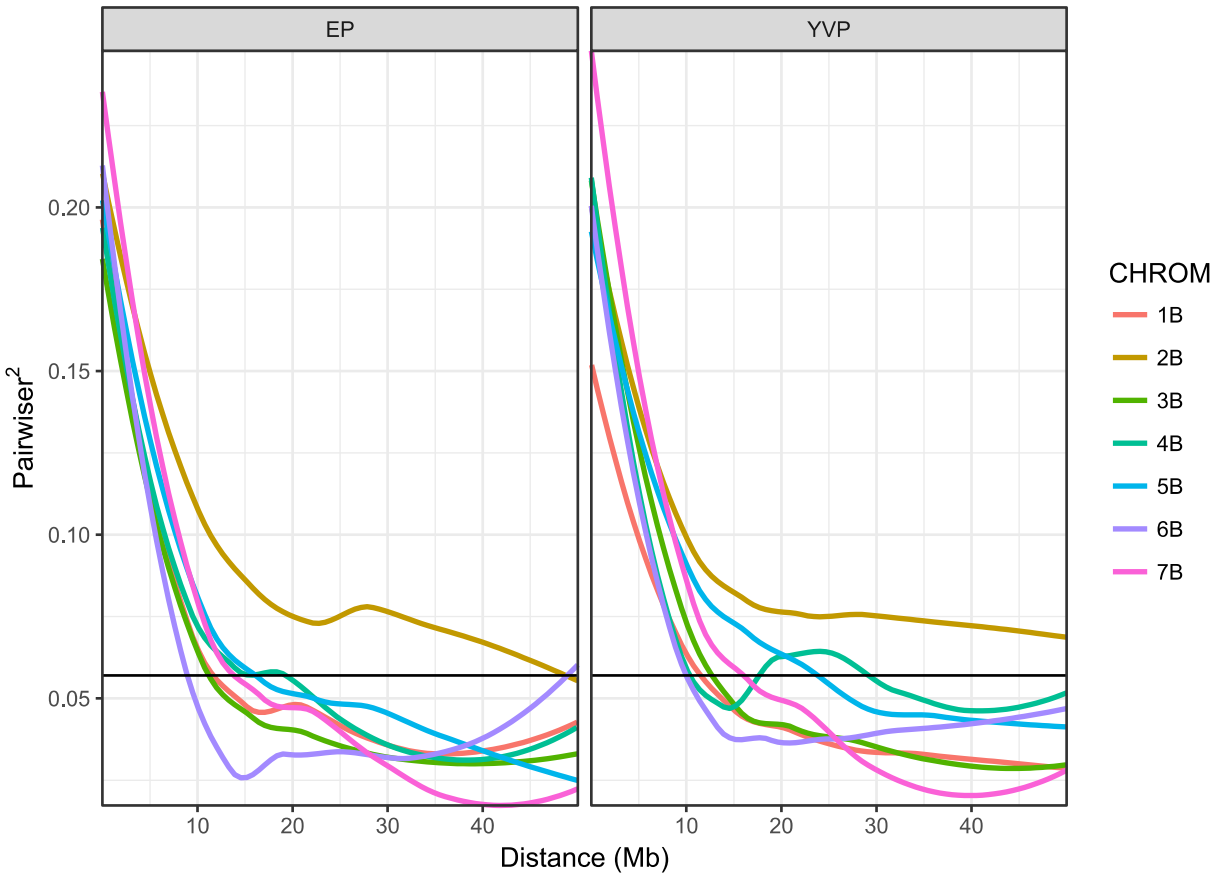


Figure 3.5a: Regression of GCTA-LOCO-generated -log transformed p-values of SNPs for the trait heading date in the elite panel (EP) against -log transformed p-values for the same SNPs in the yield validation panel (YVP). Diagonal line represents identity; horizontal and vertical dashed lines represent the significance thresholds as determined by permutation for the YVP and EP, respectively.

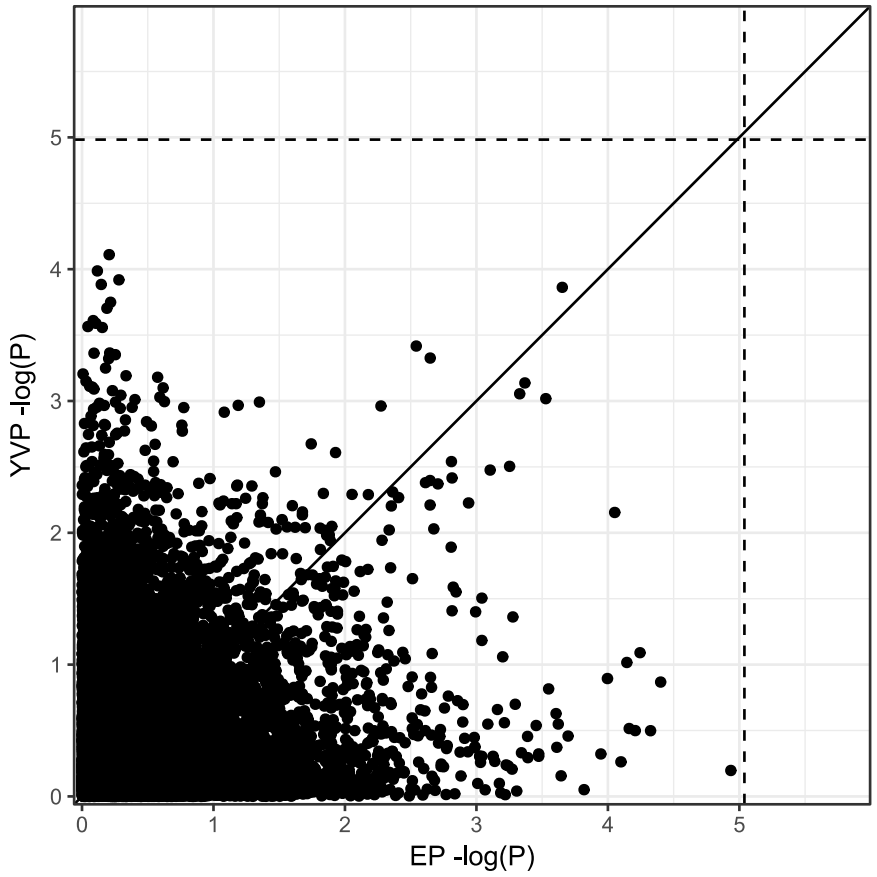


Figure 3.5b: Regression of FarmCPU-generated -log transformed p-values of SNPs for the trait heading date in the elite panel (EP) against -log transformed p-values for the same SNPs in the yield validation panel (YVP). Diagonal line represents identity; horizontal and vertical dashed lines represent the significance thresholds as determined by permutation for the YVP and EP, respectively.

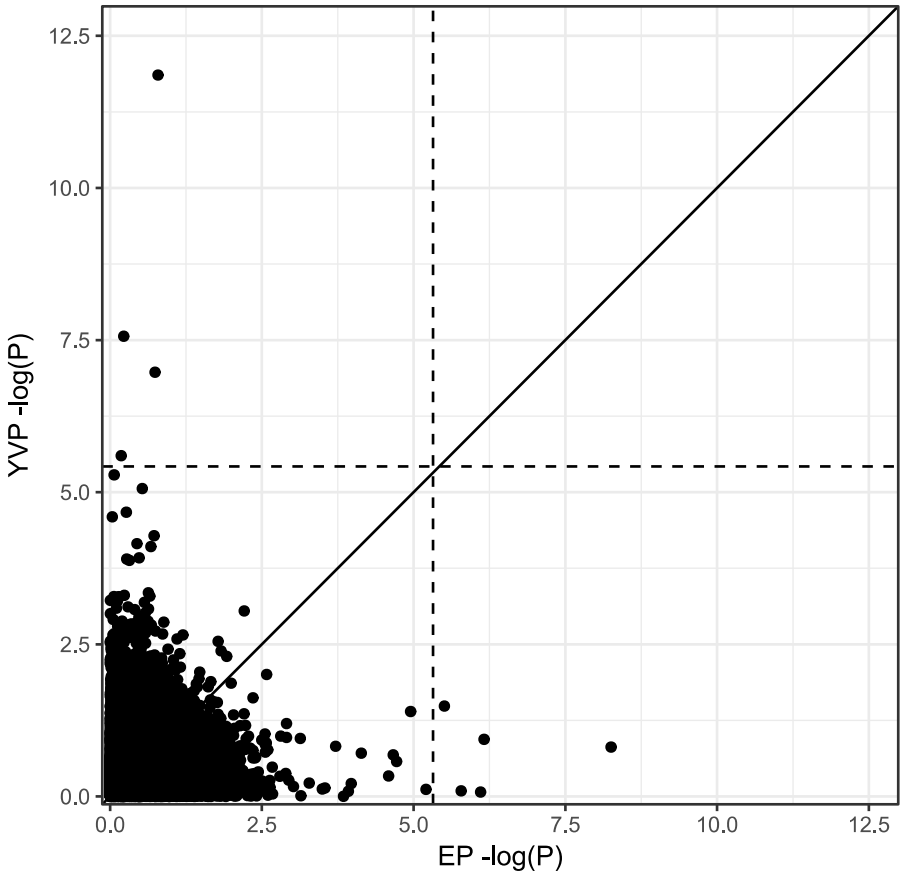
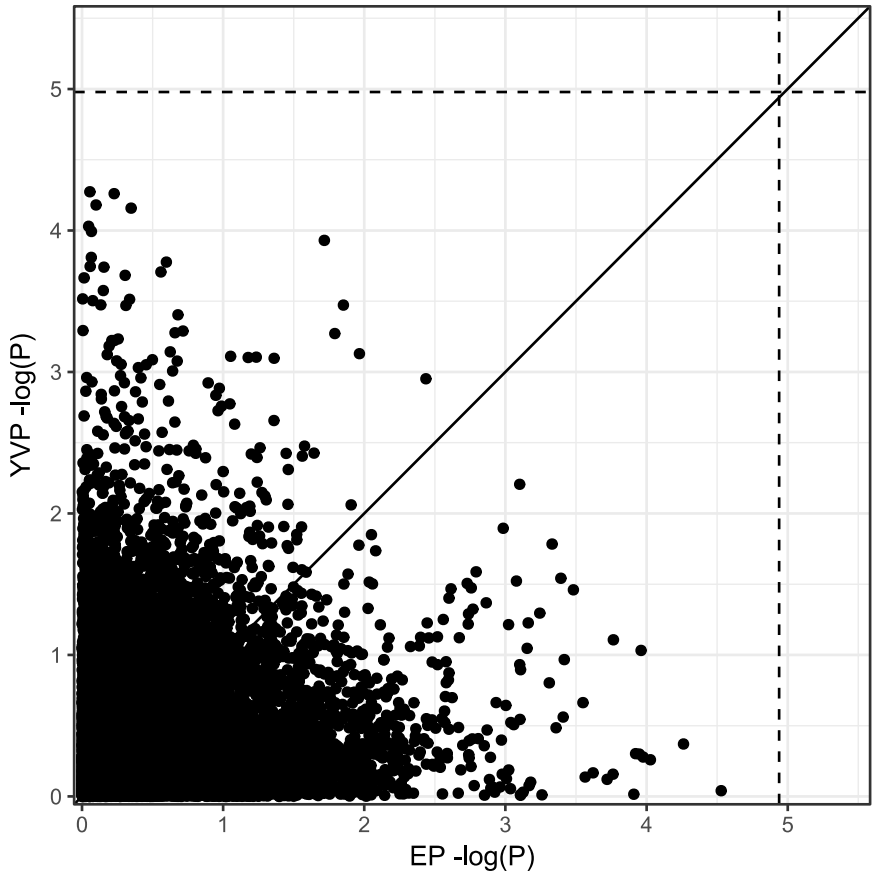


Table 3.5c: Regression of GCTA-LOCO-generated -log transformed p-values of SNPs for the trait test weight in the elite panel (EP) against -log transformed p-values for the same SNPs in the yield validation panel (YVP). Diagonal line represents identity; horizontal and vertical dashed lines represent the significance thresholds as determined by permutation for the YVP and EP, respectively.



APPENDIX A: Chemical and fertilizer applications, planting dates, and harvest dates for all environments included in chapters I & II

Table A.1: Chemical and fertilizer applications, planting dates, and harvest dates for all environments included in chapters I & II

Environment	Event	Application Rate	Date
Blacksburg, VA; 2013-2014 season	Fall Granular Fertilizer Application	34-52-67 (N-P-K) kg ha^{-1}	9/25/2013
	Planting Date	-	9/30/2013
	Harmony® Extra SG Herbicide Treatment	21 g a.i. ha^{-1}	11/21/2013
	Spring N Application GS25	28 kg ha^{-1}	2/11/2014
	Palisade® 2EC Growth Regulator	90 g a.i. ha^{-1}	4/1/2014
	Tilt® Fungicide Application	62.5 g a.i. ha^{-1}	4/1/2014
	Spring N Application GS30	67 kg ha^{-1}	4/2/2014
	Harmony® Extra SG Herbicide Treatment	26 g a.i. ha^{-1}	4/2/2014
	Tilt® Fungicide Application	125 g a.i. ha^{-1}	4/25/2014
	Prosaro® Fungicide Application	207 g a.i. ha^{-1}	5/13/2014
	Tilt® Fungicide Application	62.5 g a.i. ha^{-1}	5/27/2014
	Harvest	-	6/30/2014
Warsaw, VA; 2013-2014 season	Fall Granular Fertilizer Application	34-67-67-5.6 (N-P-K-S) kg ha^{-1}	10/18/2013
	Planting Date	-	10/22/2013
	Winter N Application	28 kg ha^{-1}	11/21/2013
	Spring N Application GS25	34 kg ha^{-1}	3/1/2014
	Starane® broadleaf herbicide treatment	159 g a.e. ha^{-1}	3/11/2014
	Harmony® Extra SG Herbicide Treatment	26 g a.i. ha^{-1}	3/11/2014
	Spring N Application GS30	67.3 kg ha^{-1}	4/4/2014
	Palisade® 2EC Growth Regulator	90 g a.i. ha^{-1}	4/4/2014
	Tilt® Fungicide Application	125 g a.i. ha^{-1}	4/10/2014
	Prosaro® Fungicide Application	245 g a.i. ha^{-1}	5/8/2014
	Tilt® Fungicide Application	125 g a.i. ha^{-1}	5/8/2014
	Harvest	-	6/24/2014
Blacksburg, VA 2014-2015	Fall Granular Fertilizer Application	34-56-56 (N-P-K) kg ha^{-1}	9/25/2014
	Planting Date	-	9/22/2014
	Harmony Extra SG Herbicide Treatment	26 g a.i. ha^{-1}	10/31/2014
	Spring N Application GS25	34 kg ha^{-1}	3/18/2015
	Spring N Application GS30	56 kg ha^{-1}	4/8/2015
	Harmony Extra SG Herbicide Treatment	26 g a.i. ha^{-1}	4/8/2015
	Manni-Plex® 3.8% Boron Application	0.95 L ha^{-1}	4/11/2015
	Crop Smart 16.7% Zinc Application	0.95 L ha^{-1}	4/11/2015
	Tilt® Fungicide Application	62.5 g a.i. ha^{-1}	4/13/2015

	Palisade® 2EC Growth Regulator	72 g a.i. ha^{-1}	4/13/2015
	Prosaro® Fungicide Application	207 g a.i. ha^{-1}	5/8/2015
	Tilt® Fungicide Application	62.5 g a.i. ha^{-1}	5/11/2015
	Harvest	-	6/29/2015
Warsaw, VA 2014-2015	Fall Granular Fertilizer Application	34-67-67-5.6 (N-P-K-S) kg ha^{-1}	10/6/2014
	Planting Date	-	10/21/2014
	Winter N application	34 kg ha^{-1}	12/15/2014
	Spring N Application GS25	34 kg ha^{-1}	2/6/2015
	Starane® broadleaf herbicide treatment	159 g a.e. ha^{-1}	3/24/2015
	Harmony® Extra SG Herbicide Treatment	26 g a.i. ha^{-1}	3/24/2015
	Spring N Application GS30	67.3 kg ha^{-1}	4/5/2015
	Palisade® 2EC Growth Regulator	72 g a.i. ha^{-1}	4/6/2015
	Tilt® Fungicide Application	125 g a.i. ha^{-1}	4/10/2015
	Manni-Plex® 3.8% Boron Application	1.9 L ha^{-1}	4/12/2015
	Crop Smart 16.7% Zinc Application	0.59 L ha^{-1}	4/12/2015
	Prosaro® Fungicide Application	245 g a.i. ha^{-1}	5/14/2015
	Tilt® Fungicide Application	125 g a.i. ha^{-1}	5/14/2015
	Harvest	-	6/22/2015

APPENDIX B: List of genotypes included in chapters I & II

The table below lists all genotypes included in chapters I & II, with their state of origin and pedigree (if known)

Table B.1: List of genotypes included in chapters I & II

GENOTYPE	ORIGIN	PEDIGREE
BESS	Virginia	MO11769/MADISON
BRANSON	Virginia	PIO2737W/891-4584A
DH11SRW061-5	Virginia	LA01139D-56-1/VA08W-294//TRIBUTE]
DH11SRW063-1	Virginia	VA08W-294//TRIBUTE]/GA08279-G3-G1-G8
DH11SRW063-14	Virginia	VA08W-294//TRIBUTE]/GA08279-G3-G1-G8
DH11SRW063-2	Virginia	VA08W-294//TRIBUTE]/GA08279-G3-G1-G8
DH11SRW065-8	Virginia	AGS2038/VA08W-294//TRIBUTE]
DH11SRW066-2	Virginia	PIONEER25R32/VA08W-176
DH11SRW067-7	Virginia	GA00067-8E35/IL04-8445/IL97-3578
DH11SRW068-1	Virginia	IL04-8445/IL97-3578]/SHIRLEY
DH11SRW070-14	Virginia	GA00067-8E35/SHIRLEY
DH11SRW070-28	Virginia	GA00067-8E35/SHIRLEY
FEATHERSTONE VA258	Virginia	VA05W-258=VA98W-130/GORE)//CK9835/SS520
FFR 555W	Virginia	COKER76-35/3/DOUBLECROP//VA72-54-14/VA76-52-12//COKER76-35/3/VA76-52-24/COKER65-20/ARTHUR/KAVKAZ//COKER65-20/ARTHUR
IL00-8530	Illinois	IL89-1687//IL90-6364/IL93-2489
IL11-33669-1	Illinois	01-11934/VA-FE24-13
IL12-10819	Illinois	04-7942/02-19463
IL12-10971	Illinois	04-8445/00-8530
IL12-12368	Illinois	04-10721/P0172A1-12-1
IL12-13178	Illinois	04-11003/04-7874
IL12-14458	Illinois	06-7155/M99-3098
IL12-14670	Illinois	06-7155/06-14303
IL12-15121	Illinois	06-8209/M99-3098
IL12-15424	Illinois	06-8209/02-18228
IL12-16207	Illinois	06-12909/79-002T
IL12-17334	Illinois	06-14303/06-11448
IL12-1862	Illinois	M03-3616-11B/02-18228
IL12-18670	Illinois	MO050699/02-19463//00-8061
IL12-20965	Illinois	97-1828/00-8641//06-14303
IL12-24232	Illinois	01-16170/OH02-12686//06-14303
IL12-25066	Illinois	01-16170/01-6262//06-13878
IL12-25758	Illinois	01-34159/79-002T//06-7155
IL12-26707	Illinois	02-7735/00-8530//00-8530
IL12-27938	Illinois	04-10721/99-26442//04-7874
IL12-29068	Illinois	05-28110/04-10741//PIO25W60
IL12-31210	Illinois	94-6727/01-34159//01-11934
IL12-3229	Illinois	P0179A1-17/79-002T

GENOTYPE	ORIGIN	PEDIGREE
IL12-37468	Illinois	01-34159/04-8445
IL12-3748	Illinois	P03112A1-7-14/04-10145
IL12-4448	Illinois	97-1828/MO061041
IL12-5169	Illinois	97-1828/06-14303
IL12-6376	Illinois	00-8641/06-11448
IL12-8119	Illinois	01-11934/02-18228
IL12-8643	Illinois	01-34159/02-18228
IL12-9329	Illinois	02-18228/MO061041
IL12-9637	Illinois	02-18228/04-11003
IL13-10367	Illinois	06-8223/04-10741
IL13-11145	Illinois	06-9607/06-18051
IL13-11556	Illinois	06-13721/PO128A1-22-22
IL13-12179	Illinois	06-14262/06-12109
IL13-12333	Illinois	06-16639/02-18228
IL13-15308	Illinois	VA03W-409/04-8445//068223
IL13-19403	Illinois	01-16170/VA03W-409//02-18228
IL13-20171	Illinois	01-11934/02-18228//06-14262
IL13-21219	Illinois	02-19463/00-8633//06-13024
IL13-227	Illinois	NCO6-27/04-10741
IL13-23803	Illinois	06-14303/01-6262/SHIRLEY
IL13-26034	Illinois	VA01W-476/03-1009//01-34159/04-8445
IL13-26521	Illinois	01-6262/02-7735
IL13-26648	Illinois	BESS//01-6262/05-38426
IL13-27973	Illinois	97-3632/97-7010//01-11934
IL13-28449	Illinois	03-18438/VA01W-476//01-6262/03-18438
IL13-29257	Illinois	00-8530//00-8530/VA01W-4766
IL13-4504	Illinois	ARS04-1249/97-1828
IL13-4606	Illinois	ARS04-1249/02-19463
IL13-5421	Illinois	00-8061/06-18051
IL13-5462	Illinois	00-8061/06-25634
IL13-5564	Illinois	00-8530/04-24668
IL13-6421	Illinois	01-16170/M0080104
IL13-7346	Illinois	02-18228/01-6262
IL13-7361	Illinois	02-18228/04-24668
IL13-7526	Illinois	02-18228/06-7034
IL13-8124	Illinois	04-9942/02-18228
IL13-9332	Illinois	06-7034/01-34159
IL13-9656	Illinois	06-7550/02-18228
KY03C-1002-02	Kentucky	25W33/25W60//25W33/KY90C-042-37-1
KY03C-1195-10-1-5	Kentucky	KY92C-0010-17//25R18/KY92C-0010-17
KY03C-1195-10-8-5	Kentucky	KY92C-0010-17//25R18/KY92C-0010-17
KY03C-1237-32	Kentucky	25R18/92C-0010-17//KY96C-0767-1
KY04C-1008-68-16-3	Kentucky	KY93C-0876-66/KY94C-0325-40-2//NC98-26192/KY93C-1238-17
KY04C-1128-38-1-5	Kentucky	SX1411/NC98-26192//25R78
KY04C-2004-1-1-1	Kentucky	ROANE/ALLEGIANCE
KY04C-2006-41-1-1	Kentucky	ROANE/KY93C-1238-17-1
KY04C-2031-29-6-1	Kentucky	TRUMAN/VA97W-375WS
KY04C-3051-4-2-1	Kentucky	25R23/SS560

GENOTYPE	ORIGIN	PEDIGREE
KY05C-1020-4-6-5	Kentucky	981517A1-1-5-2/25R18
KY05C-1126-50-12-1	Kentucky	MCCORMICK/ALLEGIANCE//KY97C-0554-02
KY05C-1169-45-10-5	Kentucky	25R78/KY93C-1238-17-1//VA01W-476
KY05C-1169-45-9-3	Kentucky	25R78/KY93C-1238-17-1//VA01W-476
KY05C-1169-61-14-1	Kentucky	25R78/KY93C-1238-17-1//VA01W-476
KY05C-1282-17-1-1	Kentucky	26R15/25R18//KY96C-0769-7-1
KY05C-1287-25-1-3	Kentucky	26R15/25R18//KY98C-1517-01
KY05C-1287-28-10-1	Kentucky	26R15/25R18//KY98C-1517-01
KY05C-1287-28-13-5	Kentucky	26R15/25R18//KY98C-1517-01
KY06C-1003-139-17-5	Kentucky	TRUMAN/MCCORMICK//25R37
KY06C-1003-140-3-3	Kentucky	TRUMAN/MCCORMICK//25R37
KY06C-1003-140-4-3	Kentucky	TRUMAN/MCCORMICK//25R37
KY06C-1177-15-18-1	Kentucky	KY93C-0004-22-1/NC03-11458//KY97C-0546-17-01
KY06C-2020-11-6-1	Kentucky	IL99-15867/B990081
KY06C-2067-15-1-1	Kentucky	KY97C-0519-04-05/AGRIPROCOOPER
KY06C-2093-34-19-1	Kentucky	KY98C-1169-06/B990081
KY09-1572-59-6-5	Kentucky	AGRIPROCOKER9511/IL04-8445
KY09C-0118-5-16-5	Kentucky	SSMPV-57//IL02-19463/KY97C-0546-20-05
KY09C-1073-54-5-1	Kentucky	VA03W-412/AGRIPROCOKER9511
KY09C-1119-46-14-1	Kentucky	KY00C-2109-01/KY01C-1537-05
KY09C-1245-100-9-3	Kentucky	LA01-425/VA06W-558
KY09C-1245-98-3-5	Kentucky	LA01-425/VA06W-558
KY09C-1501-93-17-3	Kentucky	KY00C-2039-15/VA05W-534
KY09C-1528-87-17-5	Kentucky	IL02-19463/IL01-34159
KY09C-1572-60-11-5	Kentucky	AGRIPROCOKER9511/IL04-8445
KY09C-1581-102-10-1	Kentucky	KY00C-2039-15/IL04-8445
KY16-1-28-1	Kentucky	26R15/25R18//KY98C-1517-01
KY221123	Kentucky	KY97C-0321-05-2*3/VA01W-476
KY409616	Kentucky	KY97C-0540-01-03*4/VA01W476
KY716506	Kentucky	KY98C-1474-02*4/VA01W476
MADISON	Virginia	ABE//BLUEBOY/VA71-54-147/3/VA72-54-14
MCCORMICK	Virginia	VA98W-591=92-51-39/AL870365
ROANE	Virginia	VA71-54-147/C68-15//IN65309C1-18-2-3-2
SALUDA	Virginia	VA71-54-147/COKER68-15
SHIRLEY	Virginia	VA94-52-25/COKER9835//VA96-54-234
SISSON	Virginia	COKER9803/FREEDOM
TRIBUTE	Virginia	VA98W-593=92-51-39/AL870365
TRIBUTE	Virginia	VA98W-593=92-51-39/AL870365
TYLER	Virginia	BLUEBOY //5* THORNE / 199-4 /3/ BLUEBOY SEL. 68-24-42 ; 199-4 = ASOSAN /3/ SUPREZA / REDHARD // CHANCELLOR /4/ P55-47.1 -5
USG 3555	Virginia	VA02W-555=VA94-52-60/PION2643//USG3209
VA07MAS12-8752-4-1-4	Virginia	U3960-3R-3-11-6/VA02W-398//GA96693-4E16
VA07MAS13-8833-1-4-3-1	Virginia	NC03-11458/TRIBUTE//SS5205
VA07MAS14-9260-8-2-2	Virginia	NC03-11458/IL99-15867//VA05W-436
VA07MAS1-7031-7-1-2-1	Virginia	MCCORMICK/AGS2060//SS5205
VA07MAS1-7031-7-1-2-4	Virginia	MCCORMICK/AGS2060//SS5205
VA07MAS1-7047-1-1-4-2	Virginia	MCCORMICK/AGS2060//SS5205
VA07MAS1-7054-3-3-2-1	Virginia	MCCORMICK/AGS2060//SS5205
VA07MAS1-7054-3-3-2-4	Virginia	MCCORMICK/AGS2060//SS5205

GENOTYPE	ORIGIN	PEDIGREE
VA07MAS3-7304-3-1-2-3	Virginia	SHIRLEY/AGS2060//SS8404
VA07MAS3-7304-3-2-2-3	Virginia	SHIRLEY/AGS2060//SS8404
VA07MAS3-7304-3-2-4	Virginia	SHIRLEY/GA951231-4E26//SS8404
VA07MAS3-7304-3-2-4-2	Virginia	SHIRLEY/AGS2060//SS8404
VA07MAS3-7304-3-2-4-3	Virginia	SHIRLEY/AGS2060//SS8404
VA07MAS3-7304-8-1-4	Virginia	SHIRLEY/GA951231-4E26//SS8404
VA07MAS4-7416-5-4-2	Virginia	GA951231-4E25/SS8404//SHIRLEY
VA07MAS4-7417-1-3-3	Virginia	GA951231-4E25/SS8404//SHIRLEY
VA07MAS4-7454-3-3-4	Virginia	GA951231-4E25/SS8404//SHIRLEY
VA07MAS4-7463-6-2-2-2	Virginia	OGLETHORPE/SS8404//SHIRLEY
VA07MAS4-7463-6-2-2-4	Virginia	OGLETHORPE/SS8404//SHIRLEY
VA07MAS4-7520-2-3-1	Virginia	GA951231-4E25/SS8404//SHIRLEY
VA07MAS9-8189-7-2-3-2	Virginia	AGS2026/PEMBROKE//SHIRLEY
VA08MAS1-188-6-4	Virginia	VA05W-640/VA05W-693//SHIRLEY
VA08MAS1-188-6-4-1	Virginia	VA05W-640/VA05W-693//SHIRLEY
VA08MAS1-188-6-4-3	Virginia	VA05W-640/VA05W-693//SHIRLEY
VA08MAS1-190-4-1	Virginia	VA05W-640/VA05W-693//SHIRLEY
VA08MAS1-85-8-2	Virginia	VA05W-640/VA05W-693//SHIRLEY
VA08MAS-369	Virginia	MCCORMICK/GA881130LE5
VA08MAS5-157-6-1-2	Virginia	BALDWIN/SHIRLEY//VA04W-360
VA08MAS5-18-3-1	Virginia	GA981621-1-3-5/SHIRLEY//VA04W-360
VA08MAS5-39-6-4	Virginia	GA981621-1-3-5/SHIRLEY//VA04W-360
VA08MAS6-174-7-4	Virginia	VA05W-693/SHIRLEY//GA001532-6E26
VA09MAS1-12-5-1	Virginia	GA991371-6E13/USG3555//OAKES
VA09MAS1-12-8-4	Virginia	GA991371-6E13/USG3555//OAKES
VA09MAS2-131-6-2	Virginia	GA991227-6A33/SHIRLEY//G41730
VA09MAS3-10-2-3	Virginia	VA05W-139/SS5205//GA031238-DH7-7A28
VA09MAS3-34-2-1	Virginia	VA05W-139/SS5205//GA031238-DH7-7A28
VA09MAS6-122-7-1	Virginia	SHIRLEY/GA991371-6E13//SS5205
VA09MAS6-55-4-4	Virginia	SHIRLEY/GA991371-6E13//SS5205
VA09MAS7-166-8-2	Virginia	VA06W-256/SS8641//12V51
VA09MAS7-168-6-3	Virginia	VA06W-256/SS8641//12V51
VA09MAS7-190-7-3	Virginia	VA06W-256/SS8641//12V51
VA09MAS7-30-2-1	Virginia	VA06W-256/SS8641//12V51
VA09MAS7-52-3-1	Virginia	VA06W-256/SS8641//12V51
VA09MAS7-61-2-1	Virginia	VA06W-256/SS8641//12V51
VA09MAS7-80-4-1	Virginia	VA06W-256/SS8641//12V51
VA09MAS8-34-2-4	Virginia	SHIRLEY/USG3120//P992231A1-2-1
VA09MAS8-34-5-2	Virginia	SHIRLEY/USG3120//P992231A1-2-1
VA09W-192WS	Virginia	TW137-003/VA97W-375WS
VA10W-119	Virginia	KY97C-0540-04/G/F951079-2E31
VA10W-123	Virginia	PIONEER25R47/G/F951079-2E31
VA10W-140	Virginia	VA01W-210/SS520//TRIBUTE
VA10W-21	Virginia	Z00-5018/VA01W-158
VA10W-21_BSR124	Virginia	Z00-5018/VA01W-158
VA10W-42	Virginia	JAMESTOWN/M99*3098
VA10W-96	Virginia	F/G95195/JAMESTOWN
VA11MAS-7383-6-3-155	Virginia	SHIRLEY/AGS2060//SS8404

GENOTYPE	ORIGIN	PEDIGREE
VA11MAS-7520-2-3-255	Virginia	OGLETHORPE/SS8404//SHIRLEY
VA11MAS-9409-8-2-798	Virginia	VA04W-433/SS8404//VA02W-398
VA11W-106	Virginia	PIONEER25R47/JAMESTOWN
VA11W-108	Virginia	PIONEER25R47/JAMESTOWN
VA11W-111	Virginia	PIONEER25R47/JAMESTOWN
VA11W-182	Virginia	BRANSON/SHIRLEY
VA11W-194	Virginia	SSMPV57/M99*3098//VA03W-434
VA11W-230	Virginia	SS520/G/F951208-2E35//JAMESTOWN
VA11W-278	Virginia	NC00-15389/GF951079-2E31//USG3555
VA11W-279	Virginia	NC00-15389/GF951079-2E31//USG3555
VA11W-301	Virginia	PIONEER25R47/NC00-15389//JAMESTOWN
VA11W-31	Virginia	F/G95195/JAMESTOWN
VA11W-313	Virginia	PIONEER25R47/G/F951079-2E31//USG3555
VA11W-95	Virginia	PIONEER25R47/JAMESTOWN
VA11W-FHB4	Virginia	P97397B1-4-5/MCCORMICK//COKER9511
VA11W-FHB60	Virginia	VA97W-375RS/FG95195//VA04W-547
VA12FHB-34	Virginia	GA991109-4-1-3/PIONEER26R15
VA12FHB-37	Virginia	VA04W-433/SS8404
VA12FHB-4	Virginia	IL99-15867/VA04W-433
VA12FHB-53	Virginia	VA04W-433/BRANSON
VA12FHB-55	Virginia	VA04W-433/BRANSON
VA12FHB-77	Virginia	IL99-15867/VA04W-433//SS8404
VA12FHB-8	Virginia	IL99-27048/VA04W-486//SHIRLEY
VA12FHB-85	Virginia	IL96-24851-1/VA03W-434[ROANE/CK9835//96-54-270)//SS8404
VA12W-100	Virginia	VA03W-436/IL99-15867
VA12W-101	Virginia	VA03W-436/IL99-15867
VA12W-102	Virginia	VA03W-436/IL99-15867
VA12W-104	Virginia	VA03W-436/IL99-15867
VA12W-150	Virginia	IL99-15867/JAMESTOWN
VA12W-209	Virginia	VA03W-249/SS8641//USG3315
VA12W-22	Virginia	KY93C-1238-17-1/VA03W-436
VA12W-232	Virginia	VA01W-353/SS8641//RENWOOD3706
VA12W-241	Virginia	KY93C-1238-17-1/VA03W-436//SS8404
VA12W-248	Virginia	KY97C-0574-01/USG3555//USG3295
VA12W-26	Virginia	SSMPV57/M99*3098//RENWOOD3434
VA12W-272	Virginia	VA03W-235/SS8641//VA04W-86
VA12W-283	Virginia	CHESAPEAKE/SS8641//VA04W-439
VA12W-31	Virginia	SSMPV57/M99*3098//VA03W-434
VA12W-45	Virginia	SS520/G/F951208-2E35//JAMESTOWN
VA12W-49	Virginia	NC00-15389/GF951079-2E31//USG3555
VA12W-54	Virginia	NC00-15389/GF951079-2E31//USG3555
VA12W-68	Virginia	PIONEER25R47/G/F951079-2E31//USG3555
VA12W-69	Virginia	PIONEER25R47/G/F951079-2E31//USG3555
VA12W-72	Virginia	PIONEER25R47/G/F951079-2E31//USG3555
VA12W-97	Virginia	MERL/AGS2026
VA13FHB-1	Virginia	P97397J1-4-1-4-7/VA04W-433//VA02W-398
VA13FHB-11	Virginia	IL99-15867/VA04W-433//SS8404
VA13FHB-13	Virginia	IL99-15867/VA04W-433//SS8404

GENOTYPE	ORIGIN	PEDIGREE
VA13FHB-17	Virginia	IL99-15867/VA04W-433//SS8404
VA13FHB-18	Virginia	IL99-15867/VA04W-433//SS8404
VA13FHB-26	Virginia	VA05W-436/VA05W-641
VA13FHB-32	Virginia	VA05W-641/AGS2020
VA13FHB-5	Virginia	IL99-15867/VA04W-433//SS8404
VA13FHB-8	Virginia	IL99-15867/VA04W-433//SS8404
VA13FHB-9	Virginia	IL99-15867/VA04W-433//SS8404
VA13W-111	Virginia	SHIRLEY/GA98249G1-G1-2
VA13W-124	Virginia	12V51/AGS2026
VA13W-138	Virginia	VA05W-363/VA97-51-26//CK9835/SS520]/VA03W-310[VA95-51-21/CK9904//RENWOOD3260
VA13W-144	Virginia	JAMESTOWN/SS8404//VA04W-259
VA13W-148	Virginia	JAMESTOWN/SS8404//AGS2020
VA13W-150	Virginia	JAMESTOWN/SS8404//AGS2020
VA13W-154	Virginia	JAMESTOWN/SS8404//AGS2020
VA13W-162	Virginia	JAMESTOWN/USG3295//VA02W-398
VA13W-165	Virginia	JAMESTOWN/USG3295//VA02W-398
VA13W-169	Virginia	JAMESTOWN/AGS2026//SS8404
VA13W-174	Virginia	SHIRLEY/BRANSON//JAMESTOWN
VA13W-177	Virginia	SHIRLEY/BRANSON//JAMESTOWN
VA13W-178	Virginia	SHIRLEY/BRANSON//JAMESTOWN
VA13W-179	Virginia	SHIRLEY/BRANSON//JAMESTOWN
VA13W-180	Virginia	SHIRLEY/BRANSON//JAMESTOWN
VA13W-191	Virginia	VA04W-259[VA97W-533/NC95-11612/SS8404//AGS2026
VA13W-20	Virginia	SS520/GF951208-2E35//JAMESTOWN
VA13W-212	Virginia	W-1377/VA03W-310//JAMESTOWN
VA13W-217	Virginia	M01*1019/VA03W-203//AGS2020
VA13W-220	Virginia	KEY/VA02W-398//VA05W-436
VA13W-37	Virginia	IL99-15867/JAMESTOWN
VA13W-38	Virginia	IL99-15867/JAMESTOWN
VA13W-41	Virginia	IL99-15867/JAMESTOWN
VA13W-42	Virginia	IL99-15867/JAMESTOWN
VA13W-47	Virginia	IL99-15867/JAMESTOWN
VA13W-52	Virginia	USG3555/SHIRLEY//JAMESTOWN
VA13W-56	Virginia	USG3555/SHIRLEY//JAMESTOWN
VA13W-57	Virginia	USG3555/SHIRLEY//JAMESTOWN
VA13W-8	Virginia	FG95195/JAMESTOWN
VA13W-9	Virginia	FG95195/JAMESTOWN
VA13W-97	Virginia	SHIRLEY/GA98249G1-G1-2
VA13W-99	Virginia	SHIRLEY/GA98249G1-G1-2
VA14FHB-12	Virginia	VA05W-436/VA05W-641
VA14FHB-13	Virginia	VA05W-436/VA05W-641
VA14FHB-14	Virginia	VA05W-436/VA05W-641
VA14FHB-15	Virginia	VA05W-510/GA991336-6E9
VA14FHB-17	Virginia	IL03-18438/VA04W-360
VA14FHB-18	Virginia	IL03-18438/VA04W-360
VA14FHB-21	Virginia	UNKNOWN
VA14FHB-22	Virginia	UNKNOWN
VA14FHB-23	Virginia	UNKNOWN

GENOTYPE	ORIGIN	PEDIGREE
VA14FHB-26	Virginia	UNKNOWN
VA14FHB-28	Virginia	UNKNOWN
VA14FHB-29	Virginia	UNKNOWN
VA14FHB-31	Virginia	UNKNOWN
VA14FHB-7	Virginia	VA05W-436/VA05W-641
VA14FHB-9	Virginia	VA05W-436/VA05W-641
VA14W-11	Virginia	SHIRLEY/AGS2020//JAMESTOWN
VA14W-12	Virginia	SHIRLEY/AGS2020//JAMESTOWN
VA14W-17	Virginia	VA04W-259/NC95-11612)]/B020815
VA14W-26	Virginia	VA05W-139/MCCORMICK)]/SS5205
VA14W-28	Virginia	VA05W-139/MCCORMICK)]/SS5205
VA14W-29	Virginia	VA05W-139/MCCORMICK)]/SS5205
VA14W-3	Virginia	JAMESTOWN/SHIRLEY//SS8404
VA14W-32	Virginia	VA05W-139/MCCORMICK)]/SS5205
VA14W-34	Virginia	VA05W-139/MCCORMICK)]/SS5205
VA14W-35	Virginia	VA05W-139/MCCORMICK)]/SS5205
VA14W-4	Virginia	JAMESTOWN/SHIRLEY//SS8404
VA14W-40	Virginia	VA06W-112/VA06W-256
VA14W-41	Virginia	VA06W-112/VA06W-256
VA14W-46	Virginia	VA06W-256/JAMESTOWN
VA14W-50	Virginia	VA06W-256/JAMESTOWN
VA14W-55	Virginia	VA06W-256/JAMESTOWN
VA14W-57	Virginia	VA06W-627/VA06W-256
VA14W-59	Virginia	VA06W-627/VA06W-256
VA14W-6	Virginia	JAMESTOWN/SHIRLEY//SS8404
VA14W-60	Virginia	VA06W-627/VA06W-256
VA14W-62	Virginia	MO-011126/VA03W-203//VA06W-6/
VA14W-69_YR	Virginia	[YR 15/6*AVOCET(S) / RENWOOD3434 [VA03W-434=ROANE(93-54-429)/ CK9835// 96-54-270 (88-54-612 (MSY*2/ BALKAN)/FFR511W)] //(YR 5/6*AVOCET(S)/COKER9553[D00*6874-2=89M-4035A (IL77-2656/ NK79W810/ PIO2580)] /3/ [NC04-20814(NC94-6275/ P86958// SISSON"S" (VA96-54-234)], F6
VA14W-7	Virginia	JAMESTOWN/SHIRLEY//SS8404
VA14W-70_YR	Virginia	[YR 15/6*AVOCET(S) / RENWOOD3434 [VA03W-434=ROANE(93-54-429)/ CK9835// 96-54-270 (88-54-612 (MSY*2/ BALKAN)/FFR511W)] //(YR 5/6*AVOCET(S)/COKER9553[D00*6874-2=89M-4035A (IL77-2656/ NK79W810/ PIO2580)] /3/ [NC04-20814(NC94-6275/ P86958// SISSON"S" (VA96-54-234)], F6
VA14W-73_YR	Virginia	[YR 15/6*AVOCET(S) / RENWOOD3434 [VA03W-434=ROANE(93-54-429)/ CK9835// 96-54-270 (88-54-612 (MSY*2/ BALKAN)/FFR511W)] //(YR 5/6*AVOCET(S)/COKER9553[D00*6874-2=89M-4035A (IL77-2656/ NK79W810/ PIO2580)] /3/ [NC04-20814(NC94-6275/ P86958// SISSON"S" (VA96-54-234)], F6
VA14W-8	Virginia	JAMESTOWN/SHIRLEY//SS8404
WAKEFIELD	Virginia	ARTHUR // CI 13836 /8* CHANCELLOR , VA 68-22 -7// CI 13836 /8* CHANCELLOR , DOUBLECROP // ABE / VA 68-24 - 42 /3/ CI 13836 /8* CHANCELLOR , AND OASIS / VA 68-24 - 42 // CI 13836 /8* CHANCELLOR
X08C-1039-58-9-1	Kentucky	VA03W-409/IL01-34159
X08C-1074-46-14-5	Kentucky	AGRIPROCOKERBRANSON/P.03630A1-18

GENOTYPE	ORIGIN	PEDIGREE
X08C-1077-10-13-5	Kentucky	SSMPV-57/P.03630A1
X08C-1089-77-13-1	Kentucky	KY97C-0321-02-01/P.981129A1--17
X08C-1089-78-14-3	Kentucky	KY97C-0321-02-01/P.981129A1--17
X08C-1181-62-14-5	Kentucky	OH02-12686/IL02-19463
X08C-1501-21-8-3	Kentucky	IL02-19463/KY97C-0554-03-02
X08C-1502-26-4-1	Kentucky	MD01W233-06-1/KY97C-0554-03-02
X08C-1502-26-8-3	Kentucky	MD01W233-06-1/KY97C-0554-03-02
X08C-1502-28-7-3	Kentucky	MD01W233-06-1/KY97C-0554-03-02
X10-0003-16-8-5	Kentucky	SSMPV-57//KY97C-0519-04-07/KY02C-3005-25
X10-0049-3-11-3	Kentucky	SSMPV-57//KY97C-0519-04-07/KY01C-1537-05
X10-0049-3-1-5	Kentucky	SSMPV-57//KY97C-0519-04-07/KY01C-1537-05
X10-0049-4-7-1	Kentucky	SSMPV-57//KY97C-0519-04-07/KY01C-1537-05
X10-0060-5-18-5	Kentucky	SSMPV-57//KY97C-0519-04-07/IL02-19463
X10-0218-25-7-5	Kentucky	KY02C-3004-07//PEMBROKE/KY03C-2170-24
X10-0225-27-14-1	Kentucky	SSMPV-57//PEMBROKE/COKER9511
X10-0267-33-5-3	Kentucky	KY97C-0519-04-07//PEMBROKE/BESS
X10-0269-34-13-5	Kentucky	SSMPV-57//PEMBROKE/BESS
X10-0503-45-17-5	Kentucky	KY02C-3007-41//KY97C-0321-02-01/MD99W483-06-11

APPENDIX C: Predicted translational effects and distance to closest genes for significant marker-trait associations identified in chapter I

NOTE: The SNP S4A_739598141 resides within two predicted genes; not all scaffolds containing a significant marker-trait association contain genes

The code below may be copied and pasted into a text editor and saved with a .csv extension to create a data table that can readily be opened as a spreadsheet:

```
MODEL,TRAIT,SNP,TGAC_POS,ALLELES,SNP_CLASS,CLOSEST_GENE,DISTANCE_TO_GENE,RESIDUES,CODONS

farmcpu,TKW,S1A_22935081,TGACv1_scaffold_019544_1AS:35763,C/T,intergenic,TRIAE_CS42_1AS_TGACv1_019544_AA0067890,6153,-,-

farmcpu,SSQM,S1A_50513589,TGACv1_scaffold_019020_1AS:184277,G/A,3' proximal intergenic,TRIAE_CS42_1AS_TGACv1_019020_AA0058200,1388,-,-

gcta; farmcpu,TWT,S1A_583587147,TGACv1_scaffold_000910_1AL:23788,C/T,5' proximal intergenic,TRIAE_CS42_1AL_TGACv1_000910_AA0021640,3543,-,-

farmcpu,STARCH,S1B_39294256,TGACv1_scaffold_049505_1BS:21361,G/A,intronic,TRIAE_CS42_1BS_TGACv1_049505_AA0155360,0,-,-

farmcpu,MAT,S1B_44010984,TGACv1_scaffold_049567_1BS:4059,A/G,intergenic,TRIAE_CS42_1BS_TGACv1_049567_AA0157010,26087,-,-

farmcpu,HD,S1B_50850397,TGACv1_scaffold_050247_1BS:70050,A/G,3' proximal intergenic,TRIAE_CS42_1BS_TGACv1_050247_AA0169640,1230,-,-

farmcpu,STARCH,S1B_659857468,TGACv1_scaffold_031997_1BL:15691,C/T,5' proximal intergenic,TRIAE_CS42_1BL_TGACv1_031997_AA0124020,756,-,-

farmcpu,STARCH,S1D_6674498,TGACv1_scaffold_080394_1DS:80273,G/C,3' UTR,TRIAE_CS42_1DS_TGACv1_080394_AA0247280,0,-,-

farmcpu,HT,S2A_764941637,TGACv1_scaffold_093694_2AL:9813,G/A,intergenic,TRIAE_CS42_2AL_TGACv1_093694_AA0285290,26304,-,-

farmcpu,GSQM,S2B_35041282,TGACv1_scaffold_145905_2BS:18213,T/G,intergenic,TRIAE_CS42_2BS_TGACv1_145905_AA0448870,27333,-,-

farmcpu,HP,S2B_66522933,TGACv1_scaffold_146004_2BS:81183,A/C,intergenic,TRIAE_CS42_2BS_TGACv1_146004_AA0452740,18071,-,-

farmcpu,HD; MAT,S2D_35084672,TGACv1_scaffold_177832_2DS:20809,G/A,5' proximal intergenic,TRIAE_CS42_2DS_TGACv1_177832_AA0585380,2957,-,-

farmcpu,HD,S2D_9872868,TGACv1_scaffold_177916_2DS:19528,A/T,5' proximal intergenic,TRIAE_CS42_2DS_TGACv1_177916_AA0587150,255,-,-

farmcpu,GSQM,S3A_20232500,TGACv1_scaffold_211731_3AS:33827,C/A,5' proximal intergenic,TRIAE_CS42_3AS_TGACv1_211731_AA0693510,2932,-,-

farmcpu,SSQM,S3A_2493807,TGACv1_scaffold_213057_3AS:28456,C/T,intergenic,TRIAE_CS42_3AS_TGACv1_213057_AA0705070,9411,-,-
```

gcta,SSQM,S3A_3273716,TGACv1_scaffold_213380_3AS:13893,T/C,intergenic,-,-,-,-

farmcpu,FLS,S3A_569991635,TGACv1_scaffold_202710_3AL:2242,T/C,intergenic,-,-,-,-

farmcpu,GSQM,S3A_691815364,TGACv1_scaffold_198180_3AL:8087,G/A,intergenic,-,-,-,-

farmcpu,TWT,S3B_695966897,TGACv1_scaffold_223320_3B:37476,A/G,missense,TRIAE_CS42_3B_TGACv1_223320_AA0780190,0,T/A,Acg/Gcg

farmcpu,YLD,S3D_511264206,TGACv1_scaffold_251777_3DL:4545,C/T,intergenic,-,-,-,-

farmcpu,YLD,S4A_726716318,TGACv1_scaffold_288738_4AL:72223,G/A,intronic,TRIAE_CS42_4AL_TGACv1_288738_AA0956970,0,-,-

farmcpu,GSQM,S4A_739598141,TGACv1_scaffold_288585_4AL:104660,C/T,synonymous,TRIAE_CS42_4AL_TGACv1_288585_AA0953000,0,K,aaG/aaA

farmcpu,GSQM,S4A_739598141,TGACv1_scaffold_288585_4AL:104660,C/T,intronic,TRIAE_CS42_4AL_TGACv1_288585_AA0953010,0,-,-

farmcpu,HI;
STARCH,S4B_626390000,TGACv1_scaffold_320403_4BL:205740,A/C,intergenic,TRIAE_CS42_4BL_TGACv1_320403_AA1037940,19266,-,-

farmcpu,STARCH,S5A_9462259,TGACv1_scaffold_393202_5AS:84259,A/C,intronic,TRIAE_CS42_5AS_TGACv1_393202_AA1269790,0,-,-

farmcpu,HT,S5B_261134879,TGACv1_scaffold_407709_5BL:36111,G/A,3' proximal intergenic,TRIAE_CS42_5BL_TGACv1_407709_AA1359020,1901,-,-

farmcpu,STARCH,S5B_34721398,TGACv1_scaffold_424234_5BS:41438,A/G,premature stop,TRIAE_CS42_5BS_TGACv1_424234_AA1387820,0,stop/Q,Taa/Caa

farmcpu,HI,S5B_394707451,TGACv1_scaffold_404873_5BL:52196,G/A,intergenic,TRIAE_CS42_5BL_TGACv1_404873_AA1313490,35687,-,-

gcta;
farmcpu,GSQM,S5B_396479359,TGACv1_scaffold_641320_U:51261,C/T,intergenic,TRIAE_CS42_U_TGACv1_641320_AA2091830,9205,-,-

farmcpu,HI,S5B_644947034,TGACv1_scaffold_405672_5BL:24458,T/A,5' proximal intergenic,TRIAE_CS42_5BL_TGACv1_405672_AA1332610,3083,-,-

farmcpu,STARCH,S5D_365732020,TGACv1_scaffold_433921_5DL:59862,A/C,3' UTR,TRIAE_CS42_5DL_TGACv1_433921_AA1425280,0,-,-

farmcpu,HT,S5D_451607895,TGACv1_scaffold_445874_5DL:694,G/C,intergenic,-,-,-,-

farmcpu,SSQM,S5D_499069158,TGACv1_scaffold_434033_5DL:22481,G/A,intergenic,-,-,-,-

farmcpu,SSQM,S6A_614373502,TGACv1_scaffold_470886_6AL:28755,A/G,intergenic,TRIAE_CS42_6AL_TGACv1_470886_AA1497490,6007,-,-

farmcpu,TWT,S6A_614660970,TGACv1_scaffold_471361_6AL:101992,T/C,intronic,TRIAE_CS42_6AL_TGACv1_471361_AA1507680,0,-,-

farmcpu,FLS,S6A_63296169,TGACv1_scaffold_488607_6AS:3941,A/G,intergenic,-,-,-,-

gcta,FLS,S6B_181128808,TGACv1_scaffold_513186_6BS:5592,A/G,intergenic,TRIAE_CS42_6BS_TGACv1_513186_AA1633420,23564,-,-

farmcpu,MAT,S6B_32730233,TGACv1_scaffold_513471_6BS:85717,G/C,intergenic,-,-,-,-

farmcpu,YLD,S6B_656279771,TGACv1_scaffold_499780_6BL:147271,C/T,3' UTR,TRIAE_CS42_6BL_TGACv1_499780_AA1591550,0,-,-

farmcpu,HT,S6B_706326554,TGACv1_scaffold_500675_6BL:59205,A/G,intronic,TRIAE_CS42_6BL_TGACv1_500675_AA1608280,0,-,-

farmcpu,HD,S6D_127384672,TGACv1_scaffold_543118_6DS:58895,G/A,missense,TRIAE_CS42_6DS_TGACv1_543118_AA1735700,0,A/V,gCt/gTt

farmcpu,TKW,S6D_468113959,TGACv1_scaffold_526541_6DL:60185,A/T,intergenic,TRIAE_CS42_6DL_TGACv1_526541_AA1686520,13781,-,-

farmcpu,GSQM,S7A_673387152,TGACv1_scaffold_558731_7AL:5864,C/T,intergenic,-,-,-,-

farmcpu,TKW,S7A_673436887,TGACv1_scaffold_556548_7AL:99192,T/C,intergenic,-,-,-,-

farmcpu,HP,S7B_41890395,TGACv1_scaffold_592761_7BS:63052,A/C,intergenic,-,-,-,-

farmcpu,FLSG,S7B_63999446,TGACv1_scaffold_593873_7BS:8658,A/T,intergenic,-,-,-,-

gcta,FLSG,S7B_64393207,TGACv1_scaffold_592109_7BS:67699,G/A,intergenic,TRIAE_CS42_7BS_TGACv1_592109_AA1930670,21488,-,-

farmcpu,HD,S7D_553110861,TGACv1_scaffold_604806_7DL:19877,C/T,intergenic,-,-,-,-

gcta;
farmcpu,FLS,S7D_58449294,TGACv1_scaffold_623041_7DS:21730,A/G,missense,TRIAE_CS42_7DS_TGACv1_623041_AA2048950,0,M/V,Atg/Gtg

gcta; farmcpu,HD;
MAT,S7D_58589271,TGACv1_scaffold_622088_7DS:62338,A/G,intronic,TRIAE_CS42_7DS_TGACv1_622088_AA2032520,0,-,-

APPENDIX D: Protein-protein BLAST results for all transcripts located on genomic (TGACv1 assembly) scaffolds containing significant marker-trait associations identified in chapter I

The code below may be copied and pasted into a text editor and saved with a .csv extension to create a data table that can readily be opened as a spreadsheet:

```
QUERY_ENSEMBL_ID,SUBJECT_ENSEMBL_ID,SUBJECT_DESCRIPTION,QUERY_START,QUERY_END
,SUBJECT_START,SUBJECT_END,E_VALUE,PERCENT_IDENT,QUERY_COVERAGE,BITSCORE,CLOS
EST_SNP,SNP_TRAIT,SNP_DIST_FROM_QUERY,SNP_CLASS

TRIAE_CS42_1AL_TGACv1_000910_AA0021640.1,TRIUR3_21554-P1,L-type lectin-domain
containing receptor kinase IX.1
[Source:UniProtKB/TrEMBL;Acc:M7YC97],1,424,1,380,0,84.24,88,624,S1A_583587147
,TWT,3543,5' proximal intergenic

TRIAE_CS42_1AL_TGACv1_000910_AA0021650.1,TRIUR3_09155-P1,Ankyrin-3
[Source:UniProtKB/TrEMBL;Acc:M8A2S2],6,468,448,887,1.00E-
102,44.08,99,332,S1A_583587147,TWT,-968,5' proximal intergenic

TRIAE_CS42_1AL_TGACv1_000910_AA0021670.1,MLOC_75180.4,Endoplasmin homolog
[Source:UniProtKB/Swiss-Prot;Acc:P36183],81,493,1,449,1.00E-
144,53.74,81,430,S1A_583587147,TWT,-64431,5' proximal intergenic

TRIAE_CS42_1AL_TGACv1_000910_AA0021670.2,MLOC_75180.4,Endoplasmin homolog
[Source:UniProtKB/Swiss-Prot;Acc:P36183],58,470,1,449,3.00E-
145,53.74,84,431,S1A_583587147,TWT,-64431,5' proximal intergenic

TRIAE_CS42_1AS_TGACv1_019020_AA0058140.1,EMT11556,SKP1-like protein 1B
[Source:UniProtKB/TrEMBL;Acc:M8BC34],57,140,218,300,4.00E-
07,35.71,29,53.9,S1A_50513589,SSQM,141072,3' proximal intergenic

TRIAE_CS42_1AS_TGACv1_019020_AA0058160.1,AES91877,ubiquitin-like-specific
protease ESD4-like protein,307,492,304,486,1.00E-
13,27.23,36,76.3,S1A_50513589,SSQM,127332,3' proximal intergenic

TRIAE_CS42_1AS_TGACv1_019020_AA0058200.1,GSMUA_Achr10P13520_001,Myb-related
protein MYBAS1
[Source:GMGC_GENE;Acc:GSMUA_Achr10G13520_001],15,211,14,230,5.00E-
74,55.96,91,229,S1A_50513589,SSQM,1388,3' proximal intergenic

TRIAE_CS42_1AS_TGACv1_019020_AA0058200.2,GSMUA_Achr10P13520_001,Myb-related
protein MYBAS1
[Source:GMGC_GENE;Acc:GSMUA_Achr10G13520_001],1,228,29,230,2.00E-
52,48.07,98,175,S1A_50513589,SSQM,1388,3' proximal intergenic

TRIAE_CS42_1AS_TGACv1_019020_AA0058200.3,GSMUA_Achr10P13520_001,Myb-related
protein MYBAS1
[Source:GMGC_GENE;Acc:GSMUA_Achr10G13520_001],15,258,14,230,6.00E-
74,52.46,93,231,S1A_50513589,SSQM,1388,3' proximal intergenic

TRIAE_CS42_1AS_TGACv1_019544_AA0067890.1,EMT24409,SKP1-like protein 4
[Source:UniProtKB/TrEMBL;Acc:M8BQZ8],225,276,265,314,0.001,42.31,15,43.9,S1A_
22935081,TKW,-6153,intergenic
```

TRIAE_CS42_1AS_TGACv1_019544_AA0067890.2,EMT24409,SKP1-like protein 4 [Source:UniProtKB/TrEMBL;Acc:M8BQZ8],225,282,265,322,7.00E-04,40,20,43.9,S1A_22935081,TKW,-6153,intergenic

TRIAE_CS42_1AS_TGACv1_019544_AA0067900.1,EMT00716,Protein FAR-RED ELONGATED HYPOCOTYL 3 [Source:UniProtKB/TrEMBL;Acc:N1QTR2],1,154,149,289,2.00E-58,65.58,67,196,S1A_22935081,TKW,-10316,intergenic

TRIAE_CS42_1BL_TGACv1_031997_AA0124020.1,EMT02554,LRR receptor-like serine/threonine-protein kinase FLS2 [Source:UniProtKB/TrEMBL;Acc:N1QRD0],1,542,1,600,1.00E-147,49.67,100,469,S1B_659857468,STARCH,756,5' proximal intergenic

TRIAE_CS42_1BL_TGACv1_031997_AA0124030.1,TRIUR3_09155-P1,Ankyrin-3 [Source:UniProtKB/TrEMBL;Acc:M8A2S2],23,463,476,887,2.00E-109,48.46,95,350,S1B_659857468,STARCH,-2636,5' proximal intergenic

TRIAE_CS42_1BL_TGACv1_031997_AA0124050.1,TRIUR3_09155-P1,Ankyrin-3 [Source:UniProtKB/TrEMBL;Acc:M8A2S2],30,497,480,890,7.00E-91,42.8,94,302,S1B_659857468,STARCH,-8385,5' proximal intergenic

TRIAE_CS42_1BL_TGACv1_031997_AA0124050.2,TRIUR3_09155-P1,Ankyrin-3 [Source:UniProtKB/TrEMBL;Acc:M8A2S2],41,471,492,890,6.00E-100,45.91,91,325,S1B_659857468,STARCH,-8385,5' proximal intergenic

TRIAE_CS42_1BS_TGACv1_049505_AA0155330.1,TRIUR3_09155-P1,Ankyrin-3 [Source:UniProtKB/TrEMBL;Acc:M8A2S2],38,457,477,861,1.00E-71,39.49,79,250,S1B_39294256,STARCH,10824,intronic

TRIAE_CS42_1BS_TGACv1_049505_AA0155330.2,TRIUR3_09155-P1,Ankyrin-3 [Source:UniProtKB/TrEMBL;Acc:M8A2S2],38,494,477,886,3.00E-72,37.85,92,251,S1B_39294256,STARCH,10824,intronic

TRIAE_CS42_1BS_TGACv1_049505_AA0155350.1,TRIUR3_22088-P1,5'-methylthioadenosine/S-adenosylhomocysteine nucleosidase [Source:UniProtKB/TrEMBL;Acc:M7YX26],13,54,298,339,6.00E-07,59.52,60,48.5,S1B_39294256,STARCH,5346,intronic

TRIAE_CS42_1BS_TGACv1_049505_AA0155360.1,EMT23143,Protein STIP1-like protein [Source:UniProtKB/TrEMBL;Acc:M8BMT1],31,338,9,344,0,89.88,91,602,S1B_39294256,STARCH,0,intronic

TRIAE_CS42_1BS_TGACv1_049505_AA0155360.2,EMT23143,Protein STIP1-like protein [Source:UniProtKB/TrEMBL;Acc:M8BMT1],31,338,9,344,0,89.88,91,602,S1B_39294256,STARCH,0,intronic

TRIAE_CS42_1BS_TGACv1_049505_AA0155370.1,BGIOSGA008089-PA,Protein FON2 SPARE1 [Source:UniProtKB/Swiss-Prot;Acc:A2X462],21,136,20,131,3.00E-15,52.5,85,71.6,S1B_39294256,STARCH,-67150,intronic

TRIAE_CS42_1BS_TGACv1_049567_AA0157010.1,TRIUR3_05946-P1,"Calcium-transporting ATPase 4, endoplasmic reticulum-type [Source:UniProtKB/TrEMBL;Acc:M8AS38]",184,1057,1,848,0,96.68,83,1682,S1B_44010984,MAT,-26087,intergenic

TRIAE_CS42_1BS_TGACv1_049567_AA0157020.1,EMJ09213,Cam-binding protein 60-like G [Source:Projected from Arabidopsis thaliana (AT5G26920)

TAIR;Acc:AT5G26920],55,279,83,309,7.00E-41,37.44,39,159,S1B_44010984,MAT,-83305,intergenic

TRIAE_CS42_1BS_TGACv1_050247_AA0169620.1,EMT23334,Tropinone reductase 1 [Source:UniProtKB/TrEMBL;Acc:M8BN31],526,574,260,317,0.015,50,8,41.2,S1B_50850397,HD,7789,3' proximal intergenic

TRIAE_CS42_1BS_TGACv1_050247_AA0169640.1,TRIUR3_05731-P1,Tetratricopeptide repeat (TPR)-like superfamily protein [Source:Projected from Arabidopsis thaliana (AT5G03560) TAIR;Acc:AT5G03560],1,269,1,269,9.00E-171,91.45,100,479,S1B_50850397,HD,-1230,3' proximal intergenic

TRIAE_CS42_1DS_TGACv1_080394_AA0247270.1,TRIUR3_35063-P1,Fatty acyl-CoA reductase 2 [Source:UniProtKB/TrEMBL;Acc:M7YFR6],1,568,1,580,0,90.34,100,1083,S1D_6674498,STARCH,74998,3' UTR

TRIAE_CS42_1DS_TGACv1_080394_AA0247280.1,OS06T0520600-01,Zinc finger CCCH domain-containing protein 43 [Source:Uniprot/SWISSPROT;Acc:Q5Z5Q3],514,693,534,711,9.00E-41,50.28,77,162,S1D_6674498,STARCH,0,3' UTR

TRIAE_CS42_2AL_TGACv1_093694_AA0285290.1,TRIUR3_25882-P1,Werner Syndrome-like exonuclease [Source:UniProtKB/TrEMBL;Acc:M8B238],1,231,1,232,4.00E-148,96.55,99,419,S2A_764941637,HT,-26304,intergenic

TRIAE_CS42_2AL_TGACv1_093694_AA0285310.1,EMT15100,Heat shock cognate 70 kDa protein 1 [Source:UniProtKB/TrEMBL;Acc:N1R361],12,625,780,1417,0,52.04,96,597,S2A_764941637,HT,-39816,intergenic

TRIAE_CS42_2AL_TGACv1_093694_AA0285330.1,EMT28374,Speckle-type POZ protein-like protein [Source:UniProtKB/TrEMBL;Acc:M8C2B6],1,301,1,300,0,90.7,85,571,S2A_764941637,HT,-77147,intergenic

TRIAE_CS42_2BS_TGACv1_145905_AA0448870.1,TRIUR3_31175-P1,UDP-glycosyltransferase 74F2 [Source:UniProtKB/TrEMBL;Acc:M7ZVE2],14,158,10,158,2.00E-36,52,86,137,S2B_35041282,GSQM,-27333,intergenic

TRIAE_CS42_2BS_TGACv1_145905_AA0448900.1,TRIUR3_31175-P1,UDP-glycosyltransferase 74F2 [Source:UniProtKB/TrEMBL;Acc:M7ZVE2],212,392,3,186,7.00E-35,43.48,38,140,S2B_35041282,GSQM,-177178,intergenic

TRIAE_CS42_2BS_TGACv1_145905_AA0448930.1,EMT07704,Lectin-domain containing receptor kinase A4.2 [Source:UniProtKB/TrEMBL;Acc:M8BLR0],271,596,331,656,0,93.56,98,635,S2B_35041282,GSQM,-275503,intergenic

TRIAE_CS42_2BS_TGACv1_146004_AA0452740.1,OS07T0691800-01,26S protease regulatory subunit 4 homolog [Source:Uniprot/SWISSPROT;Acc:P46466],1,449,1,448,0,98,100,882,S2B_66522933,HP,-18071,intergenic

TRIAE_CS42_2BS_TGACv1_146004_AA0452740.2,OS07T0691800-01,26S protease regulatory subunit 4 homolog [Source:Uniprot/SWISSPROT;Acc:P46466],1,423,1,422,0,97.87,94,830,S2B_66522933,HP,-18071,intergenic

TRIAE_CS42_2BS_TGACv1_146004_AA0452750.1,TRIUR3_09269-P1,Leucine-rich repeat receptor-like protein kinase PEPR2 [Source:UniProtKB/TrEMBL;Acc:M7Z8I6],1,576,212,787,0,94.79,100,1123,S2B_66522933,HP,-33616,intergenic

TRIAE_CS42_2DS_TGACv1_177832_AA0585380.1,POPTR_0008s07580.1,Dicer-like protein [source:UniProtKB/TrEMBL;Acc:B9HHX6_POPTR],28,76,656,712,1.2,29.82,60,30,S2D_35084672,HD; MAT,-2957,5' proximal intergenic

TRIAE_CS42_2DS_TGACv1_177832_AA0585400.1,MLOC_36445.2,plasmodesmata-located protein 8 [Source:Projected from Arabidopsis thaliana (AT3G60720) TAIR;Acc:AT3G60720],31,68,74,107,5.9,44.74,55,27.3,S2D_35084672,HD; MAT,-29907,5' proximal intergenic

TRIAE_CS42_2DS_TGACv1_177832_AA0585410.1,scaffold_501429.1,Transducin family protein [Source:UniProtKB/TrEMBL;Acc:D7LMP6],46,107,389,449,2.9,27.42,48,30.4,S2D_35084672,HD; MAT,-51428,5' proximal intergenic

TRIAE_CS42_2DS_TGACv1_177832_AA0585420.1,MLOC_69399.1,MLO-like protein [Source:UniProtKB/TrEMBL;Acc:M0YHK1],1,487,1,487,0,95.07,100,935,S2D_35084672,HD; MAT,-60367,5' proximal intergenic

TRIAE_CS42_2DS_TGACv1_177916_AA0587120.1,OB04G11740.1,Serine/threonine-protein kinase [Source:UniProtKB/TrEMBL;Acc:J3LVJ9],3,813,4,808,0,67.81,99,1136,S2D_9872868,HD,13537,5' proximal intergenic

TRIAE_CS42_2DS_TGACv1_177916_AA0587140.1,MLOC_55457.3,Haloacid dehalogenase-like hydrolase (HAD) superfamily protein [Source:Projected from Arabidopsis thaliana (AT1G79790) TAIR;Acc:AT1G79790],66,141,5,75,0.074,31.58,33,36.2,S2D_9872868,HD,2253,5' proximal intergenic

TRIAE_CS42_2DS_TGACv1_177916_AA0587150.1,EMT11322,Anthocyanin 3'-O-beta-glucosyltransferase [Source:UniProtKB/TrEMBL;Acc:M8B3M4],5,493,1,489,0,99.59,99,1003,S2D_9872868,HD,-255,5' proximal intergenic

TRIAE_CS42_2DS_TGACv1_177916_AA0587160.1,EMT20319,Disease resistance protein RGA2 [Source:UniProtKB/TrEMBL;Acc:N1R0R0],97,405,180,495,3.00E-109,55.14,99,336,S2D_9872868,HD,-5249,5' proximal intergenic

TRIAE_CS42_2DS_TGACv1_177916_AA0587170.1,EMT05736,Cytochrome P450 71D7 [Source:UniProtKB/TrEMBL;Acc:M8AMJ6],14,522,1,502,0,68.33,97,677,S2D_9872868,HD,-9759,5' proximal intergenic

TRIAE_CS42_2DS_TGACv1_177916_AA0587180.1,EMT11319,Obtusifoliol 14-alpha demethylase

[Source:UniProtKB/TrEMBL;Acc:M8BWY3],5,511,1,507,0,99.8,98,1056,S2D_9872868,H
D,-26448,5' proximal intergenic

TRIAE_CS42_3AS_TGACv1_211731_AA0693510.1,TRIUR3_10402-P1,Deoxyhypusine
synthase
[Source:UniProtKB/TrEMBL;Acc:M7ZR55],193,590,1,403,0,97.77,84,757,S3A_2023250
0,GSQM,2932,5' proximal intergenic

TRIAE_CS42_3AS_TGACv1_211731_AA0693510.2,TRIUR3_10402-P1,Deoxyhypusine
synthase
[Source:UniProtKB/TrEMBL;Acc:M7ZR55],203,641,1,403,0,88.74,84,735,S3A_2023250
0,GSQM,2932,5' proximal intergenic

TRIAE_CS42_3AS_TGACv1_213057_AA0705060.1,EMT17330,LRR receptor-like
serine/threonine-protein kinase EFR
[Source:UniProtKB/TrEMBL;Acc:N1R1Y3],21,637,1,610,0,83.63,84,1016,S3A_2493807
,SSQM,12251,intergenic

TRIAE_CS42_3AS_TGACv1_213057_AA0705060.2,EMT17330,LRR receptor-like
serine/threonine-protein kinase EFR
[Source:UniProtKB/TrEMBL;Acc:N1R1Y3],21,637,1,610,0,83.63,81,1013,S3A_2493807
,SSQM,12251,intergenic

TRIAE_CS42_3AS_TGACv1_213057_AA0705070.1,TRIUR3_05393-P1,60S ribosomal
protein L9 [Source:UniProtKB/TrEMBL;Acc:M7YJ37],1,185,1,185,4.00E-
132,98.92,84,376,S3A_2493807,SSQM,9411,intergenic

TRIAE_CS42_3AS_TGACv1_213057_AA0705070.2,TRIUR3_05393-P1,60S ribosomal
protein L9 [Source:UniProtKB/TrEMBL;Acc:M7YJ37],1,189,1,189,7.00E-
137,100,100,387,S3A_2493807,SSQM,9411,intergenic

TRIAE_CS42_3B_TGACv1_223320_AA0780170.1,EMT30985,Serine/threonine-protein
kinase CTR1
[Source:UniProtKB/TrEMBL;Acc:M8C131],1,437,1,409,0,66,100,533,S3B_695966897,T
WT,5276,missense

TRIAE_CS42_3B_TGACv1_223320_AA0780190.1,EMT32145,Signal peptide peptidase-
like 2B
[Source:UniProtKB/TrEMBL;Acc:M8C4C0],222,594,1,376,0,96.54,63,738,S3B_6959668
97,TWT,0,missense

TRIAE_CS42_3B_TGACv1_223320_AA0780190.2,EMT32145,Signal peptide peptidase-
like 2B
[Source:UniProtKB/TrEMBL;Acc:M8C4C0],192,564,1,376,0,96.54,66,734,S3B_6959668
97,TWT,0,missense

TRIAE_CS42_4AL_TGACv1_288585_AA0952930.1,Bra036804.1-P,AT4G23160 (E=9e-043) |
protein kinase family protein ,15,259,3,248,5.00E-
69,45.93,94,219,S4A_739598141,GSQM,85897,synonymous

TRIAE_CS42_4AL_TGACv1_288585_AA0952930.2,Bra036804.1-P,AT4G23160 (E=9e-043) |
protein kinase family protein ,15,259,3,248,5.00E-
69,45.93,94,219,S4A_739598141,GSQM,85897,synonymous

TRIAE_CS42_4AL_TGACv1_288585_AA0952940.1,EMT33735,LRR receptor-like
serine/threonine-protein kinase FLS2

[Source:UniProtKB/TrEMBL;Acc:N1R5T1],1,550,1,557,0,77.24,100,759,S4A_739598141,GSQM,75055,synonymous

TRIAE_CS42_4AL_TGACv1_288585_AA0952940.2,EMT33735,LRR receptor-like serine/threonine-protein kinase FLS2
[Source:UniProtKB/TrEMBL;Acc:N1R5T1],1,550,1,557,0,77.24,100,759,S4A_739598141,GSQM,75055,synonymous

TRIAE_CS42_4AL_TGACv1_288585_AA0952960.1,TRIUR3_09775-P1,F-box/kelch-repeat protein SKIP11
[Source:UniProtKB/TrEMBL;Acc:M7ZHU0],1,1172,1,1146,0,80.75,99,1825,S4A_739598141,GSQM,45095,synonymous

TRIAE_CS42_4AL_TGACv1_288585_AA0952970.1,EMT15955,Disease resistance protein RPP13
[Source:UniProtKB/TrEMBL;Acc:N1R150],546,984,423,935,0,70.82,82,666,S4A_739598141,GSQM,29009,synonymous

TRIAE_CS42_4AL_TGACv1_288585_AA0952990.1,EMT24075,Disease resistance protein RPP13
[Source:UniProtKB/TrEMBL;Acc:M8BGH7],1,905,1,900,0,82.84,97,1421,S4A_739598141,GSQM,5795,synonymous

TRIAE_CS42_4AL_TGACv1_288585_AA0953000.1,GSMUA_Achr10P08010_001,expressed protein
[Source:GMGC_GENE;Acc:GSMUA_Achr10G08010_001],27,740,22,725,0,53.81,96,743,S4A_739598141,GSQM,0,synonymous

TRIAE_CS42_4AL_TGACv1_288585_AA0953010.1,MLOC_53088.1,Terpenoid synthases superfamily protein [Source:Projected from Arabidopsis thaliana (AT1G62730) TAIR;Acc:AT1G62730],14,315,1,302,0,94.7,96,564,S4A_739598141,GSQM,0,synonymous

TRIAE_CS42_4AL_TGACv1_288585_AA0953020.1,TRIUR3_08851-P1,NAC domain-containing protein 7
[Source:UniProtKB/TrEMBL;Acc:M8ALU3],1,336,1,341,0,96.77,100,666,S4A_739598141,GSQM,-8862,synonymous

TRIAE_CS42_4AL_TGACv1_288738_AA0956940.1,EMT29067,Disease resistance protein RPS2
[Source:UniProtKB/TrEMBL;Acc:M8CNJ7],5,629,55,682,0,75.48,98,958,S4A_726716318,YLD,43929,intronic

TRIAE_CS42_4AL_TGACv1_288738_AA0956970.1,EMT29067,Disease resistance protein RPS2
[Source:UniProtKB/TrEMBL;Acc:M8CNJ7],1,634,463,1066,0,63.79,100,774,S4A_726716318,YLD,0,intronic

TRIAE_CS42_4AL_TGACv1_288738_AA0956970.2,EMT29067,Disease resistance protein RPS2
[Source:UniProtKB/TrEMBL;Acc:M8CNJ7],20,699,417,1066,0,64.62,97,848,S4A_726716318,YLD,0,intronic

TRIAE_CS42_4BL_TGACv1_320403_AA1037940.1,AET01348,pathogenic type III effector avirulence factor Avr AvrRpt-cleavage: cleavage site

protein,11,169,12,146,0.97,23.49,94,32,S4B_626390000,HI; STARCH,-19266,intergenic

TRIAE_CS42_5AS_TGACv1_393202_AA1269750.1,PGSC0003DMT400062565,Ripening regulated protein [Source:PGSC_GENE;Acc:PGSC0003DMG400024346],11,343,6,328,3.00E-142,63.17,96,412,S5A_9462259,STARCH,16024,intronic

TRIAE_CS42_5AS_TGACv1_393202_AA1269750.2,PGSC0003DMT400062565,Ripening regulated protein [Source:PGSC_GENE;Acc:PGSC0003DMG400024346],11,295,6,328,7.00E-115,57.41,96,341,S5A_9462259,STARCH,16024,intronic

TRIAE_CS42_5AS_TGACv1_393202_AA1269760.1,EOY19851,"Zinc finger, CCHC-type-like protein",92,267,50,221,2.00E-15,28.25,28,79.7,S5A_9462259,STARCH,12920,intronic

TRIAE_CS42_5AS_TGACv1_393202_AA1269770.1,KEH29019,endonuclease/exonuclease/phosphatase family protein,2,407,40,430,3.00E-58,29.31,97,204,S5A_9462259,STARCH,11429,intronic

TRIAE_CS42_5AS_TGACv1_393202_AA1269790.1,OS01T0624000-01,Neutral ceramidase [Source:Uniprot/SWISSPROT;Acc:Q0JL46],1,781,1,781,0,83.97,97,1335,S5A_9462259,STARCH,0,intronic

TRIAE_CS42_5AS_TGACv1_393202_AA1269790.2,OS01T0624000-01,Neutral ceramidase [Source:Uniprot/SWISSPROT;Acc:Q0JL46],1,783,1,785,0,84.52,100,1352,S5A_9462259,STARCH,0,intronic

TRIAE_CS42_5AS_TGACv1_393202_AA1269790.3,OS01T0624000-01,Neutral ceramidase [Source:Uniprot/SWISSPROT;Acc:Q0JL46],1,783,1,785,0,84.52,100,1352,S5A_9462259,STARCH,0,intronic

TRIAE_CS42_5AS_TGACv1_393202_AA1269790.4,OS01T0624000-01,Neutral ceramidase [Source:Uniprot/SWISSPROT;Acc:Q0JL46],1,783,1,785,0,84.52,100,1352,S5A_9462259,STARCH,0,intronic

TRIAE_CS42_5AS_TGACv1_393202_AA1269790.5,OS01T0624000-01,Neutral ceramidase [Source:Uniprot/SWISSPROT;Acc:Q0JL46],1,544,1,547,0,86.13,89,948,S5A_9462259,STARCH,0,intronic

TRIAE_CS42_5BL_TGACv1_404873_AA1313490.1,EOY31522,AP2 domain-containing transcription factor,39,396,30,350,2.00E-137,60.34,90,403,S5B_394707451,HI,-35687,intergenic

TRIAE_CS42_5BL_TGACv1_404873_AA1313500.1,BRADI4G30610.1,DnaJ/Hsp40 cysteine-rich domain superfamily protein [Source:Projected from Arabidopsis thaliana (AT2G24860) TAIR;Acc:AT2G24860],1,148,1,148,9.00E-90,87.16,100,264,S5B_394707451,HI,-43858,intergenic

TRIAE_CS42_5BL_TGACv1_405672_AA1332610.1,EMT22165,Agamous-like MADS-box protein AGL61 [Source:UniProtKB/TrEMBL;Acc:M8BBC6],1,173,1,173,2.00E-116,95.38,100,333,S5B_644947034,HI,3083,5' proximal intergenic

TRIAE_CS42_5BL_TGACv1_407709_AA1359020.1,AET02817,Myb/SANT-like DNA-binding domain protein,4,187,22,215,2.00E-14,29.35,62,73.9,S5B_261134879,HT,1901,3' proximal intergenic

TRIAE_CS42_5BS_TGACv1_424234_AA1387800.1,TRIUR3_21256-P1,WD repeat-containing protein 26 [Source:UniProtKB/TrEMBL;Acc:M7YN51],91,219,555,683,3.00E-19,37.98,13,97.4,S5B_34721398,STARCH,24703,premature stop

TRIAE_CS42_5BS_TGACv1_424234_AA1387810.1,EOY24040,RING/U-box superfamily protein isoform 1,235,312,74,179,2.00E-04,27.36,44,46.6,S5B_34721398,STARCH,21313,premature stop

TRIAE_CS42_5BS_TGACv1_424234_AA1387820.1,EMT17640,Disease resistance protein RPM1 [Source:UniProtKB/TrEMBL;Acc:N1R0E7],1,927,1,926,0,84.57,99,1606,S5B_34721398,STARCH,0,premature stop

TRIAE_CS42_5DL_TGACv1_433921_AA1425260.1,KEH32063,transcription factor BIM2-like protein,2,456,12,524,3.00E-73,34.43,90,248,S5D_365732020,STARCH,34762,3' UTR

TRIAE_CS42_5DL_TGACv1_433921_AA1425270.1,EOY22172,O-fucosyltransferase family protein isoform 1,12,495,33,499,0,72.16,96,747,S5D_365732020,STARCH,29557,3' UTR

TRIAE_CS42_5DL_TGACv1_433921_AA1425280.1,EMT16494,Prolyl oligopeptidase family protein [Source:Projected from Arabidopsis thaliana (AT1G69020) TAIR;Acc:AT1G69020],42,732,108,819,0,92.3,99,1268,S5D_365732020,STARCH,0,3' UTR

TRIAE_CS42_6AL_TGACv1_470886_AA1497500.1,EOX99084,TCP family transcription factor,1,187,50,257,2.00E-58,56.48,94,190,S6A_614373502,SSQM,-7719,intergenic

TRIAE_CS42_6AL_TGACv1_470886_AA1497540.1,PGSC0003DMT400002341,Thylakoid-bound ascorbate peroxidase 6 [Source:PGSC_GENE;Acc:PGSC0003DMG400000894],33,431,42,420,0,71.57,93,584,S6A_614373502,SSQM,-98281,intergenic

TRIAE_CS42_6AL_TGACv1_470886_AA1497540.2,OS04T0434800-01,"Probable L-ascorbate peroxidase 7, chloroplastic [Source:Uniprot/SWISSPROT;Acc:Q7XJ02]",6,340,32,358,5.00E-179,72.41,86,509,S6A_614373502,SSQM,-98281,intergenic

TRIAE_CS42_6AL_TGACv1_470886_AA1497550.1,EMT14678,Transcriptional corepressor LEUNIG [Source:UniProtKB/TrEMBL;Acc:N1QVQ4],43,91,977,1016,1.8,40.82,30,32,S6A_614373502,SSQM,-103371,intergenic

TRIAE_CS42_6AL_TGACv1_470886_AA1497560.1,OGLUM04G24450.6,DNA ligase IV [Source:Projected from Arabidopsis thaliana (AT5G57160) TAIR;Acc:AT5G57160],19,87,519,588,0.31,31.43,77,32.3,S6A_614373502,SSQM,-116361,intergenic

TRIAE_CS42_6AL_TGACv1_470886_AA1497580.1,EOX95444,P-glycoprotein 21,34,84,783,833,4,31.37,59,28.9,S6A_614373502,SSQM,-147006,intergenic

TRIAE_CS42_6AL_TGACv1_470886_AA1497590.1,EMT27748,"Chloroplastic group IIA intron splicing facilitator CRS1, chloroplastic [Source:UniProtKB/TrEMBL;Acc:M8C133]",87,143,134,190,0.86,41.38,33,33.1,S6A_614373502,SSQM,-161571,intergenic

TRIAE_CS42_6AL_TGACv1_470886_AA1497600.1,MLOC_49780.1,GAST1 protein homolog 4 [Source:Projected from Arabidopsis thaliana (AT5G15230) TAIR;Acc:AT5G15230],10,116,7,113,2.00E-16,38.53,92,73.9,S6A_614373502,SSQM,-175072,intergenic

TRIAE_CS42_6AL_TGACv1_470886_AA1497600.2,MLOC_36360.1,Gibberellin-regulated family protein [Source:Projected from Arabidopsis thaliana (AT2G30810) TAIR;Acc:AT2G30810],44,143,12,103,1.00E-16,42,70,75.1,S6A_614373502,SSQM,-175072,intergenic

TRIAE_CS42_6AL_TGACv1_470886_AA1497610.1,BRADI3G60400.1,polyprenyltransferase 1 [Source:Projected from Arabidopsis thaliana (AT4G23660) TAIR;Acc:AT4G23660],17,402,20,408,0,84.91,96,598,S6A_614373502,SSQM,-177572,intergenic

TRIAE_CS42_6AL_TGACv1_470886_AA1497610.2,BRADI3G60400.1,polyprenyltransferase 1 [Source:Projected from Arabidopsis thaliana (AT4G23660) TAIR;Acc:AT4G23660],17,402,20,408,0,84.91,96,598,S6A_614373502,SSQM,-177572,intergenic

TRIAE_CS42_6AL_TGACv1_470886_AA1497620.1,EMT08397,Replication protein A 32 kDa subunit [Source:UniProtKB/TrEMBL;Acc:M8AVA3],1,277,1,296,0,86.56,100,519,S6A_614373502,SSQM,-237991,intergenic

TRIAE_CS42_6AL_TGACv1_470886_AA1497630.1,PGSC0003DMT400000123,Kiwellin [Source:PGSC_GENE;Acc:PGSC0003DMG400000034],9,184,8,189,7.00E-60,54.1,96,191,S6A_614373502,SSQM,-249701,intergenic

TRIAE_CS42_6AL_TGACv1_470886_AA1497640.1,EOY26276,"Tyrosyl-tRNA synthetase, class Ib, bacterial/mitochondrial",17,365,37,385,0,69.17,95,534,S6A_614373502,SSQM,-254484,intergenic

TRIAE_CS42_6AL_TGACv1_471361_AA1507650.1,TRIUR3_06763-P1,Disease resistance RPP13-like protein 4 [Source:UniProtKB/TrEMBL;Acc:M7YKV0],1,540,1,539,0,79.07,98,871,S6A_614660970,TWT,28584,intronic

TRIAE_CS42_6AL_TGACv1_471361_AA1507650.2,TRIUR3_06763-P1,Disease resistance RPP13-like protein 4 [Source:UniProtKB/TrEMBL;Acc:M7YKV0],1,495,1,494,0,78.79,98,798,S6A_614660970,TWT,28584,intronic

TRIAE_CS42_6AL_TGACv1_471361_AA1507650.3,TRIUR3_06763-P1,Disease resistance RPP13-like protein 4 [Source:UniProtKB/TrEMBL;Acc:M7YKV0],1,495,1,494,0,78.79,98,798,S6A_614660970,TWT,28584,intronic

TRIAE_CS42_6AL_TGACv1_471361_AA1507660.1,EMT31140,F-box protein [Source:UniProtKB/TrEMBL;Acc:M8C1J6],18,363,82,422,1.00E-52,35.47,94,185,S6A_614660970,TWT,6387,intronic

TRIAE_CS42_6AL_TGACv1_471361_AA1507660.2,EMT31140,F-box protein [Source:UniProtKB/TrEMBL;Acc:M8C1J6],18,363,82,422,1.00E-52,35.47,94,185,S6A_614660970,TWT,6387,intronic

TRIAE_CS42_6AL_TGACv1_471361_AA1507660.3,EMT31140,F-box protein [Source:UniProtKB/TrEMBL;Acc:M8C1J6],18,298,82,363,2.00E-45,36.18,92,164,S6A_614660970,TWT,6387,intronic

TRIAE_CS42_6AL_TGACv1_471361_AA1507690.1,POPTR_0010s17740.1,Integral membrane single C2 domain protein [source:UniProtKB/TrEMBL;Acc:B9HY18_POPTR],94,675,111,669,0,64.77,86,715,S6A_614660970,TWT,-10174,intronic

TRIAE_CS42_6BL_TGACv1_499780_AA1591540.1,EMT03461,Glutamate receptor 2.7 [Source:UniProtKB/TrEMBL;Acc:N1QUK2],1,1014,1,992,0,80.75,100,1540,S6B_656279771,YLD,275,3' UTR

TRIAE_CS42_6BL_TGACv1_499780_AA1591550.1,Bo1g014280.1,"diacylglycerol kinase 7 [Source:TAIR;Acc:AT4G30340](projected from arabidopsis_thaliana,AT4G30340)",28,501,21,490,0,62.42,94,632,S6B_656279771,YLD,0,3' UTR

TRIAE_CS42_6BL_TGACv1_499780_AA1591550.2,Bo1g014280.1,"diacylglycerol kinase 7 [Source:TAIR;Acc:AT4G30340](projected from arabidopsis_thaliana,AT4G30340)",28,501,21,490,0,62.42,94,632,S6B_656279771,YLD,0,3' UTR

TRIAE_CS42_6BL_TGACv1_500675_AA1608280.1,TRIUR3_33630-P1,Cysteine-rich receptor-like protein kinase 19 [Source:UniProtKB/TrEMBL;Acc:M7YHJ8],1,315,1,332,8.00E-133,58.56,61,396,S6B_706326554,HT,0,intronic

TRIAE_CS42_6BL_TGACv1_500675_AA1608280.2,TRIUR3_33630-P1,Cysteine-rich receptor-like protein kinase 19 [Source:UniProtKB/TrEMBL;Acc:M7YHJ8],1,315,1,332,8.00E-133,58.56,61,396,S6B_706326554,HT,0,intronic

TRIAE_CS42_6BL_TGACv1_500675_AA1608290.1,EMT29395,Disease resistance protein RGA2 [Source:UniProtKB/TrEMBL;Acc:M8BWJ8],18,768,145,899,0,81.85,97,1244,S6B_706326554,HT,-7221,intronic

TRIAE_CS42_6DL_TGACv1_526541_AA1686500.1,EMT22911,Wall-associated receptor kinase 2 [Source:UniProtKB/TrEMBL;Acc:M8CI25],8,641,2,636,0,84.41,84,1112,S6D_468113959,TKW,51870,intergenic

TRIAE_CS42_6DL_TGACv1_526541_AA1686500.2,EMT22915,Wall-associated receptor kinase 1 [Source:UniProtKB/TrEMBL;Acc:M8BDC6],8,702,2,741,0,72.74,98,1052,S6D_468113959,TKW,51870,intergenic

TRIAE_CS42_6DL_TGACv1_526541_AA1686510.1,TRIUR3_12904-P1,Alpha-latroinsectotoxin-Lt1a [Source:UniProtKB/TrEMBL;Acc:M7YXF1],25,499,26,456,2.00E-111,44.87,66,352,S6D_468113959,TKW,21580,intergenic

TRIAE_CS42_6DL_TGACv1_526541_AA1686510.2,TRIUR3_12904-P1,Alpha-latroinsectotoxin-Lt1a

[Source:UniProtKB/TrEMBL;Acc:M7YXF1],25,471,26,456,2.00E-117,47.55,65,367,S6D_468113959,TKW,21580,intergenic

TRIAE_CS42_6DL_TGACv1_526541_AA1686520.1,EMT22911,Wall-associated receptor kinase 2 [Source:UniProtKB/TrEMBL;Acc:M8CI25],1,635,1,636,0,98.9,85,1289,S6D_468113959,TKW,13781,intergenic

TRIAE_CS42_6DL_TGACv1_526541_AA1686560.1,EMT22915,Wall-associated receptor kinase 1 [Source:UniProtKB/TrEMBL;Acc:M8BDC6],1,743,1,741,0,72.55,99,1075,S6D_468113959,TKW,-30214,intergenic

TRIAE_CS42_6DL_TGACv1_526541_AA1686570.1,EMT22911,Wall-associated receptor kinase 2 [Source:UniProtKB/TrEMBL;Acc:M8CI25],1,247,390,636,2.00E-160,91.9,98,466,S6D_468113959,TKW,-51314,intergenic

TRIAE_CS42_6DL_TGACv1_526541_AA1686580.1,EMT22911,Wall-associated receptor kinase 2 [Source:UniProtKB/TrEMBL;Acc:M8CI25],118,419,1,304,2.00E-156,75,69,463,S6D_468113959,TKW,-53992,intergenic

TRIAE_CS42_6DL_TGACv1_526541_AA1686590.1,EMT22915,Wall-associated receptor kinase 1 [Source:UniProtKB/TrEMBL;Acc:M8BDC6],1,721,1,741,0,96.09,100,1429,S6D_468113959,TKW,-64138,intergenic

TRIAE_CS42_6DL_TGACv1_526541_AA1686590.2,EMT22915,Wall-associated receptor kinase 1 [Source:UniProtKB/TrEMBL;Acc:M8BDC6],1,741,1,741,0,98.79,100,1520,S6D_468113959,TKW,-64138,intergenic

TRIAE_CS42_6DL_TGACv1_526541_AA1686590.3,EMT22915,Wall-associated receptor kinase 1 [Source:UniProtKB/TrEMBL;Acc:M8BDC6],1,569,173,741,0,98.59,100,1158,S6D_468113959,TKW,-64138,intergenic

TRIAE_CS42_6DS_TGACv1_543118_AA1735680.1,EMT26516,Fumarylacetoacetase [Source:UniProtKB/TrEMBL;Acc:M8CS44],1,427,1,427,0,100,100,883,S6D_127384672,HD,5621,missense

TRIAE_CS42_6DS_TGACv1_543118_AA1735680.2,EMT26516,Fumarylacetoacetase [Source:UniProtKB/TrEMBL;Acc:M8CS44],1,348,80,427,0,100,100,724,S6D_127384672,HD,5621,missense

TRIAE_CS42_6DS_TGACv1_543118_AA1735690.1,TRIUR3_12627-P1,Exopolygalacturonase [Source:UniProtKB/TrEMBL;Acc:M7Y8Y4],52,416,1,365,0,98.36,88,736,S6D_127384672,HD,3960,missense

TRIAE_CS42_6DS_TGACv1_543118_AA1735700.1,POPTR_0001s05650.1,Heavy metal ATPase [Source:UniProtKB/TrEMBL;Acc:B9GM73],1,672,300,971,0,75.15,98,1047,S6D_127384672,HD,0,missense

TRIAE_CS42_6DS_TGACv1_543118_AA1735700.2,POPTR_0001s05650.1,Heavy metal ATPase [Source:UniProtKB/TrEMBL;Acc:B9GM73],1,672,300,971,0,75.15,98,1047,S6D_127384672,HD,0,missense

TRIAE_CS42_7DS_TGACv1_622088_AA2032510.1,EMT22045,"Cysteine synthase, chloroplastic/chromoplastic [Source:UniProtKB/TrEMBL;Acc:M8BB06]",1,386,1,439,0,87.24,97,726,S7D_58589271,HD; MAT,5823,intronic

TRIAE_CS42_7DS_TGACv1_622088_AA2032510.2,EMT22045,"Cysteine synthase, chloroplastic/chromoplastic [Source:UniProtKB/TrEMBL;Acc:M8BB06]",13,273,179,439,0,99.23,92,528,S7D_58589271,HD; MAT,5823,intronic

TRIAE_CS42_7DS_TGACv1_622088_AA2032510.3,TRIUR3_02991-P1,"Cysteine synthase, chloroplastic/chromoplastic [Source:UniProtKB/TrEMBL;Acc:M7YWM0]",1,258,97,354,0,99.22,96,520,S7D_58589271,HD; MAT,5823,intronic

TRIAE_CS42_7DS_TGACv1_622088_AA2032520.1,BRADI3G36140.1,3-methylcrotonyl-CoA carboxylase [Source:Projected from Arabidopsis thaliana (AT4G34030) TAIR;Acc:AT4G34030],64,628,1,581,0,90.53,90,1072,S7D_58589271,HD; MAT,0,intronic

TRIAE_CS42_7DS_TGACv1_622088_AA2032520.2,BRADI3G36140.1,3-methylcrotonyl-CoA carboxylase [Source:Projected from Arabidopsis thaliana (AT4G34030) TAIR;Acc:AT4G34030],64,422,1,375,0,85.87,80,647,S7D_58589271,HD; MAT,0,intronic

TRIAE_CS42_7DS_TGACv1_622088_AA2032530.1,Si036043m,catalytics;transferases;[acyl-carrier-protein] S-malonyltransferases;binding [Source:Projected from Arabidopsis thaliana (AT2G30200) TAIR;Acc:AT2G30200],1,76,156,237,2.2,36.59,64,30.4,S7D_58589271,HD; MAT,-9314,intronic

TRIAE_CS42_7DS_TGACv1_623041_AA2048930.1,TRIUR3_05385-P1,Nucleosome assembly protein 1-like 1-A [Source:UniProtKB/TrEMBL;Acc:M7ZTA3],38,382,132,476,0,97.1,90,558,S7D_58449294,FLS,7085,missense

TRIAE_CS42_7DS_TGACv1_623041_AA2048930.2,TRIUR3_05385-P1,Nucleosome assembly protein 1-like 1-A [Source:UniProtKB/TrEMBL;Acc:M7ZTA3],38,362,132,456,2.00E-180,97.23,89,516,S7D_58449294,FLS,7085,missense

TRIAE_CS42_7DS_TGACv1_623041_AA2048940.1,EDO97709,3-ketoacyl-CoA-synthase [Source:UniProtKB/TrEMBL;Acc:A8JEF7],9,49,469,509,3.8,41.46,63,28.1,S7D_58449294,FLS,2534,missense

TRIAE_CS42_7DS_TGACv1_623041_AA2048950.1,KEH23680,PPR containing plant-like protein,4,636,14,659,0,46.19,99,603,S7D_58449294,FLS,0,missense

TRIAE_CS42_7DS_TGACv1_623041_AA2048960.1,TRIUR3_06168-P1,Disease resistance protein RPM1 [Source:UniProtKB/TrEMBL;Acc:M7YDZ0],1,1277,1,1276,0,78.4,100,2022,S7D_58449294,FLS,-1094,missense

TRIAE_CS42_7DS_TGACv1_623041_AA2048970.1,GSMUA_Achr11P19570_001,Probable inactive purple acid phosphatase 27

[Source:GMGC_GENE;Acc:GSMUA_Achr11G19570_001],26,621,34,622,0,69.01,96,885,S7D_58449294,FLS,-9177,missense

TRIAE_CS42_7DS_TGACv1_623041_AA2048970.2,GSMUA_Achr11P19570_001,Probable inactive purple acid phosphatase 27 [Source:GMGC_GENE;Acc:GSMUA_Achr11G19570_001],26,613,34,622,0,68.01,96,866,S7D_58449294,FLS,-9177,missense

TRIAE_CS42_U_TGACv1_641320_AA2091830.1,EMT17640,Disease resistance protein RPM1 [Source:UniProtKB/TrEMBL;Acc:N1R0E7],36,98,642,707,0.31,31.82,64,32.3,S5B_396479359,GSQM,9205,intergenic

TRIAE_CS42_U_TGACv1_641320_AA2091850.1,EMT30163,Callose synthase 4 [Source:UniProtKB/TrEMBL;Acc:M8C7W7],13,421,41,456,0,78.95,83,685,S5B_396479359,GSQM,-27515,intergenic

TRIAE_CS42_U_TGACv1_641320_AA2091850.2,EMT30163,Callose synthase 4 [Source:UniProtKB/TrEMBL;Acc:M8C7W7],9,355,106,456,0,78.37,82,573,S5B_396479359,GSQM,-27515,intergenic

TRIAE_CS42_U_TGACv1_641320_AA2091850.3,EMT30163,Callose synthase 4 [Source:UniProtKB/TrEMBL;Acc:M8C7W7],9,354,106,456,0,78.31,82,576,S5B_396479359,GSQM,-27515,intergenic

TRIAE_CS42_U_TGACv1_641320_AA2091860.1,BGIOSGA005297-PA,"Outer envelope pore protein 21, chloroplastic [Source:UniProtKB/Swiss-Prot;Acc:B8AFI8]",60,232,1,185,3.00E-71,61.62,75,221,S5B_396479359,GSQM,-34236,intergenic

TRIAE_CS42_U_TGACv1_641320_AA2091870.1,EMT28537,Callose synthase 3 [Source:UniProtKB/TrEMBL;Acc:M8CM74],49,1973,1,1859,0,79.54,97,3112,S5B_396479359,GSQM,-38175,intergenic

TRIAE_CS42_U_TGACv1_641320_AA2091900.1,PGSC0003DMT400033774,F-box family protein [Source:PGSC_GENE;Acc:PGSC0003DMG400012976],10,110,8,107,0.08,32.11,28,37.4,S5B_396479359,GSQM,-77429,intergenic

TRIAE_CS42_U_TGACv1_641320_AA2091910.1,TRIUR3_15204-P1,RING-H2 finger protein ATL44 [Source:UniProtKB/TrEMBL;Acc:M7ZBG6],1,174,34,204,7.00E-90,93.1,100,267,S5B_396479359,GSQM,-88485,intergenic

APPENDIX E: Within-environment mean GEBV-phenotype correlations across 500 Monte Carlo replications for univariate and multi-environment models used in chapter II

The table below presents the within-environment mean GEBV-phenotype correlations (Mean r), their standard deviations (SD), standard errors (SE) and 95% confidence intervals (CI) calculated across 500 Monte Carlo replications for various traits, ratios of validation population size:training population size (Prop_Miss) and models.

Table E.1: Within-environment mean GEBV-phenotype correlations across 500 Monte Carlo replications for univariate and multi-environment models used in chapter II

Trait †	Prop_Miss	Env ‡	Mean r	SD	SE	CI	Model §
BIOM	0.2	14Bb	0.12	0.1389	0.0062	0.0122	Adj Means
BIOM	0.2	14War	0.16	0.1486	0.0066	0.0131	Adj Means
BIOM	0.2	15Bb	0.27	0.1531	0.0068	0.0135	Adj Means
BIOM	0.2	15War	0.27	0.1483	0.0066	0.0130	Adj Means
BIOM	0.4	14Bb	0.12	0.0856	0.0038	0.0075	Adj Means
BIOM	0.4	14War	0.15	0.0886	0.0040	0.0078	Adj Means
BIOM	0.4	15Bb	0.26	0.0915	0.0041	0.0080	Adj Means
BIOM	0.4	15War	0.27	0.0914	0.0041	0.0080	Adj Means
BIOM	0.6	14Bb	0.11	0.0642	0.0029	0.0056	Adj Means
BIOM	0.6	14War	0.15	0.0659	0.0029	0.0058	Adj Means
BIOM	0.6	15Bb	0.25	0.0661	0.0030	0.0058	Adj Means
BIOM	0.6	15War	0.25	0.0714	0.0032	0.0063	Adj Means
BIOM	0.8	14Bb	0.09	0.0563	0.0025	0.0049	Adj Means
BIOM	0.8	14War	0.14	0.0557	0.0025	0.0049	Adj Means
BIOM	0.8	15Bb	0.22	0.0655	0.0029	0.0058	Adj Means
BIOM	0.8	15War	0.22	0.0696	0.0031	0.0061	Adj Means
FLSG	0.2	14Bb	0.22	0.1547	0.0069	0.0136	Adj Means
FLSG	0.2	14War	0.51	0.1137	0.0051	0.0100	Adj Means
FLSG	0.2	15Bb	0.52	0.1091	0.0049	0.0096	Adj Means
FLSG	0.2	15War	0.48	0.1167	0.0052	0.0103	Adj Means
FLSG	0.4	14Bb	0.20	0.1062	0.0047	0.0093	Adj Means
FLSG	0.4	14War	0.48	0.0713	0.0032	0.0063	Adj Means
FLSG	0.4	15Bb	0.51	0.0686	0.0031	0.0060	Adj Means
FLSG	0.4	15War	0.47	0.0740	0.0033	0.0065	Adj Means
FLSG	0.6	14Bb	0.18	0.0728	0.0033	0.0064	Adj Means
FLSG	0.6	14War	0.46	0.0597	0.0027	0.0052	Adj Means
FLSG	0.6	15Bb	0.48	0.0525	0.0023	0.0046	Adj Means
FLSG	0.6	15War	0.44	0.0524	0.0023	0.0046	Adj Means
FLSG	0.8	14Bb	0.14	0.0602	0.0027	0.0053	Adj Means
FLSG	0.8	14War	0.41	0.0667	0.0030	0.0059	Adj Means
FLSG	0.8	15Bb	0.43	0.0525	0.0023	0.0046	Adj Means

Trait †	Prop_Miss	Env ‡	Mean r	SD	SE	CI	Model §
FLSG	0.8	15War	0.40	0.0547	0.0024	0.0048	Adj Means
GSQM	0.2	14Bb	0.46	0.1247	0.0056	0.0110	Adj Means
GSQM	0.2	14War	0.29	0.1572	0.0070	0.0138	Adj Means
GSQM	0.2	15Bb	0.15	0.1429	0.0064	0.0126	Adj Means
GSQM	0.2	15War	0.35	0.1346	0.0060	0.0118	Adj Means
GSQM	0.4	14Bb	0.42	0.0896	0.0040	0.0079	Adj Means
GSQM	0.4	14War	0.26	0.0952	0.0043	0.0084	Adj Means
GSQM	0.4	15Bb	0.11	0.0949	0.0042	0.0083	Adj Means
GSQM	0.4	15War	0.30	0.0976	0.0044	0.0086	Adj Means
GSQM	0.6	14Bb	0.36	0.0781	0.0035	0.0069	Adj Means
GSQM	0.6	14War	0.22	0.0766	0.0034	0.0067	Adj Means
GSQM	0.6	15Bb	0.07	0.0759	0.0034	0.0067	Adj Means
GSQM	0.6	15War	0.24	0.0858	0.0038	0.0075	Adj Means
GSQM	0.8	14Bb	0.24	0.0982	0.0044	0.0086	Adj Means
GSQM	0.8	14War	0.14	0.0781	0.0035	0.0069	Adj Means
GSQM	0.8	15Bb	0.02	0.0682	0.0031	0.0060	Adj Means
GSQM	0.8	15War	0.16	0.0872	0.0039	0.0077	Adj Means
GW	0.2	14Bb	0.14	0.1393	0.0062	0.0122	Adj Means
GW	0.2	14War	0.14	0.1466	0.0066	0.0129	Adj Means
GW	0.2	15Bb	0.34	0.1389	0.0062	0.0122	Adj Means
GW	0.2	15War	0.31	0.1277	0.0057	0.0112	Adj Means
GW	0.4	14Bb	0.13	0.0908	0.0041	0.0080	Adj Means
GW	0.4	14War	0.14	0.0964	0.0043	0.0085	Adj Means
GW	0.4	15Bb	0.33	0.0886	0.0040	0.0078	Adj Means
GW	0.4	15War	0.30	0.0851	0.0038	0.0075	Adj Means
GW	0.6	14Bb	0.12	0.0681	0.0030	0.0060	Adj Means
GW	0.6	14War	0.13	0.0653	0.0029	0.0057	Adj Means
GW	0.6	15Bb	0.32	0.0677	0.0030	0.0059	Adj Means
GW	0.6	15War	0.29	0.0703	0.0031	0.0062	Adj Means
GW	0.8	14Bb	0.08	0.0675	0.0030	0.0059	Adj Means
GW	0.8	14War	0.12	0.0589	0.0026	0.0052	Adj Means
GW	0.8	15Bb	0.29	0.0827	0.0037	0.0073	Adj Means
GW	0.8	15War	0.26	0.0779	0.0035	0.0068	Adj Means
HD	0.2	14Bb	0.31	0.1344	0.0060	0.0118	Adj Means
HD	0.2	14War	0.44	0.1267	0.0057	0.0111	Adj Means
HD	0.2	15Bb	0.63	0.0884	0.0040	0.0078	Adj Means
HD	0.2	15War	0.60	0.0928	0.0041	0.0082	Adj Means
HD	0.4	14Bb	0.30	0.0923	0.0041	0.0081	Adj Means
HD	0.4	14War	0.42	0.0798	0.0036	0.0070	Adj Means
HD	0.4	15Bb	0.59	0.0641	0.0029	0.0056	Adj Means
HD	0.4	15War	0.56	0.0654	0.0029	0.0057	Adj Means
HD	0.6	14Bb	0.29	0.0691	0.0031	0.0061	Adj Means
HD	0.6	14War	0.40	0.0645	0.0029	0.0057	Adj Means

Trait †	Prop_Miss	Env ‡	Mean r	SD	SE	CI	Model §
HD	0.6	15Bb	0.54	0.0551	0.0025	0.0048	Adj Means
HD	0.6	15War	0.51	0.0566	0.0025	0.0050	Adj Means
HD	0.8	14Bb	0.26	0.0710	0.0032	0.0062	Adj Means
HD	0.8	14War	0.35	0.0742	0.0033	0.0065	Adj Means
HD	0.8	15Bb	0.45	0.0615	0.0028	0.0054	Adj Means
HD	0.8	15War	0.42	0.0621	0.0028	0.0055	Adj Means
HI	0.2	14Bb	0.45	0.1515	0.0068	0.0133	Adj Means
HI	0.2	14War	0.40	0.1452	0.0065	0.0128	Adj Means
HI	0.2	15Bb	0.38	0.1277	0.0057	0.0112	Adj Means
HI	0.2	15War	0.20	0.1382	0.0062	0.0121	Adj Means
HI	0.4	14Bb	0.45	0.0856	0.0038	0.0075	Adj Means
HI	0.4	14War	0.39	0.0901	0.0040	0.0079	Adj Means
HI	0.4	15Bb	0.36	0.0924	0.0041	0.0081	Adj Means
HI	0.4	15War	0.20	0.0877	0.0039	0.0077	Adj Means
HI	0.6	14Bb	0.43	0.0678	0.0030	0.0060	Adj Means
HI	0.6	14War	0.36	0.0662	0.0030	0.0058	Adj Means
HI	0.6	15Bb	0.33	0.0801	0.0036	0.0070	Adj Means
HI	0.6	15War	0.17	0.0704	0.0031	0.0062	Adj Means
HI	0.8	14Bb	0.37	0.0785	0.0035	0.0069	Adj Means
HI	0.8	14War	0.29	0.0707	0.0032	0.0062	Adj Means
HI	0.8	15Bb	0.28	0.1188	0.0053	0.0104	Adj Means
HI	0.8	15War	0.15	0.0782	0.0035	0.0069	Adj Means
HT	0.2	14Bb	0.46	0.1268	0.0057	0.0111	Adj Means
HT	0.2	14War	0.40	0.1528	0.0068	0.0134	Adj Means
HT	0.2	15Bb	0.63	0.1095	0.0049	0.0096	Adj Means
HT	0.2	15War	0.55	0.1259	0.0056	0.0111	Adj Means
HT	0.4	14Bb	0.44	0.0765	0.0034	0.0067	Adj Means
HT	0.4	14War	0.38	0.0940	0.0042	0.0083	Adj Means
HT	0.4	15Bb	0.61	0.0607	0.0027	0.0053	Adj Means
HT	0.4	15War	0.52	0.0733	0.0033	0.0064	Adj Means
HT	0.6	14Bb	0.42	0.0599	0.0027	0.0053	Adj Means
HT	0.6	14War	0.35	0.0716	0.0032	0.0063	Adj Means
HT	0.6	15Bb	0.57	0.0499	0.0022	0.0044	Adj Means
HT	0.6	15War	0.49	0.0571	0.0026	0.0050	Adj Means
HT	0.8	14Bb	0.37	0.0628	0.0028	0.0055	Adj Means
HT	0.8	14War	0.29	0.0755	0.0034	0.0066	Adj Means
HT	0.8	15Bb	0.51	0.0641	0.0029	0.0056	Adj Means
HT	0.8	15War	0.42	0.0616	0.0028	0.0054	Adj Means
MAT	0.2	14Bb	0.33	0.1393	0.0062	0.0122	Adj Means
MAT	0.2	14War	0.34	0.1312	0.0059	0.0115	Adj Means
MAT	0.2	15Bb	0.52	0.1110	0.0050	0.0098	Adj Means
MAT	0.2	15War	0.50	0.1080	0.0048	0.0095	Adj Means
MAT	0.4	14Bb	0.29	0.0885	0.0040	0.0078	Adj Means

Trait †	Prop_Miss	Env ‡	Mean r	SD	SE	CI	Model §
MAT	0.4	14War	0.30	0.0875	0.0039	0.0077	Adj Means
MAT	0.4	15Bb	0.48	0.0807	0.0036	0.0071	Adj Means
MAT	0.4	15War	0.46	0.0798	0.0036	0.0070	Adj Means
MAT	0.6	14Bb	0.26	0.0696	0.0031	0.0061	Adj Means
MAT	0.6	14War	0.26	0.0680	0.0030	0.0060	Adj Means
MAT	0.6	15Bb	0.42	0.0722	0.0032	0.0063	Adj Means
MAT	0.6	15War	0.41	0.0663	0.0030	0.0058	Adj Means
MAT	0.8	14Bb	0.22	0.0622	0.0028	0.0055	Adj Means
MAT	0.8	14War	0.21	0.0724	0.0032	0.0064	Adj Means
MAT	0.8	15Bb	0.35	0.0813	0.0036	0.0071	Adj Means
MAT	0.8	15War	0.32	0.0727	0.0033	0.0064	Adj Means
NDVI	0.2	14Bb	0.30	0.1377	0.0062	0.0121	Adj Means
NDVI	0.2	14War	0.37	0.1193	0.0053	0.0105	Adj Means
NDVI	0.2	15Bb	0.35	0.1327	0.0059	0.0117	Adj Means
NDVI	0.2	15War	0.26	0.1516	0.0068	0.0133	Adj Means
NDVI	0.4	14Bb	0.28	0.0858	0.0038	0.0075	Adj Means
NDVI	0.4	14War	0.35	0.0822	0.0037	0.0072	Adj Means
NDVI	0.4	15Bb	0.33	0.0857	0.0038	0.0075	Adj Means
NDVI	0.4	15War	0.25	0.1026	0.0046	0.0090	Adj Means
NDVI	0.6	14Bb	0.26	0.0647	0.0029	0.0057	Adj Means
NDVI	0.6	14War	0.33	0.0650	0.0029	0.0057	Adj Means
NDVI	0.6	15Bb	0.28	0.0694	0.0031	0.0061	Adj Means
NDVI	0.6	15War	0.24	0.0752	0.0034	0.0066	Adj Means
NDVI	0.8	14Bb	0.21	0.0697	0.0031	0.0061	Adj Means
NDVI	0.8	14War	0.27	0.0747	0.0033	0.0066	Adj Means
NDVI	0.8	15Bb	0.23	0.0783	0.0035	0.0069	Adj Means
NDVI	0.8	15War	0.18	0.0762	0.0034	0.0067	Adj Means
SPH	0.2	14Bb	0.35	0.1361	0.0061	0.0120	Adj Means
SPH	0.2	14War	0.32	0.1321	0.0059	0.0116	Adj Means
SPH	0.2	15Bb	0.35	0.1247	0.0056	0.0110	Adj Means
SPH	0.2	15War	0.36	0.1332	0.0060	0.0117	Adj Means
SPH	0.4	14Bb	0.27	0.0926	0.0041	0.0081	Adj Means
SPH	0.4	14War	0.25	0.0890	0.0040	0.0078	Adj Means
SPH	0.4	15Bb	0.31	0.0941	0.0042	0.0083	Adj Means
SPH	0.4	15War	0.30	0.0959	0.0043	0.0084	Adj Means
SPH	0.6	14Bb	0.18	0.0900	0.0040	0.0079	Adj Means
SPH	0.6	14War	0.18	0.0756	0.0034	0.0066	Adj Means
SPH	0.6	15Bb	0.26	0.0805	0.0036	0.0071	Adj Means
SPH	0.6	15War	0.22	0.0798	0.0036	0.0070	Adj Means
SPH	0.8	14Bb	0.11	0.0860	0.0038	0.0076	Adj Means
SPH	0.8	14War	0.12	0.0715	0.0032	0.0063	Adj Means
SPH	0.8	15Bb	0.19	0.0978	0.0044	0.0086	Adj Means
SPH	0.8	15War	0.14	0.0807	0.0036	0.0071	Adj Means

Trait †	Prop_Miss	Env ‡	Mean r	SD	SE	CI	Model §
SSQM	0.2	14Bb	0.45	0.1126	0.0050	0.0099	Adj Means
SSQM	0.2	14War	0.43	0.1329	0.0059	0.0117	Adj Means
SSQM	0.2	15Bb	0.39	0.1233	0.0055	0.0108	Adj Means
SSQM	0.2	15War	0.47	0.1204	0.0054	0.0106	Adj Means
SSQM	0.4	14Bb	0.40	0.0873	0.0039	0.0077	Adj Means
SSQM	0.4	14War	0.39	0.0891	0.0040	0.0078	Adj Means
SSQM	0.4	15Bb	0.36	0.0823	0.0037	0.0072	Adj Means
SSQM	0.4	15War	0.44	0.0826	0.0037	0.0073	Adj Means
SSQM	0.6	14Bb	0.33	0.0716	0.0032	0.0063	Adj Means
SSQM	0.6	14War	0.34	0.0721	0.0032	0.0063	Adj Means
SSQM	0.6	15Bb	0.30	0.0686	0.0031	0.0060	Adj Means
SSQM	0.6	15War	0.39	0.0666	0.0030	0.0059	Adj Means
SSQM	0.8	14Bb	0.22	0.0766	0.0034	0.0067	Adj Means
SSQM	0.8	14War	0.25	0.0710	0.0032	0.0062	Adj Means
SSQM	0.8	15Bb	0.22	0.0690	0.0031	0.0061	Adj Means
SSQM	0.8	15War	0.29	0.0797	0.0036	0.0070	Adj Means
STARCH	0.2	14Bb	0.44	0.1259	0.0056	0.0111	Adj Means
STARCH	0.2	14War	0.42	0.1169	0.0052	0.0103	Adj Means
STARCH	0.2	15Bb	0.30	0.1392	0.0062	0.0122	Adj Means
STARCH	0.2	15War	0.32	0.1362	0.0061	0.0120	Adj Means
STARCH	0.4	14Bb	0.42	0.0838	0.0037	0.0074	Adj Means
STARCH	0.4	14War	0.40	0.0802	0.0036	0.0070	Adj Means
STARCH	0.4	15Bb	0.27	0.1010	0.0045	0.0089	Adj Means
STARCH	0.4	15War	0.33	0.0919	0.0041	0.0081	Adj Means
STARCH	0.6	14Bb	0.39	0.0609	0.0027	0.0054	Adj Means
STARCH	0.6	14War	0.37	0.0662	0.0030	0.0058	Adj Means
STARCH	0.6	15Bb	0.25	0.0872	0.0039	0.0077	Adj Means
STARCH	0.6	15War	0.32	0.0814	0.0036	0.0072	Adj Means
STARCH	0.8	14Bb	0.33	0.0682	0.0031	0.0060	Adj Means
STARCH	0.8	14War	0.30	0.0760	0.0034	0.0067	Adj Means
STARCH	0.8	15Bb	0.22	0.1025	0.0046	0.0090	Adj Means
STARCH	0.8	15War	0.29	0.0838	0.0037	0.0074	Adj Means
TKW	0.2	14Bb	0.63	0.1103	0.0049	0.0097	Adj Means
TKW	0.2	14War	0.69	0.0959	0.0043	0.0084	Adj Means
TKW	0.2	15Bb	0.62	0.1017	0.0045	0.0089	Adj Means
TKW	0.2	15War	0.65	0.0979	0.0044	0.0086	Adj Means
TKW	0.4	14Bb	0.60	0.0779	0.0035	0.0068	Adj Means
TKW	0.4	14War	0.66	0.0686	0.0031	0.0060	Adj Means
TKW	0.4	15Bb	0.57	0.0680	0.0030	0.0060	Adj Means
TKW	0.4	15War	0.62	0.0710	0.0032	0.0062	Adj Means
TKW	0.6	14Bb	0.54	0.0696	0.0031	0.0061	Adj Means
TKW	0.6	14War	0.60	0.0676	0.0030	0.0059	Adj Means
TKW	0.6	15Bb	0.53	0.0554	0.0025	0.0049	Adj Means

Trait †	Prop_Miss	Env ‡	Mean r	SD	SE	CI	Model §
TKW	0.6	15War	0.57	0.0551	0.0025	0.0048	Adj Means
TKW	0.8	14Bb	0.43	0.0773	0.0035	0.0068	Adj Means
TKW	0.8	14War	0.49	0.0789	0.0035	0.0069	Adj Means
TKW	0.8	15Bb	0.46	0.0537	0.0024	0.0047	Adj Means
TKW	0.8	15War	0.49	0.0578	0.0026	0.0051	Adj Means
TWT	0.2	14Bb	0.55	0.1253	0.0056	0.0110	Adj Means
TWT	0.2	14War	0.45	0.1439	0.0064	0.0126	Adj Means
TWT	0.2	15Bb	0.57	0.1041	0.0047	0.0091	Adj Means
TWT	0.2	15War	0.56	0.0997	0.0045	0.0088	Adj Means
TWT	0.4	14Bb	0.54	0.0869	0.0039	0.0076	Adj Means
TWT	0.4	14War	0.45	0.0833	0.0037	0.0073	Adj Means
TWT	0.4	15Bb	0.53	0.0691	0.0031	0.0061	Adj Means
TWT	0.4	15War	0.52	0.0684	0.0031	0.0060	Adj Means
TWT	0.6	14Bb	0.51	0.0669	0.0030	0.0059	Adj Means
TWT	0.6	14War	0.41	0.0688	0.0031	0.0060	Adj Means
TWT	0.6	15Bb	0.50	0.0578	0.0026	0.0051	Adj Means
TWT	0.6	15War	0.47	0.0640	0.0029	0.0056	Adj Means
TWT	0.8	14Bb	0.45	0.0709	0.0032	0.0062	Adj Means
TWT	0.8	14War	0.35	0.0769	0.0034	0.0068	Adj Means
TWT	0.8	15Bb	0.44	0.0660	0.0030	0.0058	Adj Means
TWT	0.8	15War	0.39	0.0685	0.0031	0.0060	Adj Means
WCPROT	0.2	14Bb	0.22	0.1559	0.0070	0.0137	Adj Means
WCPROT	0.2	14War	0.17	0.1406	0.0063	0.0124	Adj Means
WCPROT	0.2	15Bb	0.29	0.1347	0.0060	0.0118	Adj Means
WCPROT	0.2	15War	0.34	0.1367	0.0061	0.0120	Adj Means
WCPROT	0.4	14Bb	0.20	0.1017	0.0045	0.0089	Adj Means
WCPROT	0.4	14War	0.15	0.0933	0.0042	0.0082	Adj Means
WCPROT	0.4	15Bb	0.26	0.0923	0.0041	0.0081	Adj Means
WCPROT	0.4	15War	0.32	0.0917	0.0041	0.0081	Adj Means
WCPROT	0.6	14Bb	0.17	0.0876	0.0039	0.0077	Adj Means
WCPROT	0.6	14War	0.13	0.0726	0.0032	0.0064	Adj Means
WCPROT	0.6	15Bb	0.22	0.0761	0.0034	0.0067	Adj Means
WCPROT	0.6	15War	0.29	0.0727	0.0033	0.0064	Adj Means
WCPROT	0.8	14Bb	0.12	0.0965	0.0043	0.0085	Adj Means
WCPROT	0.8	14War	0.09	0.0710	0.0032	0.0062	Adj Means
WCPROT	0.8	15Bb	0.17	0.0871	0.0039	0.0077	Adj Means
WCPROT	0.8	15War	0.23	0.0872	0.0039	0.0077	Adj Means
YLD	0.2	14Bb	0.26	0.1440	0.0064	0.0127	Adj Means
YLD	0.2	14War	0.23	0.1450	0.0065	0.0127	Adj Means
YLD	0.2	15Bb	0.33	0.1244	0.0056	0.0109	Adj Means
YLD	0.2	15War	0.41	0.1132	0.0051	0.0099	Adj Means
YLD	0.4	14Bb	0.24	0.0954	0.0043	0.0084	Adj Means
YLD	0.4	14War	0.20	0.0944	0.0042	0.0083	Adj Means

Trait †	Prop_Miss	Env ‡	Mean r	SD	SE	CI	Model §
YLD	0.4	15Bb	0.32	0.0856	0.0038	0.0075	Adj Means
YLD	0.4	15War	0.39	0.0833	0.0037	0.0073	Adj Means
YLD	0.6	14Bb	0.20	0.0791	0.0035	0.0070	Adj Means
YLD	0.6	14War	0.17	0.0779	0.0035	0.0068	Adj Means
YLD	0.6	15Bb	0.30	0.0716	0.0032	0.0063	Adj Means
YLD	0.6	15War	0.36	0.0761	0.0034	0.0067	Adj Means
YLD	0.8	14Bb	0.13	0.0875	0.0039	0.0077	Adj Means
YLD	0.8	14War	0.13	0.0784	0.0035	0.0069	Adj Means
YLD	0.8	15Bb	0.26	0.0923	0.0041	0.0081	Adj Means
YLD	0.8	15War	0.32	0.0988	0.0044	0.0087	Adj Means
BIOM	0.2	14Bb	0.09	0.1506	0.0067	0.0132	GEI CV1
BIOM	0.2	14Bb	0.22	0.1465	0.0066	0.0129	GEI CV2
BIOM	0.4	14Bb	0.07	0.0915	0.0041	0.0080	GEI CV1
BIOM	0.4	14Bb	0.22	0.0854	0.0038	0.0075	GEI CV2
BIOM	0.6	14Bb	0.06	0.0738	0.0033	0.0065	GEI CV1
BIOM	0.6	14Bb	0.15	0.0767	0.0034	0.0067	GEI CV2
BIOM	0.8	14Bb	0.05	0.0670	0.0030	0.0059	GEI CV1
BIOM	0.8	14Bb	0.06	0.0696	0.0031	0.0061	GEI CV2
BIOM	0.2	14War	0.00	0.1564	0.0070	0.0137	GEI CV1
BIOM	0.2	14War	0.14	0.1515	0.0068	0.0133	GEI CV2
BIOM	0.4	14War	-0.01	0.1091	0.0049	0.0096	GEI CV1
BIOM	0.4	14War	0.15	0.0969	0.0043	0.0085	GEI CV2
BIOM	0.6	14War	-0.01	0.0978	0.0044	0.0086	GEI CV1
BIOM	0.6	14War	0.06	0.1007	0.0045	0.0089	GEI CV2
BIOM	0.8	14War	-0.01	0.0940	0.0042	0.0083	GEI CV1
BIOM	0.8	14War	0.00	0.0931	0.0042	0.0082	GEI CV2
BIOM	0.2	15Bb	0.25	0.1454	0.0065	0.0128	GEI CV1
BIOM	0.2	15Bb	0.28	0.1433	0.0064	0.0126	GEI CV2
BIOM	0.4	15Bb	0.25	0.0812	0.0036	0.0071	GEI CV1
BIOM	0.4	15Bb	0.26	0.0895	0.0040	0.0079	GEI CV2
BIOM	0.6	15Bb	0.23	0.0591	0.0026	0.0052	GEI CV1
BIOM	0.6	15Bb	0.20	0.0946	0.0042	0.0083	GEI CV2
BIOM	0.8	15Bb	0.18	0.0940	0.0042	0.0083	GEI CV1
BIOM	0.8	15Bb	0.11	0.1251	0.0056	0.0110	GEI CV2
BIOM	0.2	15War	0.21	0.1392	0.0062	0.0122	GEI CV1
BIOM	0.2	15War	0.25	0.1278	0.0057	0.0112	GEI CV2
BIOM	0.4	15War	0.18	0.1022	0.0046	0.0090	GEI CV1
BIOM	0.4	15War	0.23	0.1018	0.0046	0.0089	GEI CV2
BIOM	0.6	15War	0.14	0.1232	0.0055	0.0108	GEI CV1
BIOM	0.6	15War	0.17	0.1113	0.0050	0.0098	GEI CV2
BIOM	0.8	15War	0.10	0.1561	0.0070	0.0137	GEI CV1
BIOM	0.8	15War	0.09	0.1341	0.0060	0.0118	GEI CV2
FLSG	0.2	14Bb	0.10	0.1539	0.0069	0.0135	GEI CV1

Trait †	Prop_Miss	Env ‡	Mean r	SD	SE	CI	Model §
FLSG	0.2	14Bb	0.24	0.1561	0.0070	0.0137	GEI CV2
FLSG	0.4	14Bb	0.08	0.0996	0.0045	0.0088	GEI CV1
FLSG	0.4	14Bb	0.24	0.1047	0.0047	0.0092	GEI CV2
FLSG	0.6	14Bb	0.09	0.0749	0.0033	0.0066	GEI CV1
FLSG	0.6	14Bb	0.19	0.0844	0.0038	0.0074	GEI CV2
FLSG	0.8	14Bb	0.06	0.0700	0.0031	0.0062	GEI CV1
FLSG	0.8	14Bb	0.14	0.0738	0.0033	0.0065	GEI CV2
FLSG	0.2	14War	0.53	0.1153	0.0052	0.0101	GEI CV1
FLSG	0.2	14War	0.58	0.1078	0.0048	0.0095	GEI CV2
FLSG	0.4	14War	0.51	0.0659	0.0029	0.0058	GEI CV1
FLSG	0.4	14War	0.55	0.0733	0.0033	0.0064	GEI CV2
FLSG	0.6	14War	0.49	0.0535	0.0024	0.0047	GEI CV1
FLSG	0.6	14War	0.48	0.0718	0.0032	0.0063	GEI CV2
FLSG	0.8	14War	0.43	0.0636	0.0028	0.0056	GEI CV1
FLSG	0.8	14War	0.38	0.0869	0.0039	0.0076	GEI CV2
FLSG	0.2	15Bb	0.53	0.0993	0.0044	0.0087	GEI CV1
FLSG	0.2	15Bb	0.60	0.0882	0.0039	0.0078	GEI CV2
FLSG	0.4	15Bb	0.51	0.0637	0.0028	0.0056	GEI CV1
FLSG	0.4	15Bb	0.59	0.0606	0.0027	0.0053	GEI CV2
FLSG	0.6	15Bb	0.49	0.0477	0.0021	0.0042	GEI CV1
FLSG	0.6	15Bb	0.51	0.0554	0.0025	0.0049	GEI CV2
FLSG	0.8	15Bb	0.43	0.0542	0.0024	0.0048	GEI CV1
FLSG	0.8	15Bb	0.42	0.0659	0.0029	0.0058	GEI CV2
FLSG	0.2	15War	0.48	0.1183	0.0053	0.0104	GEI CV1
FLSG	0.2	15War	0.55	0.1045	0.0047	0.0092	GEI CV2
FLSG	0.4	15War	0.48	0.0757	0.0034	0.0067	GEI CV1
FLSG	0.4	15War	0.55	0.0664	0.0030	0.0058	GEI CV2
FLSG	0.6	15War	0.45	0.0552	0.0025	0.0049	GEI CV1
FLSG	0.6	15War	0.49	0.0601	0.0027	0.0053	GEI CV2
FLSG	0.8	15War	0.40	0.0580	0.0026	0.0051	GEI CV1
FLSG	0.8	15War	0.40	0.0744	0.0033	0.0065	GEI CV2
GSQM	0.2	14Bb	0.47	0.1330	0.0059	0.0117	GEI CV1
GSQM	0.2	14Bb	0.68	0.0833	0.0037	0.0073	GEI CV2
GSQM	0.4	14Bb	0.43	0.0901	0.0040	0.0079	GEI CV1
GSQM	0.4	14Bb	0.68	0.0525	0.0023	0.0046	GEI CV2
GSQM	0.6	14Bb	0.35	0.0878	0.0039	0.0077	GEI CV1
GSQM	0.6	14Bb	0.53	0.0667	0.0030	0.0059	GEI CV2
GSQM	0.8	14Bb	0.25	0.0959	0.0043	0.0084	GEI CV1
GSQM	0.8	14Bb	0.33	0.0891	0.0040	0.0078	GEI CV2
GSQM	0.2	14War	0.29	0.1511	0.0068	0.0133	GEI CV1
GSQM	0.2	14War	0.55	0.1120	0.0050	0.0098	GEI CV2
GSQM	0.4	14War	0.26	0.0988	0.0044	0.0087	GEI CV1
GSQM	0.4	14War	0.56	0.0656	0.0029	0.0058	GEI CV2

Trait †	Prop_Miss	Env ‡	Mean r	SD	SE	CI	Model §
GSQM	0.6	14War	0.21	0.0815	0.0036	0.0072	GEI CV1
GSQM	0.6	14War	0.42	0.0696	0.0031	0.0061	GEI CV2
GSQM	0.8	14War	0.14	0.0806	0.0036	0.0071	GEI CV1
GSQM	0.8	14War	0.27	0.0758	0.0034	0.0067	GEI CV2
GSQM	0.2	15Bb	0.17	0.1625	0.0073	0.0143	GEI CV1
GSQM	0.2	15Bb	0.35	0.1339	0.0060	0.0118	GEI CV2
GSQM	0.4	15Bb	0.13	0.1034	0.0046	0.0091	GEI CV1
GSQM	0.4	15Bb	0.35	0.0862	0.0039	0.0076	GEI CV2
GSQM	0.6	15Bb	0.09	0.0819	0.0037	0.0072	GEI CV1
GSQM	0.6	15Bb	0.25	0.0778	0.0035	0.0068	GEI CV2
GSQM	0.8	15Bb	0.04	0.0734	0.0033	0.0064	GEI CV1
GSQM	0.8	15Bb	0.18	0.0682	0.0031	0.0060	GEI CV2
GSQM	0.2	15War	0.37	0.1530	0.0068	0.0134	GEI CV1
GSQM	0.2	15War	0.50	0.1279	0.0057	0.0112	GEI CV2
GSQM	0.4	15War	0.31	0.1044	0.0047	0.0092	GEI CV1
GSQM	0.4	15War	0.50	0.0761	0.0034	0.0067	GEI CV2
GSQM	0.6	15War	0.24	0.0966	0.0043	0.0085	GEI CV1
GSQM	0.6	15War	0.37	0.0785	0.0035	0.0069	GEI CV2
GSQM	0.8	15War	0.15	0.0909	0.0041	0.0080	GEI CV1
GSQM	0.8	15War	0.23	0.0839	0.0038	0.0074	GEI CV2
GW	0.2	14Bb	0.18	0.1512	0.0068	0.0133	GEI CV1
GW	0.2	14Bb	0.30	0.1363	0.0061	0.0120	GEI CV2
GW	0.4	14Bb	0.17	0.0961	0.0043	0.0084	GEI CV1
GW	0.4	14Bb	0.29	0.0855	0.0038	0.0075	GEI CV2
GW	0.6	14Bb	0.14	0.0722	0.0032	0.0063	GEI CV1
GW	0.6	14Bb	0.21	0.0788	0.0035	0.0069	GEI CV2
GW	0.8	14Bb	0.10	0.0664	0.0030	0.0058	GEI CV1
GW	0.8	14Bb	0.11	0.0740	0.0033	0.0065	GEI CV2
GW	0.2	14War	0.15	0.1487	0.0067	0.0131	GEI CV1
GW	0.2	14War	0.23	0.1419	0.0063	0.0125	GEI CV2
GW	0.4	14War	0.15	0.0850	0.0038	0.0075	GEI CV1
GW	0.4	14War	0.23	0.0866	0.0039	0.0076	GEI CV2
GW	0.6	14War	0.13	0.0662	0.0030	0.0058	GEI CV1
GW	0.6	14War	0.19	0.0705	0.0032	0.0062	GEI CV2
GW	0.8	14War	0.11	0.0636	0.0028	0.0056	GEI CV1
GW	0.8	14War	0.16	0.0529	0.0024	0.0047	GEI CV2
GW	0.2	15Bb	0.33	0.1373	0.0061	0.0121	GEI CV1
GW	0.2	15Bb	0.42	0.1262	0.0056	0.0111	GEI CV2
GW	0.4	15Bb	0.34	0.0824	0.0037	0.0072	GEI CV1
GW	0.4	15Bb	0.40	0.0790	0.0035	0.0069	GEI CV2
GW	0.6	15Bb	0.33	0.0573	0.0026	0.0050	GEI CV1
GW	0.6	15Bb	0.36	0.0636	0.0028	0.0056	GEI CV2
GW	0.8	15Bb	0.30	0.0693	0.0031	0.0061	GEI CV1

Trait †	Prop_Miss	Env ‡	Mean r	SD	SE	CI	Model §
GW	0.8	15Bb	0.34	0.0748	0.0033	0.0066	GEI CV2
GW	0.2	15War	0.30	0.1171	0.0052	0.0103	GEI CV1
GW	0.2	15War	0.37	0.1159	0.0052	0.0102	GEI CV2
GW	0.4	15War	0.29	0.0778	0.0035	0.0068	GEI CV1
GW	0.4	15War	0.36	0.0775	0.0035	0.0068	GEI CV2
GW	0.6	15War	0.28	0.0725	0.0032	0.0064	GEI CV1
GW	0.6	15War	0.29	0.0869	0.0039	0.0076	GEI CV2
GW	0.8	15War	0.26	0.0904	0.0040	0.0079	GEI CV1
GW	0.8	15War	0.24	0.1028	0.0046	0.0090	GEI CV2
HD	0.2	14Bb	0.20	0.1431	0.0064	0.0126	GEI CV1
HD	0.2	14Bb	0.55	0.1013	0.0045	0.0089	GEI CV2
HD	0.4	14Bb	0.20	0.0984	0.0044	0.0086	GEI CV1
HD	0.4	14Bb	0.58	0.0605	0.0027	0.0053	GEI CV2
HD	0.6	14Bb	0.21	0.0713	0.0032	0.0063	GEI CV1
HD	0.6	14Bb	0.42	0.0711	0.0032	0.0062	GEI CV2
HD	0.8	14Bb	0.19	0.0712	0.0032	0.0063	GEI CV1
HD	0.8	14Bb	0.28	0.0823	0.0037	0.0072	GEI CV2
HD	0.2	14War	0.42	0.1281	0.0057	0.0113	GEI CV1
HD	0.2	14War	0.58	0.0951	0.0043	0.0084	GEI CV2
HD	0.4	14War	0.39	0.0803	0.0036	0.0071	GEI CV1
HD	0.4	14War	0.56	0.0568	0.0025	0.0050	GEI CV2
HD	0.6	14War	0.35	0.0628	0.0028	0.0055	GEI CV1
HD	0.6	14War	0.42	0.0621	0.0028	0.0055	GEI CV2
HD	0.8	14War	0.28	0.0718	0.0032	0.0063	GEI CV1
HD	0.8	14War	0.35	0.0777	0.0035	0.0068	GEI CV2
HD	0.2	15Bb	0.52	0.1000	0.0045	0.0088	GEI CV1
HD	0.2	15Bb	0.75	0.0713	0.0032	0.0063	GEI CV2
HD	0.4	15Bb	0.48	0.0700	0.0031	0.0062	GEI CV1
HD	0.4	15Bb	0.76	0.0417	0.0019	0.0037	GEI CV2
HD	0.6	15Bb	0.42	0.0664	0.0030	0.0058	GEI CV1
HD	0.6	15Bb	0.62	0.0522	0.0023	0.0046	GEI CV2
HD	0.8	15Bb	0.31	0.0948	0.0042	0.0083	GEI CV1
HD	0.8	15Bb	0.44	0.0756	0.0034	0.0066	GEI CV2
HD	0.2	15War	0.57	0.0947	0.0042	0.0083	GEI CV1
HD	0.2	15War	0.83	0.0408	0.0018	0.0036	GEI CV2
HD	0.4	15War	0.53	0.0659	0.0029	0.0058	GEI CV1
HD	0.4	15War	0.81	0.0275	0.0012	0.0024	GEI CV2
HD	0.6	15War	0.46	0.0622	0.0028	0.0055	GEI CV1
HD	0.6	15War	0.66	0.0500	0.0022	0.0044	GEI CV2
HD	0.8	15War	0.34	0.0870	0.0039	0.0076	GEI CV1
HD	0.8	15War	0.47	0.0684	0.0031	0.0060	GEI CV2
HI	0.2	14Bb	0.42	0.1476	0.0066	0.0130	GEI CV1
HI	0.2	14Bb	0.70	0.1039	0.0046	0.0091	GEI CV2

Trait †	Prop_Miss	Env ‡	Mean r	SD	SE	CI	Model §
HI	0.4	14Bb	0.42	0.0918	0.0041	0.0081	GEI CV1
HI	0.4	14Bb	0.70	0.0604	0.0027	0.0053	GEI CV2
HI	0.6	14Bb	0.41	0.0622	0.0028	0.0055	GEI CV1
HI	0.6	14Bb	0.58	0.0634	0.0028	0.0056	GEI CV2
HI	0.8	14Bb	0.35	0.0847	0.0038	0.0074	GEI CV1
HI	0.8	14Bb	0.38	0.0886	0.0040	0.0078	GEI CV2
HI	0.2	14War	0.41	0.1274	0.0057	0.0112	GEI CV1
HI	0.2	14War	0.67	0.1122	0.0050	0.0099	GEI CV2
HI	0.4	14War	0.40	0.0846	0.0038	0.0074	GEI CV1
HI	0.4	14War	0.68	0.0631	0.0028	0.0055	GEI CV2
HI	0.6	14War	0.36	0.0634	0.0028	0.0056	GEI CV1
HI	0.6	14War	0.56	0.0628	0.0028	0.0055	GEI CV2
HI	0.8	14War	0.29	0.0746	0.0033	0.0066	GEI CV1
HI	0.8	14War	0.39	0.0732	0.0033	0.0064	GEI CV2
HI	0.2	15Bb	0.49	0.1049	0.0047	0.0092	GEI CV1
HI	0.2	15Bb	0.64	0.0898	0.0040	0.0079	GEI CV2
HI	0.4	15Bb	0.47	0.0734	0.0033	0.0064	GEI CV1
HI	0.4	15Bb	0.64	0.0539	0.0024	0.0047	GEI CV2
HI	0.6	15Bb	0.43	0.0670	0.0030	0.0059	GEI CV1
HI	0.6	15Bb	0.55	0.0569	0.0025	0.0050	GEI CV2
HI	0.8	15Bb	0.36	0.0915	0.0041	0.0080	GEI CV1
HI	0.8	15Bb	0.48	0.0603	0.0027	0.0053	GEI CV2
HI	0.2	15War	0.24	0.1369	0.0061	0.0120	GEI CV1
HI	0.2	15War	0.46	0.1134	0.0051	0.0100	GEI CV2
HI	0.4	15War	0.22	0.0880	0.0039	0.0077	GEI CV1
HI	0.4	15War	0.44	0.0823	0.0037	0.0072	GEI CV2
HI	0.6	15War	0.19	0.0750	0.0034	0.0066	GEI CV1
HI	0.6	15War	0.34	0.0736	0.0033	0.0065	GEI CV2
HI	0.8	15War	0.15	0.0753	0.0034	0.0066	GEI CV1
HI	0.8	15War	0.26	0.0699	0.0031	0.0061	GEI CV2
HT	0.2	14Bb	0.43	0.1228	0.0055	0.0108	GEI CV1
HT	0.2	14Bb	0.50	0.1143	0.0051	0.0100	GEI CV2
HT	0.4	14Bb	0.42	0.0765	0.0034	0.0067	GEI CV1
HT	0.4	14Bb	0.49	0.0784	0.0035	0.0069	GEI CV2
HT	0.6	14Bb	0.40	0.0619	0.0028	0.0054	GEI CV1
HT	0.6	14Bb	0.43	0.0682	0.0031	0.0060	GEI CV2
HT	0.8	14Bb	0.37	0.0569	0.0025	0.0050	GEI CV1
HT	0.8	14Bb	0.39	0.0692	0.0031	0.0061	GEI CV2
HT	0.2	14War	0.41	0.1370	0.0061	0.0120	GEI CV1
HT	0.2	14War	0.56	0.1166	0.0052	0.0102	GEI CV2
HT	0.4	14War	0.40	0.0854	0.0038	0.0075	GEI CV1
HT	0.4	14War	0.56	0.0733	0.0033	0.0064	GEI CV2
HT	0.6	14War	0.37	0.0617	0.0028	0.0054	GEI CV1

Trait †	Prop_Miss	Env ‡	Mean r	SD	SE	CI	Model §
HT	0.6	14War	0.48	0.0626	0.0028	0.0055	GEI CV2
HT	0.8	14War	0.31	0.0665	0.0030	0.0058	GEI CV1
HT	0.8	14War	0.41	0.0655	0.0029	0.0058	GEI CV2
HT	0.2	15Bb	0.52	0.1069	0.0048	0.0094	GEI CV1
HT	0.2	15Bb	0.71	0.0838	0.0037	0.0074	GEI CV2
HT	0.4	15Bb	0.50	0.0625	0.0028	0.0055	GEI CV1
HT	0.4	15Bb	0.71	0.0549	0.0025	0.0048	GEI CV2
HT	0.6	15Bb	0.48	0.0565	0.0025	0.0050	GEI CV1
HT	0.6	15Bb	0.60	0.0571	0.0026	0.0050	GEI CV2
HT	0.8	15Bb	0.43	0.0742	0.0033	0.0065	GEI CV1
HT	0.8	15Bb	0.50	0.0665	0.0030	0.0058	GEI CV2
HT	0.2	15War	0.42	0.1290	0.0058	0.0113	GEI CV1
HT	0.2	15War	0.60	0.0993	0.0044	0.0087	GEI CV2
HT	0.4	15War	0.40	0.0792	0.0035	0.0070	GEI CV1
HT	0.4	15War	0.61	0.0607	0.0027	0.0053	GEI CV2
HT	0.6	15War	0.37	0.0588	0.0026	0.0052	GEI CV1
HT	0.6	15War	0.51	0.0543	0.0024	0.0048	GEI CV2
HT	0.8	15War	0.32	0.0632	0.0028	0.0056	GEI CV1
HT	0.8	15War	0.43	0.0549	0.0025	0.0048	GEI CV2
MAT	0.2	14Bb	0.13	0.1466	0.0066	0.0129	GEI CV1
MAT	0.2	14Bb	0.32	0.1313	0.0059	0.0115	GEI CV2
MAT	0.4	14Bb	0.12	0.0896	0.0040	0.0079	GEI CV1
MAT	0.4	14Bb	0.36	0.0763	0.0034	0.0067	GEI CV2
MAT	0.6	14Bb	0.12	0.0697	0.0031	0.0061	GEI CV1
MAT	0.6	14Bb	0.25	0.0645	0.0029	0.0057	GEI CV2
MAT	0.8	14Bb	0.11	0.0668	0.0030	0.0059	GEI CV1
MAT	0.8	14Bb	0.17	0.0770	0.0034	0.0068	GEI CV2
MAT	0.2	14War	0.14	0.1451	0.0065	0.0127	GEI CV1
MAT	0.2	14War	0.33	0.1228	0.0055	0.0108	GEI CV2
MAT	0.4	14War	0.13	0.0913	0.0041	0.0080	GEI CV1
MAT	0.4	14War	0.33	0.0867	0.0039	0.0076	GEI CV2
MAT	0.6	14War	0.12	0.0711	0.0032	0.0062	GEI CV1
MAT	0.6	14War	0.23	0.0825	0.0037	0.0072	GEI CV2
MAT	0.8	14War	0.11	0.0664	0.0030	0.0058	GEI CV1
MAT	0.8	14War	0.20	0.0806	0.0036	0.0071	GEI CV2
MAT	0.2	15Bb	0.39	0.1264	0.0057	0.0111	GEI CV1
MAT	0.2	15Bb	0.52	0.1093	0.0049	0.0096	GEI CV2
MAT	0.4	15Bb	0.36	0.0762	0.0034	0.0067	GEI CV1
MAT	0.4	15Bb	0.53	0.0750	0.0034	0.0066	GEI CV2
MAT	0.6	15Bb	0.34	0.0744	0.0033	0.0065	GEI CV1
MAT	0.6	15Bb	0.44	0.0795	0.0036	0.0070	GEI CV2
MAT	0.8	15Bb	0.28	0.1028	0.0046	0.0090	GEI CV1
MAT	0.8	15Bb	0.31	0.1177	0.0053	0.0103	GEI CV2

Trait †	Prop_Miss	Env ‡	Mean r	SD	SE	CI	Model §
MAT	0.2	15War	0.36	0.1414	0.0063	0.0124	GEI CV1
MAT	0.2	15War	0.54	0.1094	0.0049	0.0096	GEI CV2
MAT	0.4	15War	0.33	0.0849	0.0038	0.0075	GEI CV1
MAT	0.4	15War	0.56	0.0700	0.0031	0.0061	GEI CV2
MAT	0.6	15War	0.29	0.0743	0.0033	0.0065	GEI CV1
MAT	0.6	15War	0.43	0.0739	0.0033	0.0065	GEI CV2
MAT	0.8	15War	0.21	0.0817	0.0037	0.0072	GEI CV1
MAT	0.8	15War	0.26	0.0906	0.0041	0.0080	GEI CV2
NDVI	0.2	14Bb	0.25	0.1388	0.0062	0.0122	GEI CV1
NDVI	0.2	14Bb	0.31	0.1243	0.0056	0.0109	GEI CV2
NDVI	0.4	14Bb	0.23	0.0866	0.0039	0.0076	GEI CV1
NDVI	0.4	14Bb	0.27	0.0870	0.0039	0.0076	GEI CV2
NDVI	0.6	14Bb	0.21	0.0642	0.0029	0.0056	GEI CV1
NDVI	0.6	14Bb	0.20	0.0875	0.0039	0.0077	GEI CV2
NDVI	0.8	14Bb	0.17	0.0690	0.0031	0.0061	GEI CV1
NDVI	0.8	14Bb	0.15	0.0801	0.0036	0.0070	GEI CV2
NDVI	0.2	14War	0.38	0.1408	0.0063	0.0124	GEI CV1
NDVI	0.2	14War	0.41	0.1265	0.0057	0.0111	GEI CV2
NDVI	0.4	14War	0.34	0.0959	0.0043	0.0084	GEI CV1
NDVI	0.4	14War	0.38	0.0863	0.0039	0.0076	GEI CV2
NDVI	0.6	14War	0.29	0.0796	0.0036	0.0070	GEI CV1
NDVI	0.6	14War	0.33	0.0767	0.0034	0.0067	GEI CV2
NDVI	0.8	14War	0.23	0.0867	0.0039	0.0076	GEI CV1
NDVI	0.8	14War	0.27	0.0667	0.0030	0.0059	GEI CV2
NDVI	0.2	15Bb	0.11	0.1551	0.0069	0.0136	GEI CV1
NDVI	0.2	15Bb	0.17	0.1615	0.0072	0.0142	GEI CV2
NDVI	0.4	15Bb	0.14	0.0940	0.0042	0.0083	GEI CV1
NDVI	0.4	15Bb	0.14	0.1096	0.0049	0.0096	GEI CV2
NDVI	0.6	15Bb	0.14	0.0799	0.0036	0.0070	GEI CV1
NDVI	0.6	15Bb	0.09	0.1022	0.0046	0.0090	GEI CV2
NDVI	0.8	15Bb	0.13	0.0885	0.0040	0.0078	GEI CV1
NDVI	0.8	15Bb	0.00	0.1104	0.0049	0.0097	GEI CV2
NDVI	0.2	15War	0.23	0.1453	0.0065	0.0128	GEI CV1
NDVI	0.2	15War	0.27	0.1376	0.0062	0.0121	GEI CV2
NDVI	0.4	15War	0.20	0.0936	0.0042	0.0082	GEI CV1
NDVI	0.4	15War	0.25	0.0920	0.0041	0.0081	GEI CV2
NDVI	0.6	15War	0.18	0.0736	0.0033	0.0065	GEI CV1
NDVI	0.6	15War	0.18	0.0855	0.0038	0.0075	GEI CV2
NDVI	0.8	15War	0.14	0.0812	0.0036	0.0071	GEI CV1
NDVI	0.8	15War	0.14	0.0896	0.0040	0.0079	GEI CV2
SPH	0.2	14Bb	0.45	0.1330	0.0059	0.0117	GEI CV1
SPH	0.2	14Bb	0.81	0.0796	0.0036	0.0070	GEI CV2
SPH	0.4	14Bb	0.39	0.0907	0.0041	0.0080	GEI CV1

Trait †	Prop_Miss	Env ‡	Mean r	SD	SE	CI	Model §
SPH	0.4	14Bb	0.80	0.0472	0.0021	0.0041	GEI CV2
SPH	0.6	14Bb	0.31	0.0813	0.0036	0.0071	GEI CV1
SPH	0.6	14Bb	0.62	0.0621	0.0028	0.0055	GEI CV2
SPH	0.8	14Bb	0.20	0.0850	0.0038	0.0075	GEI CV1
SPH	0.8	14Bb	0.43	0.0650	0.0029	0.0057	GEI CV2
SPH	0.2	14War	0.35	0.1342	0.0060	0.0118	GEI CV1
SPH	0.2	14War	0.77	0.0845	0.0038	0.0074	GEI CV2
SPH	0.4	14War	0.29	0.0979	0.0044	0.0086	GEI CV1
SPH	0.4	14War	0.75	0.0539	0.0024	0.0047	GEI CV2
SPH	0.6	14War	0.22	0.0866	0.0039	0.0076	GEI CV1
SPH	0.6	14War	0.56	0.0803	0.0036	0.0071	GEI CV2
SPH	0.8	14War	0.13	0.0760	0.0034	0.0067	GEI CV1
SPH	0.8	14War	0.39	0.0647	0.0029	0.0057	GEI CV2
SPH	0.2	15Bb	0.35	0.1288	0.0058	0.0113	GEI CV1
SPH	0.2	15Bb	0.67	0.0743	0.0033	0.0065	GEI CV2
SPH	0.4	15Bb	0.32	0.0838	0.0037	0.0074	GEI CV1
SPH	0.4	15Bb	0.66	0.0490	0.0022	0.0043	GEI CV2
SPH	0.6	15Bb	0.27	0.0721	0.0032	0.0063	GEI CV1
SPH	0.6	15Bb	0.52	0.0545	0.0024	0.0048	GEI CV2
SPH	0.8	15Bb	0.20	0.0815	0.0036	0.0072	GEI CV1
SPH	0.8	15Bb	0.39	0.0557	0.0025	0.0049	GEI CV2
SPH	0.2	15War	0.39	0.1491	0.0067	0.0131	GEI CV1
SPH	0.2	15War	0.71	0.0726	0.0032	0.0064	GEI CV2
SPH	0.4	15War	0.32	0.0920	0.0041	0.0081	GEI CV1
SPH	0.4	15War	0.68	0.0521	0.0023	0.0046	GEI CV2
SPH	0.6	15War	0.25	0.0816	0.0036	0.0072	GEI CV1
SPH	0.6	15War	0.53	0.0621	0.0028	0.0055	GEI CV2
SPH	0.8	15War	0.16	0.0855	0.0038	0.0075	GEI CV1
SPH	0.8	15War	0.36	0.0664	0.0030	0.0058	GEI CV2
SSQM	0.2	14Bb	0.38	0.1317	0.0059	0.0116	GEI CV1
SSQM	0.2	14Bb	0.58	0.1075	0.0048	0.0094	GEI CV2
SSQM	0.4	14Bb	0.34	0.0864	0.0039	0.0076	GEI CV1
SSQM	0.4	14Bb	0.59	0.0663	0.0030	0.0058	GEI CV2
SSQM	0.6	14Bb	0.28	0.0720	0.0032	0.0063	GEI CV1
SSQM	0.6	14Bb	0.45	0.0727	0.0032	0.0064	GEI CV2
SSQM	0.8	14Bb	0.20	0.0681	0.0030	0.0060	GEI CV1
SSQM	0.8	14Bb	0.34	0.0640	0.0029	0.0056	GEI CV2
SSQM	0.2	14War	0.40	0.1293	0.0058	0.0114	GEI CV1
SSQM	0.2	14War	0.57	0.1055	0.0047	0.0093	GEI CV2
SSQM	0.4	14War	0.36	0.0878	0.0039	0.0077	GEI CV1
SSQM	0.4	14War	0.59	0.0629	0.0028	0.0055	GEI CV2
SSQM	0.6	14War	0.31	0.0676	0.0030	0.0059	GEI CV1
SSQM	0.6	14War	0.45	0.0724	0.0032	0.0064	GEI CV2

Trait †	Prop_Miss	Env ‡	Mean r	SD	SE	CI	Model §
SSQM	0.8	14War	0.24	0.0661	0.0030	0.0058	GEI CV1
SSQM	0.8	14War	0.34	0.0678	0.0030	0.0060	GEI CV2
SSQM	0.2	15Bb	0.33	0.1244	0.0056	0.0109	GEI CV1
SSQM	0.2	15Bb	0.47	0.1202	0.0054	0.0106	GEI CV2
SSQM	0.4	15Bb	0.30	0.0846	0.0038	0.0074	GEI CV1
SSQM	0.4	15Bb	0.49	0.0724	0.0032	0.0064	GEI CV2
SSQM	0.6	15Bb	0.25	0.0720	0.0032	0.0063	GEI CV1
SSQM	0.6	15Bb	0.41	0.0643	0.0029	0.0057	GEI CV2
SSQM	0.8	15Bb	0.18	0.0626	0.0028	0.0055	GEI CV1
SSQM	0.8	15Bb	0.30	0.0576	0.0026	0.0051	GEI CV2
SSQM	0.2	15War	0.42	0.1276	0.0057	0.0112	GEI CV1
SSQM	0.2	15War	0.56	0.0977	0.0044	0.0086	GEI CV2
SSQM	0.4	15War	0.37	0.0851	0.0038	0.0075	GEI CV1
SSQM	0.4	15War	0.57	0.0605	0.0027	0.0053	GEI CV2
SSQM	0.6	15War	0.31	0.0750	0.0034	0.0066	GEI CV1
SSQM	0.6	15War	0.46	0.0649	0.0029	0.0057	GEI CV2
SSQM	0.8	15War	0.24	0.0734	0.0033	0.0064	GEI CV1
SSQM	0.8	15War	0.31	0.0668	0.0030	0.0059	GEI CV2
STARCH	0.2	14Bb	0.42	0.1421	0.0064	0.0125	GEI CV1
STARCH	0.2	14Bb	0.59	0.1227	0.0055	0.0108	GEI CV2
STARCH	0.4	14Bb	0.38	0.0898	0.0040	0.0079	GEI CV1
STARCH	0.4	14Bb	0.60	0.0683	0.0031	0.0060	GEI CV2
STARCH	0.6	14Bb	0.33	0.0805	0.0036	0.0071	GEI CV1
STARCH	0.6	14Bb	0.46	0.0776	0.0035	0.0068	GEI CV2
STARCH	0.8	14Bb	0.24	0.1161	0.0052	0.0102	GEI CV1
STARCH	0.8	14Bb	0.23	0.1170	0.0052	0.0103	GEI CV2
STARCH	0.2	14War	0.42	0.1128	0.0050	0.0099	GEI CV1
STARCH	0.2	14War	0.63	0.0980	0.0044	0.0086	GEI CV2
STARCH	0.4	14War	0.40	0.0730	0.0033	0.0064	GEI CV1
STARCH	0.4	14War	0.62	0.0618	0.0028	0.0054	GEI CV2
STARCH	0.6	14War	0.35	0.0781	0.0035	0.0069	GEI CV1
STARCH	0.6	14War	0.49	0.0612	0.0027	0.0054	GEI CV2
STARCH	0.8	14War	0.26	0.0892	0.0040	0.0078	GEI CV1
STARCH	0.8	14War	0.29	0.0971	0.0043	0.0085	GEI CV2
STARCH	0.2	15Bb	0.15	0.1538	0.0069	0.0135	GEI CV1
STARCH	0.2	15Bb	0.27	0.1534	0.0069	0.0135	GEI CV2
STARCH	0.4	15Bb	0.15	0.1098	0.0049	0.0096	GEI CV1
STARCH	0.4	15Bb	0.26	0.1169	0.0052	0.0103	GEI CV2
STARCH	0.6	15Bb	0.13	0.1130	0.0051	0.0099	GEI CV1
STARCH	0.6	15Bb	0.25	0.1147	0.0051	0.0101	GEI CV2
STARCH	0.8	15Bb	0.11	0.1454	0.0065	0.0128	GEI CV1
STARCH	0.8	15Bb	0.27	0.1320	0.0059	0.0116	GEI CV2
STARCH	0.2	15War	0.37	0.1183	0.0053	0.0104	GEI CV1

Trait †	Prop_Miss	Env ‡	Mean r	SD	SE	CI	Model §
STARCH	0.2	15War	0.35	0.1398	0.0063	0.0123	GEI CV2
STARCH	0.4	15War	0.34	0.0949	0.0042	0.0083	GEI CV1
STARCH	0.4	15War	0.32	0.0948	0.0042	0.0083	GEI CV2
STARCH	0.6	15War	0.30	0.0897	0.0040	0.0079	GEI CV1
STARCH	0.6	15War	0.27	0.0973	0.0044	0.0086	GEI CV2
STARCH	0.8	15War	0.24	0.1067	0.0048	0.0094	GEI CV1
STARCH	0.8	15War	0.23	0.1079	0.0048	0.0095	GEI CV2
TKW	0.2	14Bb	0.63	0.1144	0.0051	0.0101	GEI CV1
TKW	0.2	14Bb	0.89	0.0408	0.0018	0.0036	GEI CV2
TKW	0.4	14Bb	0.60	0.0788	0.0035	0.0069	GEI CV1
TKW	0.4	14Bb	0.89	0.0242	0.0011	0.0021	GEI CV2
TKW	0.6	14Bb	0.54	0.0669	0.0030	0.0059	GEI CV1
TKW	0.6	14Bb	0.75	0.0462	0.0021	0.0041	GEI CV2
TKW	0.8	14Bb	0.42	0.0749	0.0033	0.0066	GEI CV1
TKW	0.8	14Bb	0.58	0.0592	0.0026	0.0052	GEI CV2
TKW	0.2	14War	0.68	0.1004	0.0045	0.0088	GEI CV1
TKW	0.2	14War	0.93	0.0244	0.0011	0.0021	GEI CV2
TKW	0.4	14War	0.65	0.0732	0.0033	0.0064	GEI CV1
TKW	0.4	14War	0.94	0.0140	0.0006	0.0012	GEI CV2
TKW	0.6	14War	0.59	0.0644	0.0029	0.0057	GEI CV1
TKW	0.6	14War	0.80	0.0378	0.0017	0.0033	GEI CV2
TKW	0.8	14War	0.48	0.0752	0.0034	0.0066	GEI CV1
TKW	0.8	14War	0.62	0.0616	0.0028	0.0054	GEI CV2
TKW	0.2	15Bb	0.61	0.1067	0.0048	0.0094	GEI CV1
TKW	0.2	15Bb	0.78	0.1066	0.0048	0.0094	GEI CV2
TKW	0.4	15Bb	0.56	0.0737	0.0033	0.0065	GEI CV1
TKW	0.4	15Bb	0.76	0.0848	0.0038	0.0075	GEI CV2
TKW	0.6	15Bb	0.53	0.0566	0.0025	0.0050	GEI CV1
TKW	0.6	15Bb	0.64	0.0582	0.0026	0.0051	GEI CV2
TKW	0.8	15Bb	0.46	0.0580	0.0026	0.0051	GEI CV1
TKW	0.8	15Bb	0.50	0.0568	0.0025	0.0050	GEI CV2
TKW	0.2	15War	0.66	0.1029	0.0046	0.0090	GEI CV1
TKW	0.2	15War	0.80	0.0876	0.0039	0.0077	GEI CV2
TKW	0.4	15War	0.63	0.0666	0.0030	0.0059	GEI CV1
TKW	0.4	15War	0.80	0.0603	0.0027	0.0053	GEI CV2
TKW	0.6	15War	0.58	0.0554	0.0025	0.0049	GEI CV1
TKW	0.6	15War	0.70	0.0492	0.0022	0.0043	GEI CV2
TKW	0.8	15War	0.49	0.0598	0.0027	0.0053	GEI CV1
TKW	0.8	15War	0.58	0.0510	0.0023	0.0045	GEI CV2
TWT	0.2	14Bb	0.55	0.1309	0.0059	0.0115	GEI CV1
TWT	0.2	14Bb	0.76	0.0731	0.0033	0.0064	GEI CV2
TWT	0.4	14Bb	0.52	0.0905	0.0040	0.0080	GEI CV1
TWT	0.4	14Bb	0.75	0.0456	0.0020	0.0040	GEI CV2

Trait †	Prop_Miss	Env ‡	Mean r	SD	SE	CI	Model §
TWT	0.6	14Bb	0.49	0.0703	0.0031	0.0062	GEI CV1
TWT	0.6	14Bb	0.63	0.0520	0.0023	0.0046	GEI CV2
TWT	0.8	14Bb	0.42	0.0774	0.0035	0.0068	GEI CV1
TWT	0.8	14Bb	0.54	0.0630	0.0028	0.0055	GEI CV2
TWT	0.2	14War	0.55	0.1179	0.0053	0.0104	GEI CV1
TWT	0.2	14War	0.78	0.0631	0.0028	0.0055	GEI CV2
TWT	0.4	14War	0.52	0.0785	0.0035	0.0069	GEI CV1
TWT	0.4	14War	0.74	0.0500	0.0022	0.0044	GEI CV2
TWT	0.6	14War	0.46	0.0608	0.0027	0.0053	GEI CV1
TWT	0.6	14War	0.59	0.0573	0.0026	0.0050	GEI CV2
TWT	0.8	14War	0.37	0.0711	0.0032	0.0062	GEI CV1
TWT	0.8	14War	0.48	0.0599	0.0027	0.0053	GEI CV2
TWT	0.2	15Bb	0.51	0.1213	0.0054	0.0107	GEI CV1
TWT	0.2	15Bb	0.73	0.0693	0.0031	0.0061	GEI CV2
TWT	0.4	15Bb	0.50	0.0742	0.0033	0.0065	GEI CV1
TWT	0.4	15Bb	0.74	0.0431	0.0019	0.0038	GEI CV2
TWT	0.6	15Bb	0.48	0.0582	0.0026	0.0051	GEI CV1
TWT	0.6	15Bb	0.63	0.0472	0.0021	0.0041	GEI CV2
TWT	0.8	15Bb	0.42	0.0689	0.0031	0.0061	GEI CV1
TWT	0.8	15Bb	0.51	0.0653	0.0029	0.0057	GEI CV2
TWT	0.2	15War	0.56	0.0951	0.0043	0.0084	GEI CV1
TWT	0.2	15War	0.77	0.0621	0.0028	0.0055	GEI CV2
TWT	0.4	15War	0.52	0.0669	0.0030	0.0059	GEI CV1
TWT	0.4	15War	0.75	0.0407	0.0018	0.0036	GEI CV2
TWT	0.6	15War	0.47	0.0603	0.0027	0.0053	GEI CV1
TWT	0.6	15War	0.61	0.0609	0.0027	0.0054	GEI CV2
TWT	0.8	15War	0.39	0.0642	0.0029	0.0056	GEI CV1
TWT	0.8	15War	0.44	0.0818	0.0037	0.0072	GEI CV2
WCPROT	0.2	14Bb	-0.01	0.1512	0.0068	0.0133	GEI CV1
WCPROT	0.2	14Bb	0.24	0.1427	0.0064	0.0125	GEI CV2
WCPROT	0.4	14Bb	-0.04	0.0978	0.0044	0.0086	GEI CV1
WCPROT	0.4	14Bb	0.23	0.0936	0.0042	0.0082	GEI CV2
WCPROT	0.6	14Bb	-0.05	0.0856	0.0038	0.0075	GEI CV1
WCPROT	0.6	14Bb	0.12	0.0906	0.0041	0.0080	GEI CV2
WCPROT	0.8	14Bb	-0.08	0.0919	0.0041	0.0081	GEI CV1
WCPROT	0.8	14Bb	0.02	0.1002	0.0045	0.0088	GEI CV2
WCPROT	0.2	14War	0.00	0.1485	0.0066	0.0130	GEI CV1
WCPROT	0.2	14War	0.30	0.1456	0.0065	0.0128	GEI CV2
WCPROT	0.4	14War	-0.01	0.0942	0.0042	0.0083	GEI CV1
WCPROT	0.4	14War	0.30	0.0814	0.0036	0.0071	GEI CV2
WCPROT	0.6	14War	-0.01	0.0740	0.0033	0.0065	GEI CV1
WCPROT	0.6	14War	0.16	0.0808	0.0036	0.0071	GEI CV2
WCPROT	0.8	14War	-0.03	0.0716	0.0032	0.0063	GEI CV1

Trait †	Prop_Miss	Env ‡	Mean r	SD	SE	CI	Model §
WCPROT	0.8	14War	0.02	0.0686	0.0031	0.0060	GEI CV2
WCPROT	0.2	15Bb	0.38	0.1339	0.0060	0.0118	GEI CV1
WCPROT	0.2	15Bb	0.49	0.1136	0.0051	0.0100	GEI CV2
WCPROT	0.4	15Bb	0.33	0.0977	0.0044	0.0086	GEI CV1
WCPROT	0.4	15Bb	0.48	0.0690	0.0031	0.0061	GEI CV2
WCPROT	0.6	15Bb	0.27	0.0883	0.0040	0.0078	GEI CV1
WCPROT	0.6	15Bb	0.39	0.0702	0.0031	0.0062	GEI CV2
WCPROT	0.8	15Bb	0.17	0.0916	0.0041	0.0081	GEI CV1
WCPROT	0.8	15Bb	0.26	0.0978	0.0044	0.0086	GEI CV2
WCPROT	0.2	15War	0.23	0.1315	0.0059	0.0116	GEI CV1
WCPROT	0.2	15War	0.32	0.1260	0.0056	0.0111	GEI CV2
WCPROT	0.4	15War	0.21	0.0993	0.0044	0.0087	GEI CV1
WCPROT	0.4	15War	0.31	0.0777	0.0035	0.0068	GEI CV2
WCPROT	0.6	15War	0.15	0.0949	0.0042	0.0083	GEI CV1
WCPROT	0.6	15War	0.24	0.0810	0.0036	0.0071	GEI CV2
WCPROT	0.8	15War	0.08	0.1243	0.0056	0.0109	GEI CV1
WCPROT	0.8	15War	0.16	0.1159	0.0052	0.0102	GEI CV2
YLD	0.2	14Bb	-0.01	0.1542	0.0069	0.0135	GEI CV1
YLD	0.2	14Bb	0.19	0.1680	0.0075	0.0148	GEI CV2
YLD	0.4	14Bb	-0.01	0.1053	0.0047	0.0093	GEI CV1
YLD	0.4	14Bb	0.19	0.1077	0.0048	0.0095	GEI CV2
YLD	0.6	14Bb	0.00	0.0915	0.0041	0.0080	GEI CV1
YLD	0.6	14Bb	0.12	0.1015	0.0045	0.0089	GEI CV2
YLD	0.8	14Bb	-0.01	0.0797	0.0036	0.0070	GEI CV1
YLD	0.8	14Bb	0.03	0.0798	0.0036	0.0070	GEI CV2
YLD	0.2	14War	0.33	0.1341	0.0060	0.0118	GEI CV1
YLD	0.2	14War	0.33	0.1446	0.0065	0.0127	GEI CV2
YLD	0.4	14War	0.32	0.0773	0.0035	0.0068	GEI CV1
YLD	0.4	14War	0.34	0.0948	0.0042	0.0083	GEI CV2
YLD	0.6	14War	0.29	0.0634	0.0028	0.0056	GEI CV1
YLD	0.6	14War	0.28	0.0809	0.0036	0.0071	GEI CV2
YLD	0.8	14War	0.22	0.0706	0.0032	0.0062	GEI CV1
YLD	0.8	14War	0.23	0.0733	0.0033	0.0064	GEI CV2
YLD	0.2	15Bb	0.32	0.1382	0.0062	0.0121	GEI CV1
YLD	0.2	15Bb	0.44	0.1318	0.0059	0.0116	GEI CV2
YLD	0.4	15Bb	0.31	0.0806	0.0036	0.0071	GEI CV1
YLD	0.4	15Bb	0.45	0.0808	0.0036	0.0071	GEI CV2
YLD	0.6	15Bb	0.31	0.0617	0.0028	0.0054	GEI CV1
YLD	0.6	15Bb	0.34	0.0779	0.0035	0.0068	GEI CV2
YLD	0.8	15Bb	0.31	0.0629	0.0028	0.0055	GEI CV1
YLD	0.8	15Bb	0.29	0.0893	0.0040	0.0078	GEI CV2
YLD	0.2	15War	0.39	0.1271	0.0057	0.0112	GEI CV1
YLD	0.2	15War	0.49	0.1086	0.0049	0.0095	GEI CV2

Trait †	Prop_Miss	Env ‡	Mean r	SD	SE	CI	Model §
YLD	0.4	15War	0.38	0.0769	0.0034	0.0068	GEI CV1
YLD	0.4	15War	0.49	0.0757	0.0034	0.0067	GEI CV2
YLD	0.6	15War	0.36	0.0692	0.0031	0.0061	GEI CV1
YLD	0.6	15War	0.42	0.0765	0.0034	0.0067	GEI CV2
YLD	0.8	15War	0.34	0.0994	0.0044	0.0087	GEI CV1
YLD	0.8	15War	0.34	0.1037	0.0046	0.0091	GEI CV2

† BIOM above-ground biomass; FLS flag leaf senescence; FLSG flag leaf stay green; GSQM grains per square meter; GW grain weight; HD heading date; HI harvest index; HP maturity date minus heading date; HT plant height; MAT physiological maturity date; NDVI normalized-difference vegetation index at Zadok's GS25; SPH seeds per head; SSQM spikes per square meter; STARCH whole-grain starch content; TKW thousand kernel weight; TWT test weight; WCPROT wet chemistry-validated whole-grain protein content; YLD grain yield

‡ 14Bb Blacksburg, VA 2014; 14War Warsaw, VA 2014; 15Bb Blacksburg, VA 2015; 15War Warsaw, VA 2015

§ Adj Means univariate model using adjusted across-environment means as the phenotypic response variable; GEI CV1 multi-environment model using genotype/environment combinations as the phenotypic response variable with the CV1 cross-validation scheme; GEI CV2 multi-environment model using genotype/environment combinations as the phenotypic response variable with the CV2 cross-validation scheme

APPENDIX F: List of genotypes included in the elite panel (EP) of chapter III

The table below lists all genotypes included in the elite panel (EP) tested in chapter III, with their state of origin and pedigree (if known).

Table F.1: List of genotypes included in the elite panel (EP) of chapter III

GENOTYPE	ORIGIN	PEDIGREE
011007A1-14-16-50	Purdue	981477A1-10-2-1/981312A1-6-2-1//INW0316
0175A1-37-4-1	Purdue	981419/97397
02444A1-23-1-3	Purdue	981129A1-45-3/99793RE2-3//INW0301/92145E8-7-7-3-57
03207A1-7-3-1	Purdue	97395C1-1-4/RSI5//97395B1-4-3-8-1/3/981281A1-4-3-7/4/INW0315/99794RA4-14-1
03549A1-18-25	Purdue	981358/97462/3/92145/981004X48//INW0301
03633A1-69-2-5	Purdue	992059/INW0316//981358/97462
04606RA1-1-7-1	Purdue	Truman/INW0316
04606RA1-1-7-1-6	Purdue	TRUMAN/961341A3-1-4-6(INW0316)
04620A1-1-7-4	Purdue	TRUMAN/9017C1//92823A1/9218B4/3/P107/4/PATT/5/INW9811/GLD//96204A1
04702A1-18	Purdue	INW0316*2/INW0301
04719A1-16-1-1-7	Purdue	99840C4-8-3-6-1/5/INW0315/3/INW0301/MADSEN//INW0315/97395B1-8-4/6/99840C4-8-3-6-1
0513A1-1-3	Purdue	04607/04697
05219A1-8-21-2-4	Purdue	unknown
05222A1-1-2-1	Purdue	99840C4/5/99593RA1/6/97395C1/RSI5//INW0304/3/981281A1/4/981517A1/7/INW0316/8/99794RA4-14-1-5
05247A1-7-3-120	Purdue	98840*2/03726//99794
05247A1-7-3-27	Purdue	98840*2/03726//99794
05247A1-7-7-3-1	Purdue	99840C4/5/INW0315/3/INW0301MADSEN//INW0315/4/97395B1/6/99840C4//99794RA1
05251A1-1-136-9-5	Purdue	INW0412*2/03705//981312
05264A1-1-3-2	Purdue	INW0304*2/03727/5/96169/3/Tadinia/BH1146//Geneva/4/INW0316
05287A1-1-13	Purdue	99840*2/03726//INW0412*2/03705
0537A1-12	Purdue	97397/3/2754//INW0412/98134W
0537A1-3-12	Purdue	INW0411/2754//INW0412/98134
0566A1-3-1-65	Purdue	INW0412/992060
0566A1-3-1-67	Purdue	INW0412/992060
0570A1-2-39-5	Purdue	INW0412/3/9017/92823//201R/97462
06403A1-4	Purdue	INW0411/992059/3/96169//981129/981312
0722A1-1-7	Purdue	INW0731/3/981129/99793//INW0301/92145
07287RA1-14	Purdue	INW0304/INW0316//97462/3/Truman
07290A1-12	Purdue	992060G1-1-5-71/5/92829A1-1-5-1-4-2/A94-1048/4/GOLDFIELD/X117/3/ROANE//92145A2-4-6/6/TRUMAN

GENOTYPE	ORIGIN	PEDIGREE
0762A1-2-8	Purdue	981129A1-45-3/99793RE2-3//INW0301/92145E8-7-7-3-57/3/981477A1/981312A1//INW0316
91193D1	Purdue	Benhur//Arthur/Knox62/3/Arthur/NY5715AB/4/Hart/Beau//Arthur/Abe/5/Auburn/Coker 8427/3/OH256/Scotty//jClark
92201D5	Purdue	Tyler//Caldwell*2/Pioneer S76/3/Clark/CI5549/4/Roazon/Caldwell/5/Glory
9346A1-2-5-5-2-1	Purdue	83146C1-9-6-3-75//831800A1-7-2-5-2/861A1-8-X-38
ALLEGIANCE	Kentucky	Pioneer2548/FFR555
BECKER	Ohio	Hart/VA-66-54-10
BESS	Virginia	MO11769/Madison
BROMFIELD	Ohio	FOSTER/HOPEWELL//OH581/OH569
CALEDONIA	Cornell	Off type selection out of Geneva
CATOCTIN	Maryland	EarlyHolley/Pb67137B12-3/2/VA70-52-22
CAYUGA	Cornell	Geneva/Clark'sCream//Geneva
CHOPTANK	Maryland	Coker9803/Freedom
CLARK	Purdue	Beau/2/Pd65256A1-8-1/Pd67137B5-16/4/Sullivan/3/Beau/2/Pd5517B8-5-3-3/Logan
CRYSTAL	Michigan	Pioneer 2737W/MSU Line D1148
D6234	Michigan	MSUX1291/MSUC5107
D8006	Michigan	PIONEER2555/LOWELL
E2041	Michigan	PioneerBrand2552/PioneerBrand2737W
E5011	Michigan	PioneerBrand2555/LOWELL
E5024	Michigan	MSUD6234/P25W33
E6012	Michigan	Caledonia/PioneerBrand25W33
ERNIE	Missouri	Pike/MO9965(Stoddard/Blueboy//Stoddard/D1707)
FOSTER	Kentucky	Coker6520/Tyler
HOPKINS	Cornell	NY87048W-7387(84074(Ho/SuMei)/Caledonia//Caledonia-2///Caledonia9BC2S1
IL00-8109	Illinois	P81-1615-50/Foster//IL93-2489
IL00-8530	Illinois	IL89-1687/IL90-6364//IL93-2489
IL00-8633	Illinois	IL89-1687/IL90-6364//IL93-2489
IL01-11934	Illinois	IL90-6364/IL94-1909
IL02-18228	Illinois	Pio25R26/IL9634-24437//IL95-4162
IL02-19483B	Illinois	Patton/Cardinal//IL96-2550
IL04-24668	Illinois	IL98-13404/IL97-3578
IL04-9942	Illinois	IL95-2516/IL97-3578
IL05-4236	Illinois	Truman/KY93C-1238-17-1
IL06-13072	Illinois	IL00-8109/IL97-3632
IL06-13708	Illinois	IL00-8530/IL97-3632
IL06-13721	Illinois	IL00-8530/IL97-3632
IL06-14262	Illinois	IL00-8530/IL97-1828
IL06-14303	Illinois	IL00-8530/IL97-1828
IL06-14325	Illinois	IL00-8530/IL97-1828
IL06-23571	Illinois	IL96-6472/Pio25W33//IL94-1653
IL06-31053	Illinois	IL94-1909/IL97-3578//P93149A1-5-4-2-6

GENOTYPE	ORIGIN	PEDIGREE
IL06-7550	Illinois	IL97-3632/IL98-4632
IL06-7653	Illinois	IL97-3632/IL99-8879
IL07-12948	Illinois	IL00-8530/IL99-27048
IL07-16075	Illinois	IL01-13830/IL99-27048
IL07-19334	Illinois	IL01-36115/IL79-008T-B-B
IL07-20728	Illinois	McCormick/IL97-1828//IL00-8061
IL07-20743	Illinois	McCormick/IL97-1828//IL00-8061
IL07-21847	Illinois	IL99-2536/IL97-3632//IL00-8061
IL07-23420	Illinois	IL99-15867/IL96-6472//IL00-8530
IL07-24841	Illinois	IL00-8530/IL94-1653//IL01-5642
IL07-4415	Illinois	P96169RE2-3-6-4/IL01-34159
IL07-6861	Illinois	IL97-1828/BW402
IL08-12174	Illinois	IL01-36115/IL02-5030
IL08-12206	Illinois	IL01-36115/IL02-9074
IL08-22075	Illinois	IL00-8530/VA01-476//IL79-002DH
IL08-31639	Illinois	IL01-13830/IL00-1665
IL08-33373	Illinois	IL79-005T-B-B/IL00-8530
IL08-33951	Illinois	IL01-36115/IL79-008T
IL08-34020	Illinois	IL01-36115/IL79-008T
IL08-9266	Illinois	IL00-8530/IL02-5675
IL99-26442	Illinois	IL87-2894/Pion2571
INW0411	Purdue	96204A1-12//Goldfield/INW9824(=92823A1-11)
INW0412	Purdue	Huapei57-2(Acc3130)/Patterson
INW1021	Purdue	981129A1/99793RE2//INW0301/92145E8
JAMESTOWN	Virginia	ROANE/PIONEER2691
JAYPEE	Arkansas	Arthur 6/AR39-3 (Doublecrop//Forlani/Garibaldo)
KY02C-1058-03	Kentucky	25R37//Tribute/2552
KY02C-1076-07	Kentucky	25R44//VA97W-24/2552
KY02C-1121-11	Kentucky	Declaration/Tribute
KY02C-1121-75	Kentucky	Declaration/Tribute
KY02C-1122-06	Kentucky	Declaration/25R44
KY02C-2215-02	Kentucky	Tribute/25W33
KY02C-3004-07	Kentucky	25R18/Tribute
KY02C-3005-25	Kentucky	25R18/MCCORMICK
KY03C-1002-02	Kentucky	25W33/25W60//25W33/KY90C-042-37-1
KY03C-1192-37	Kentucky	KY93C-0876-66//KY96C-0059-21
KY03C-1195-10-1-5	Kentucky	KY92C-0010-17//25R18/KY92C-0017-17
KY03C-1221-01	Kentucky	25R18/McCormick//KY96C-0059-21
KY03C-1221-06	Kentucky	25R18/McCormick//KY96C-0059-21
KY03C-1221-22	Kentucky	25R18/McCormick//KY96C-0059-21
KY03C-1237-01	Kentucky	25R18/92C-0010-17//KY96C-0767-1
KY03C-1237-15	Kentucky	25R18/92C-0010-17//KY96C-0767-1
KY03C-1237-32	Kentucky	25R18/92C-0010-17//KY96C-0767-1
KY03C-2047-02	Kentucky	Roane/McCormick

GENOTYPE	ORIGIN	PEDIGREE
KY03C-2047-06	Kentucky	Roane/McCormick
KY03C-2049-02	Kentucky	Roane/VA97W-375ws
KY03C-2314-08	Kentucky	KY93C-1238-17-1/KY94C-0325-40-2
KY03C-2399-02	Kentucky	KY93C-1238-17-6/KY94C-0285-55-3
KY04C-1128-38-1-5	Kentucky	SX1411/NC98-26192//25R78
KY04C-2006-41-1-1	Kentucky	Roane/KY93C-1238-17-1
KY04C-2151-40	Kentucky	25R18/VA01W-476
KY04C-2151-41	Kentucky	25R18/VA01W-476
KY04C-3006-33-14-3	Kentucky	KY93C-0004-22-1/Tribute
KY05C-1007-2-12-5	Kentucky	IL96-24851-1/KY93C-1238-17-1
KY05C-1105-42-20-1	Kentucky	Roane/Truman//KY98C-1440-01
KY05C-1381-77-7-5	Kentucky	KY93C-1238-17-5/CG514W//KY96C-0769-7-1
KY05C-1617-17-17-3	Kentucky	KY94C-0094-11-2/26R15//KY98C-1169-06
KY06C-1003-139-8-3	Kentucky	Truman/McCormick//25R37
KY93C-1238-17-1	Kentucky	VA87-54-558/KY83C-004//2510
MALABAR	Ohio	P92118B4-2/OH561
MASSEY	Virginia	BLUEBOY/KNOX62
MD01W270-10-3	Maryland	VA98W769/USG3209
MD03W104-10-2	Maryland	KY96C-0768-2/MCCORMICK
MD03W151-10-12	Maryland	KY96C-0768-1/MCCORMICK
MD03W485-10-10	Maryland	USG3209/TRIBUTE//MD71-5(USG3342"S")
MD03W485-10-12	Maryland	USG3209/TRIBUTE//MD71-5(USG3342"S")
MD03W485-10-2	Maryland	USG3209/TRIBUTE//MD71-5(USG3342"S")
MD03W485-10-8	Maryland	USG3209/TRIBUTE//MD71-5(USG3342"S")
MD03W61-11-2	Maryland	25R42/Chesapeake
MD03W61-11-3	Maryland	25R42/Chesapeake
MD03W64-10-3	Maryland	26R61/CHESAPEAKE
MD03W665-10-3	Maryland	USG3209/TRIBUTE//CHESAPEAKE
MD03W665-10-5	Maryland	USG3209/TRIBUTE//CHESAPEAKE
MD04W1197-11-13	Maryland	KY96C-0768-2/PI612546//Sx1411/McCormick
MD04W249-11-12	Maryland	MV8-29/25R42
MD04W249-11-13	Maryland	MV8-29/25R42
MD04W249-11-5	Maryland	MV8-29/25R42
MD04W249-11-7	Maryland	MV8-29/25R42
MD04W359-11-10	Maryland	SX1411/25R49//MD71-5(USG3342"S")
MD04W8-11-4	Maryland	Chesapeake/25R42
MD05W10208-11-13	Maryland	Tribute/25R42//Chesapeake
MD05W10208-11-14	Maryland	Tribute/25R42//Chesapeake
MD05W10208-11-3	Maryland	Tribute/25R42//Chesapeake
MD05W10208-11-6	Maryland	Tribute/25R42//Chesapeake
MD05W10208-11-7	Maryland	Tribute/25R42//Chesapeake
MD05W10208-11-8	Maryland	Tribute/25R42//Chesapeake
MD05W1292-11-1	Maryland	Neuse/25R42//Dominion
MD05W1292-11-4	Maryland	Neuse/25R42//Dominion
MD05W1317-11-4	Maryland	Choptank/25R42//Dominion

GENOTYPE	ORIGIN	PEDIGREE
MD05W479-B-11-3	Maryland	26R61/25R42//Chesapeake
MD07W272-11-5	Maryland	VA02W713//USG3555/25R42
MD07W419UM5-11-11	Maryland	USG3209//SS8641/25R42
MD07W419UM5-11-12	Maryland	USG3209//SS8641/25R42
MD665-09-6	Maryland	USG3209/TRIBUTE//CHESAPEAKE
MEDINA	Cornell	MD286-21/Harus
MERL	Virginia	ROANE/PIONEER2643//SS520
MILTON	Missouri	MO94-103/PL2552
MO050921	Missouri	980521/ERNIE
MO080103	Missouri	L910097(releasedasCOKER9704)/MO92-599
MO080104	Missouri	L910097(releasedasCOKER9704)/MO92-599
MO080584	Missouri	980525//APPATTON/980525
MO080589	Missouri	KY90C-383-18-1/IL94-1653
MO080864	Missouri	981020//P92201D5-2/98072
MO081163	Missouri	980429/980525
MO081280	Missouri	980829//980725/IL95-4162
MO081537	Missouri	KY90C-383-18-1/IL94-1653
MO081559	Missouri	980725/Sumai3
MO081652	Missouri	L910097(releasedasCOKER9704)/MO92-599
MO081699	Missouri	Pioneer2552/980829
MO090574	Missouri	L910097(releasedasCOKER9704)/MO92-599
MO090581	Missouri	980521/950016
MO090821	Missouri	980829//980725/IL95-4162
MO091011	Missouri	980927/980525
MO091159	Missouri	PL25R49/000822
MO100172	Missouri	960304/960815
MO100231	Missouri	981020/IL96-346
MO100265	Missouri	980725/SUMAI3
MO100519	Missouri	001567/000442
MO100535	Missouri	980525//980725/ROANE
MO100539	Missouri	980525//980525/ROANE
MO100647	Missouri	ROANE/980525//980525/451-1-2
MO100745	Missouri	980725/SUMAI3
MO101142	Missouri	001567/000442
MO101202	Missouri	981020/IL96-346
MO101207	Missouri	000311/011174
MO101278	Missouri	002246/980829
MO101329	Missouri	980525//980725/ROANE
MO101358	Missouri	PATTON/000926
MO101361	Missouri	002762/001839
MO101571	Missouri	001655/981020
NY103-208-7263	Cornell	Cayuga/Caledonia
NY91017-8080	Cornell	U1266-4-11-6/Harus
NY96009-3037	Cornell	NYBatavia/GenevaResel
NY99066-3444	Cornell	NY87048W-7387(84074(Ho/SuMei)/Harus)/Mendon

GENOTYPE	ORIGIN	PEDIGREE
OH05-200-74	Ohio	OH629/HOPEWELL
OH06-150-57	Ohio	P.92201D5-2-29/OH708
OH06-180-57	Ohio	KY90C-042-37-1/OH687
OH07-166-41	Ohio	OH708/OH684
OH07-166-49	Ohio	OH708/OH684
OH07-174-11	Ohio	OH708/P.92145E8-7-7-1-9
OH07-238-15	Ohio	P.92145E8-7-7-1-9/TRIBUTE
OH07-254-11	Ohio	OH728/VA97W-361WS
OH07-263-3	Ohio	OH748/BRAVO
OH07-94-70	Ohio	ROANE/PATTON
OH07-95-7	Ohio	ROANE/PATTON
OH07-98-21	Ohio	FOSTER/IL95-947
OH08-101-57	Ohio	TRUMAN/IL96-6472
OH08-101-72	Ohio	TRUMAN/IL96-6472
OH08-107-16	Ohio	TRUMAN/OH751
OH08-133-25	Ohio	HONEY/COKER9663
OH08-141-6	Ohio	HONEY/ROANE
OH08-149-11	Ohio	SISSON/HOPEWELL
OH08-161-4	Ohio	OH751/OH738
OH08-161-78	Ohio	OH751/OH738
OH08-170-66	Ohio	DOUGLAS/IL97-3632
OH08-172-42	Ohio	DOUGLAS/JEKYL
OH08-178-52	Ohio	DOUGLAS/P.92226E2-5-3
OH08-180-48	Ohio	DOUGLAS/MCCORMICK
OH08-182-4	Ohio	DOUGLAS/OH708
OH08-199-1	Ohio	VA98W-706/IL96-6472
OH08-206-19	Ohio	P.92226E2-5-3/OH751
OH08-207-33	Ohio	P.92226E2-5-3/OH751
OH08-234-4	Ohio	OH738/OH740
OH08-235-33	Ohio	OH738/OH740
OH08-246-15	Ohio	P.961341A3-2-2/OH740
OH08-254-22	Ohio	TRIBUTE/P.92226E2-5-3
OH08-256-47	Ohio	TRIBUTE/P.92226E2-5-3
OH08-265-37	Ohio	DOUGLAS/P.92226E2-5-3
OH08-269-58	Ohio	P.92226E2-5-3/OH708
OH08-98-13	Ohio	TRUMAN/IL97-3632
PEMBROKE	Kentucky	VA94-52-25/KY87C-42-8-5//2552
PIONEER25R26	Pioneer Hi-Bred	PIONEER-2548(SIB)/W-9057-C//W-9018-A/(SIB)PIONEER-2555
REDRUBY	Michigan	PioneerBrand2552/PioneerBrand27W37
ROANE	Virginia	VA71-54-147/COKER68-15//IN65309C1-18-2-3-2
SHIRLEY	Virginia	VA94-52-25/COKER9835//SISSON"S"(VA96-54-234)
SISSON	Virginia	COKER9803/FREEDOM
SS520	Virginia	FFR555W/GA-GORE
SS5205	Virginia	PIONEER2684/VA93-54-185//POCAHONTAS
SSMPV57	Virginia	FFR555W/VA89-22-52

GENOTYPE	ORIGIN	PEDIGREE
TRIBUTE	Virginia	VA92-51-39/AL870365
TRUMAN	Missouri	MO11769/Madison
USG3209	Virginia	SALUDA/4/MASSEY*2/3/MASSEY*3/BALKAN//SALUDA
USG3315	Virginia	SS520/PIONEER2552//ROANE
USG3555	Virginia	VA94-52-60/PIONEER2643//USG3209
VA05W-151	Virginia	PIONEER26R24/McCORMICK
VA05W-251	Virginia	VA98W-130//VA88-54-328//VA96W-348/PIONEER26R61
VA06W-412	Virginia	TRIBUTE/AGS2000//VAN99W-20
VA07W-415	Virginia	VA98W-895/GA881130LE5//VA98W-627RS
VA08MAS-369	Virginia	McCORMICK/GA881130LE5
VA08W-176	Virginia	KY96C-0079-5/McCORMICK
VA08W-294	Virginia	SS520/VA99W-188//TRIBUTE
VA08W-613	Virginia	FREEDOM/NEUSE"S"//VA98W-688
VA09W-110	Virginia	USG3592(GA931241E16)/VA01W-303
VA09W-112	Virginia	USG3592(GA931241E16)/VA01W-303
VA09W-114	Virginia	USG3592(GA931241E16)/VA01W-303
VA09W-188WS	Virginia	PIONEER25W60//PIONEER25W33/VAN98W-170WS
VA09W-46	Virginia	GF921221E16/McCORMICK"S"(VA98W-590)//VA99W-200
VA09W-52	Virginia	GF921221E16/McCORMICK"S"(VA98W-590)//VA99W-200
VA09W-69	Virginia	SS520(VA96W-158)/VA99W-188//TRIBUTE
VA09W-73	Virginia	SS520/VA99W-188//TRIBUTE
VA09W-75	Virginia	SS520/VA99W-188//TRIBUTE
VA10W-119	Virginia	KY97C-0540-04/GF951079-2E31
VA10W-123	Virginia	PIONEER26R47/GF951079-2E31
VA10W-125	Virginia	PIONEER26R47/JAMESTOWN
VA10W-140	Virginia	VA01W-210/SS520(VA96W-158)//TRIBUTE
VA10W-21	Virginia	Z00-5018/VA01W-158
VA10W-28	Virginia	SSMPV57(VA97W-24)/M99*3098
VA10W-663	Virginia	P97397B1-4-5/MCCORMICK//COKER9511
VA96W-247	Virginia	Coker9803/Freedom

APPENDIX G: List of genotypes included in the yield validation panel (YVP) of chapter III

The table below lists all genotypes included in the yield validation panel (YVP) tested in chapter III, with their state of origin and pedigree (if known).

Table G.1: List of genotypes included in the yield validation panel (YVP) of chapter III

GENOTYPE	ORIGIN	PEDIGREE
04606RA1-1-7-1-6-3	Purdue	TRUMAN/961341A3-1-4-6
04606RA1-7-1	Purdue	TRUMAN/961341A3-1-4-6
04620A1-1-7-4-10	Purdue	TRUMAN/9017C1//92823A1/9218B4/3/P107/4/PATT/5/INW9811/GLD//96204A1
04620A1-1-7-4-17	Purdue	TRUMAN/9017C1//92823A1/9218B4/3/P107/4/PATT/5/INW9811/GLD//96204A1
05247A1-7-3-120	Purdue	99840C4/5/INW0315/3/INW0301MADSEN//INW0315/4/97395B1/6/99840C4//99794RA1
05247A1-7-3-54	Purdue	99840C4/5/INW0315/3/INW0301MADSEN//INW0315/4/97395B1/6/99840C4//99794RA1
05269A1-4-9-8	Purdue	03718A1=981542A1/5/INW0301/INW0315/6/97395C1/7/RSI5//INW0304/3/981281/4/INW0301
05269A1-4-9-83	Purdue	03718A1=981542A1/5/INW0301/INW0315/6/97395C1/7/RSI5//INW0304/3/981281/4/INW0301
0566A1-3-1-3	Purdue	INW0412/6/9017C1//92823A1/9218B4/3/P107/4/PATT/5/ACC3130/PATT/7/992060G1-1
0566A1-3-1-52	Purdue	INW0412/6/9017C1//92823A1/9218B4/3/P107/4/PATT/5/ACC3130/PATT/7/992060G1-1
081B1-12-4	Purdue	INW0411/92226E2-5-3-17-7
081B1-8-3	Purdue	INW0411/92226E2-5-3-17-7
082A1-3-1	Purdue	INW0411/961341A3-1-4-6
082A1-55-4	Purdue	INW0411/961341A3-1-4-6
BRANSON	Ohio	PIONEER-2737-W//(891-4584-A)PIKE/FL-302
IL11-12356	Illinois	IL02-23168/IL05-10461
IL11-12437	Illinois	IL02-23168/IL05-10461
IL11-12443	Illinois	IL02-23168/IL05-10461
IL11-15073	Illinois	IL05-10454/IL02-18228
IL11-15146	Illinois	IL05-10454/IL02-18228
IL11-15603	Illinois	IL05-10461/IL02-18228
IL11-15604	Illinois	IL05-10461/IL02-18228
IL11-15624	Illinois	IL05-10461/IL02-18228
IL11-15671	Illinois	IL05-10461/IL02-18228
IL11-15674	Illinois	IL05-10461/IL02-18228
IL11-19878	Illinois	PIO25R47/IL02-18228/IL/IL00-8530
IL11-19911	Illinois	PIO25R47/IL02-18228/IL/IL00-8530
IL11-19942	Illinois	PIO25R47/IL02-18228/IL/IL00-8530
IL11-19945	Illinois	PIO25R47/IL02-18228/IL/IL00-8530

GENOTYPE	ORIGIN	PEDIGREE
IL11-20960	Illinois	Pembroke/IL00-8530//OH02-12686
IL11-20964	Illinois	Pembroke/IL00-8530//OH02-12686
IL11-21016	Illinois	Pembroke/IL00-8530//OH02-12686
IL11-21037	Illinois	Pembroke/IL00-8530//OH02-12686
IL11-21046	Illinois	Pembroke/IL00-8530//OH02-12686
IL11-23417	Illinois	IL79-002T-B-B/IL00-8530/IL/IL00-8530
IL11-23452	Illinois	IL79-002T-B-B/IL00-8530/IL/IL00-8530
IL11-25566	Illinois	IL00-8530/IL00-8061/IL/IL79-002T-B-B/IL00-8530
IL11-25567	Illinois	IL00-8530/IL00-8061/IL/IL79-002T-B-B/IL00-8530
IL11-28576	Illinois	IL02-19463/IL00-8061/IL/IL00-8109
IL11-28619	Illinois	IL02-19463/IL00-8061/IL/IL00-8109
IL11-32306	Illinois	IL04-10721/IL00-8530/IL/IL00-8641/IL00-8530
IL11-32335	Illinois	IL04-10721/IL00-8530/IL/IL00-8641/IL00-8530
IL11-33027	Illinois	IL00-8530/IL97-1828
IL11-33060	Illinois	IL00-8530/IL97-1828
IL11-33105	Illinois	IL00-8530/IL97-1828
IL11-3407	Illinois	MO050699/IL02-19463
IL11-3411	Illinois	MO050699/IL02-19463
IL11-3433	Illinois	MO050699/IL02-19463
IL11-3437	Illinois	MO050699/IL02-19463
IL11-3466	Illinois	MO050699/IL05-10454
IL11-3517	Illinois	MO050699/IL05-10454
IL11-3531	Illinois	MO050699/IL05-10454
IL11-3538	Illinois	MO050699/IL05-10454
IL11-4147	Illinois	P0179A1-17/IL02-23168
IL11-4148	Illinois	P0179A1-17/IL02-23168
IL11-4166	Illinois	P0179A1-17/IL02-23168
IL11-4619	Illinois	P03207A1-7/IL02-23168
IL11-4620	Illinois	P03207A1-7/IL02-23168
IL11-4659	Illinois	P03207A1-7/IL02-23168
IL11-4663	Illinois	P03207A1-7/IL02-23168
IL11-4734	Illinois	P03207A1-7/IL05-14521
IL11-4768	Illinois	P03207A1-7/IL05-14521
IL11-6175	Illinois	IL97-1828/IL79-002T-B-B
IL11-6215	Illinois	IL97-1828/IL79-002T-B-B
IL11-6216	Illinois	IL97-1828/IL79-002T-B-B
IL11-8112	Illinois	IL00-8061/IL02-7735
IL11-8122	Illinois	IL00-8061/IL02-7735
IL11-8136	Illinois	IL00-8061/IL02-7735
IL11-8139	Illinois	IL00-8061/IL02-7735
IL11-8160	Illinois	IL00-8061/IL02-7735
KY07C-1214-160-2-3	Kentucky	25R37/Truman//Cooper
KY07C-1214-160-3-3	Kentucky	25R37/Truman//Cooper
KY07C-1214-160-8-1	Kentucky	25R37/Truman//Cooper
KY07C-1249-147-13-3	Kentucky	KY97C-0574-01-04/25R37//Pembroke

GENOTYPE	ORIGIN	PEDIGREE
KY07C-1249-147-19-3	Kentucky	KY97C-0574-01-04/25R37//Pembroke
KY07C-1249-148-3-1	Kentucky	KY97C-0574-01-04/25R37//Pembroke
KY07C-1252-150-17-3	Kentucky	KY97C-0574-01-04/25R37//Coker9511
KY07C-1252-151-1-1	Kentucky	KY97C-0574-01-04/25R37//Coker9511
KY07C-1261-151-19-1	Kentucky	KY97C-0546-17-01/Cooper//SSMPV-57
KY07C-1261-151-19-3	Kentucky	KY97C-0546-17-01/Cooper//SSMPV-57
KY07C-1261-152-16-5	Kentucky	KY97C-0546-17-01/Cooper//SSMPV-57
KY07C-1261-152-6-3	Kentucky	KY97C-0546-17-01/Cooper//SSMPV-57
KY07C-1261-152-8-1	Kentucky	KY97C-0546-17-01/Cooper//SSMPV-57
KY07C-1272-162-2-1	Kentucky	KY97C-0519-04-05/KY93C-1238-17-1//Pembroke
KY07C-1272-163-7-3	Kentucky	KY97C-0519-04-05/KY93C-1238-17-1//Pembroke
KY07C-1272-166-10-5	Kentucky	KY97C-0519-04-05/KY93C-1238-17-1//Pembroke
KY07C-1272-166-8-1	Kentucky	KY97C-0519-04-05/KY93C-1238-17-1//Pembroke
KY07C-1278-169-11-3	Kentucky	KY97C-0519-04-05/KY93C-1238-17-1//KY97C-0540-01-03
KY07C-1278-169-12-3	Kentucky	KY97C-0519-04-05/KY93C-1238-17-1//KY97C-0540-01-03
KY07C-1278-169-14-1	Kentucky	KY97C-0519-04-05/KY93C-1238-17-1//KY97C-0540-01-03
KY07C-1278-169-17-3	Kentucky	KY97C-0519-04-05/KY93C-1238-17-1//KY97C-0540-01-03
KY07C-1281-168-14-1	Kentucky	KY97C-0519-04-05/KY93C-1238-17-1//NC03-11465
KY07C-1281-168-15-1	Kentucky	KY97C-0519-04-05/KY93C-1238-17-1//NC03-11465
KY07C-1287-171-16-1	Kentucky	KY97C-0519-04-05/IL96-3073//KY97C-0540-01-03
KY07C-1287-171-6-3	Kentucky	KY97C-0519-04-05/IL96-3073//KY97C-0540-01-03
KY07C-1308-176-11-5	Kentucky	KY97C-0277-01-06/25R37//SSMPV-57
KY07C-1308-176-1-5	Kentucky	KY97C-0277-01-06/25R37//SSMPV-57
KY07C-1308-176-2-3	Kentucky	KY97C-0277-01-06/25R37//SSMPV-57
KY07C-1308-176-6-3	Kentucky	KY97C-0277-01-06/25R37//SSMPV-57
KY07C-1309-177-10-3	Kentucky	KY97C-0277-01-06/25R37//KY96C-0786-3-2
KY07C-1309-177-9-3	Kentucky	KY97C-0277-01-06/25R37//KY96C-0786-3-2
KY07C-1326-179-20-3	Kentucky	KY97C-0299-13-01/25R37//SSMPV-57
KY07C-1326-180-3-3	Kentucky	KY97C-0299-13-01/25R37//SSMPV-57
KY07C-1332-183-15-5	Kentucky	KY97C-0299-13-01/KY96C-0770-3//SSMPV-57
KY07C-1332-184-12-5	Kentucky	KY97C-0299-13-01/KY96C-0770-3//SSMPV-57
MD05W10208-12-12	Maryland	unknown
MD05W10208-12-14	Maryland	unknown
MD05W10208-12-16	Maryland	unknown
MD05W10208-12-6	Maryland	unknown
MD05W10208-12-7	Maryland	unknown
MILTON	Missouri	HART/W-8376//PIONEER-2555/3/PIONEER-2552
MO080864	Missouri	981020//P92201D5-2-80/980725
MO120104	Missouri	080864SPRS(Bess//P92201D5-2-80/MO980725)
MO120114	Missouri	080864SPRS(Bess//P92201D5-2-80/MO980725)
MO120331	Missouri	419-2-4/980829//980725/Roane
MO120487	Missouri	980525*2/IL95-4162
MO120655	Missouri	980725//980725/IL95-4162
MO120666	Missouri	980525*2/IL95-4162
MO120723	Missouri	980725//980725/IL95-4162

GENOTYPE	ORIGIN	PEDIGREE
MO120777	Missouri	980525*2/IL95-4162
MO120875	Missouri	080864SPRS(Bess//P92201D5-2-80/MO980725)
MO121058	Missouri	011126*2/PL25R47(011126=Milton)
MO121175	Missouri	980725//980725/IL95-4162
MO121207	Missouri	981020//P92001D5-2-80/980525
MO121271	Missouri	981020/002001
MO121396	Missouri	002946//980829/Ernie
MO121442	Missouri	980725//980725/IL95-4162
MO121457	Missouri	981020/002001
MO121539	Missouri	981020//981020/IL95-4162
MO121559	Missouri	981020//981020/IL95-4162
MO121624	Missouri	980525//981020/APPatton
MO121658	Missouri	Roane/000917
MO121681	Missouri	002946//980829/Ernie
MO121695	Missouri	980525//981020/APPatton
MO121947	Missouri	011126*2/PL25R47
MO121983	Missouri	Roane/000917
MO122003	Missouri	980525//981020/APPatton
MO122240	Missouri	080864SPRS(Bess//P92201D5-2-80/MO980725)
MO122312	Missouri	011126*2/PL25R47
MO122352	Missouri	419-2-4/980829//980725/Roane
OH09-204-52	Ohio	M99*3098/OH708
OH09-204-54	Ohio	M99*3098/OH708
OH09-204-66	Ohio	M99*3098/OH708
OH09-205-35	Ohio	M99*3098/OH708
OH09-207-68	Ohio	M99*3098/OH743
OH09-207-8	Ohio	M99*3098/OH743
OH09-208-47	Ohio	OH751/M99*3098
OH09-208-50	Ohio	OH751/M99*3098
OH09-208-58	Ohio	OH751/M99*3098
OH09-215-36	Ohio	P.984RE1-57-5/M99*3098
OH09-216-31	Ohio	P.984RE1-57-5/M99*3098
OH09-217-15	Ohio	P.99608C1-1-3/M99*3098
OH09-218-1	Ohio	P.99608C1-1-3/M99*3098
OH09-218-27	Ohio	P.99608C1-1-3/M99*3098
OH09-223-29	Ohio	M99*3098/VA98W-706
OH09-223-48	Ohio	M99*3098/VA98W-706
OH09-233-41	Ohio	VA97W-375WS/OH708
OH09-233-57	Ohio	VA97W-375WS/OH708
OH09-241-50	Ohio	OH743/P.984RE1-57-5
OH09-241-62	Ohio	OH743/P.984RE1-57-5
OH09-243-15	Ohio	OH743/P.99608C1-1-3
OH09-244-27	Ohio	OH743/P.99608C1-1-3
OH09-250-5	Ohio	OH751/P.984RE1-57-5
OH09-253-56	Ohio	OH751/VA97W-375WS

GENOTYPE	ORIGIN	PEDIGREE
OH09-259-57	Ohio	P.984RE1-57-5/VA97W-375WS
OH09-260-75	Ohio	P.984RE1-57-5/VA97W-375WS
OH09-261-47	Ohio	P.984RE1-57-5/VA97W-375WS
OH09-261-76	Ohio	P.984RE1-57-5/VA97W-375WS
OH09-282-59	Ohio	OH751/P.984RE1-57-5
OH09-283-23	Ohio	OH751/VA97W-375WS
OH09-283-78	Ohio	P.99608C1-1-3/P.984RE1-57-5
OH09-285-10	Ohio	P.99608C1-1-3/P.984RE1-57-5
OH09-285-15	Ohio	P.99608C1-1-3/P.984RE1-57-5
OH11-139-63	Ohio	OH02-12686/(5x695=IL96-24851/B980582)
OH11-139-75	Ohio	OH02-12686/(5x695=IL96-24851/B980582)
OH11-140-17	Ohio	OH02-12686/(5x695=IL96-24851/B980582)
OH11-140-74	Ohio	OH02-12686/(5x695=IL96-24851/B980582)
OH11-147-72	Ohio	P.99840C4-8-3-6/(5x699=IL00-8061/IL96-24851-1)
OH11-147-77	Ohio	P.99840C4-8-3-6/(5x699=IL00-8061/IL96-24851-1)
OH11-150-17	Ohio	OH02-13567/(5x699=IL00-8061/IL96-24851-1)
OH11-150-76	Ohio	OH02-13567/(5x699=IL00-8061/IL96-24851-1)
OH11-151-30	Ohio	KY97C-0067-2/M01-4377
OH11-152-38	Ohio	KY97C-0067-2/M01-4377
OH11-176-21	Ohio	M01-4377/OH01-7664
OH11-176-68	Ohio	M01-4377/OH01-7664
OH11-179-6	Ohio	M01-4377/OH02-13567
OH11-179-74	Ohio	M01-4377/OH02-13567
OH11-179-77	Ohio	M01-4377/OH02-13567
OH11-204-61	Ohio	MO030118/P.992178A3-1-1
OH11-205-18	Ohio	MO030118/P.992178A3-1-1
OH11-226-2	Ohio	P.99608C1-1-3-4/OH02-12686
OH11-227-5	Ohio	P.99608C1-1-3-4/OH02-12686
OH11-228-69	Ohio	P.99608C1-1-3-4/OH02-12686
OH11-241-55	Ohio	OH02-12686/(6x016=OH02-12686/PIO25R47)
OH11-242-70	Ohio	OH02-12686/(6x016=OH02-12686/PIO25R47)
OH11-250-44	Ohio	OH02-12686/(6x029=OH02-13567/PIO25R47)
OH11-251-37	Ohio	P.992178A3-1-1/(6x135=PIO25R47/IL99-12976)
OH11-252-27	Ohio	P.992178A3-1-1/(6x135=PIO25R47/IL99-12976)
OH11-253-13	Ohio	P.992178A3-1-1/(6x135=PIO25R47/IL99-12976)
OH11-257-41	Ohio	VA03W-409/(6x137=PIO25R47/IL99-15867)
OH11-258-53	Ohio	P.992178A3-1-1/(6x138=PIO25R47/P.99608C1-1-3-4)
OH11-259-41	Ohio	P.992178A3-1-1/(6x138=PIO25R47/P.99608C1-1-3-4)
OH11-268-42	Ohio	OH02-15978/OH02-12686
OH11-269-50	Ohio	OH02-15978/OH02-12686
OH11-270-9	Ohio	OH02-15978/OH02-12686
OH11-274-1	Ohio	VA03W-409/OH02-13567
OH11-278-31	Ohio	VA03W-409/P.992178A3-1-1
OH11-278-35	Ohio	VA03W-409/P.992178A3-1-1
OH11-279-32	Ohio	VA03W-409/P.992178A3-1-1

GENOTYPE	ORIGIN	PEDIGREE
OH11-279-76	Ohio	VA03W-409/P.992178A3-1-1
OH11-280-38	Ohio	COOPER/VA03W-409
OH11-280-73	Ohio	COOPER/VA03W-409
OH11-288-58	Ohio	VA03W-409/OH02-13567
OH11-289-44	Ohio	P.99608C1-1-3-4/OH02-12686
OH11-290-62	Ohio	VA03W-409/(6x137=PIO25R47/IL99-15867)
OH11-290-70	Ohio	VA03W-409/(6x137=PIO25R47/IL99-15867)
OH11-293-53	Ohio	COOPER/VA03W-409
OH11-293-70	Ohio	COOPER/VA03W-409
OH11-81-12	Ohio	OH02-12686/P.99840C4-8-3-6
OH11-81-3	Ohio	OH02-12686/P.99840C4-8-3-6
OH11-82-8	Ohio	OH02-12686/P.99840C4-8-3-6
PIONEER 25R47	Pioneer Hi-Bred	FRANKENMUTH/(SIB)PIONEER-2555//(SIB)PIONEER-2551(WBE-2190-B-1)/3/(WBA-416-H-2)HOUSER/MO-9545//W-4034-D/AUGUSTA/4/PIONEER-2552
SHIRLEY	Virginia	VA-94-52-25/COKER-9835//VA-96-54-234
VA07MAS10-8328-7-2-2	Virginia	VA04W-227/M01*1019//SS5205
VA07MAS10-8328-7-3-3	Virginia	VA04W-227/M01*1019//SS5205
VA07MAS12-8752-1-3-1	Virginia	U3960-3R-3-11-6/VA02W-398//GA-96693-4E16
VA07MAS12-8752-4-1-4	Virginia	U3960-3R-3-11-6/VA02W-398//GA-96693-4E16
VA07MAS13-8825-6-3-3	Virginia	NC03-11458/TRIBUTE//SS5205
VA07MAS13-8825-6-4-4	Virginia	NC03-11458/TRIBUTE//SS5205
VA07MAS1-7031-7-1-4	Virginia	McCORMICK/GA951231-4E26//SS5205
VA07MAS1-7054-2-2-4	Virginia	McCORMICK/GA951231-4E26//SS5205
VA07MAS1-7054-3-3-2	Virginia	McCORMICK/GA951231-4E26//SS5205
VA07MAS1-7151-1-2-4	Virginia	McCORMICK/GA951231-4E26//SS5205
VA07MAS2-7199-2-3-1	Virginia	Jamestown/GA951231-4E26//P992060G1-1-5
VA07MAS2-7263-1-2-2	Virginia	Jamestown/GA951231-4E26//P992060G1-1-5
VA07MAS3-7304-3-2-4	Virginia	Shirley/GA951231-4E26//SS8404
VA07MAS3-7313-7-1-2	Virginia	Shirley/GA951231-4E26//SS8404
VA07MAS4-7463-6-4-3	Virginia	GA951231-4E25/SS8404//Shirley
VA07MAS4-7520-2-3-2	Virginia	GA951231-4E25/SS8404//Shirley
VA08MAS2-18-3-3	Virginia	VA05W-693/VA04W-259//SS5205
VA08MAS2-187-7-1	Virginia	VA05W-693/VA04W-259//SS5205
VA11MAS-7313-3-2-162	Virginia	Shirley/AGS2060//SS8404
VA11MAS-7383-6-3-155	Virginia	Shirley/AGS2060//SS8404
VA12FHB-37	Virginia	VA04W-433/SS8404
VA12FHB-50	Virginia	VA04W-433/SS8404
VA12FHB-75	Virginia	IL99-15867/VA04W-433//SS8404
VA12FHB-77	Virginia	IL99-15867/VA04W-433//SS8404
VA12W-150	Virginia	IL99-15867/JAMESTOWN
VA12W-283	Virginia	CHESAPEAKE"S"/SS8641//VA04W-439
VA12W-284	Virginia	CHESAPEAKE"S"/SS8641//VA04W-439
VA12W-51	Virginia	NC00-15389/G/F951079-2E31//USG3555
VA12W-54	Virginia	NC00-15389/G/F951079-2E31//USG3555
VA12W-72	Virginia	PIONEER25R47/G/F951079-2E31//USG3555

GENOTYPE	ORIGIN	PEDIGREE
VA12W-74WS	Virginia	PIONEER25R47/G/F951079-2E31//USG3555
VA12W-95	Virginia	MERL/AGS2026
VA12W-97	Virginia	MERL/AGS2026
VA13FHB-11	Virginia	IL99-15867/VA04W-433//SS8404
VA13FHB-14	Virginia	IL99-15867/VA04W-433//SS8404
VA13FHB-22	Virginia	VA05W-436/VA05W-641
VA13FHB-24	Virginia	VA05W-436/VA05W-641
VA13FHB-31	Virginia	VA05W-641/AGS2020
VA13FHB-33	Virginia	VA05W-641/AGS2020
VA13W-106	Virginia	SHIRLEY/GA98249G1-G1-2
VA13W-121	Virginia	VA05W-251/AGS2026
VA13W-124	Virginia	VA05W-251/AGS2026
VA13W-130	Virginia	VA05W-363/VA03W-310
VA13W-132	Virginia	VA05W-363/VA03W-310
VA13W-136	Virginia	VA05W-363/VA03W-310
VA13W-138	Virginia	VA05W-363/VA03W-310
VA13W-141	Virginia	JAMESTOWN/SS8404//VA04W-259/NC95-11612
VA13W-144	Virginia	JAMESTOWN/SS8404//VA04W-259/NC95-11612
VA13W-146	Virginia	JAMESTOWN/SS8404//AGS2020
VA13W-148	Virginia	JAMESTOWN/SS8404//AGS2020
VA13W-150	Virginia	JAMESTOWN/SS8404//AGS2020
VA13W-155	Virginia	JAMESTOWN/SS8404//AGS2020
VA13W-16	Virginia	SS520/GF951208-2E35//JAMESTOWN
VA13W-176	Virginia	SHIRLEY/BRANSON//JAMESTOWN
VA13W-180	Virginia	SHIRLEY/BRANSON//JAMESTOWN
VA13W-20	Virginia	SS520/GF951208-2E35//JAMESTOWN
VA13W-207	Virginia	SS8404/VA02W-398//AGS2026
VA13W-209	Virginia	SS8404/VA02W-398//AGS2026
VA13W-217	Virginia	M01*1019/VA03W-203//AGS2020
VA13W-219	Virginia	M01*1019/VA03W-203//AGS2020
VA13W-25	Virginia	JAMESTOWN/AGS2020
VA13W-28	Virginia	JAMESTOWN/AGS2020
VA13W-29	Virginia	JAMESTOWN/AGS2020
VA13W-30	Virginia	JAMESTOWN/AGS2020
VA13W-36	Virginia	IL99-15867/JAMESTOWN
VA13W-4	Virginia	FG95195/VA02W-370
VA13W-42	Virginia	IL99-15867/JAMESTOWN
VA13W-53	Virginia	USG3555"S"/SHIRLEY//JAMESTOWN
VA13W-54	Virginia	USG3555"S"/SHIRLEY//JAMESTOWN
VA13W-56	Virginia	USG3555"S"/SHIRLEY//JAMESTOWN
VA13W-57	Virginia	USG3555"S"/SHIRLEY//JAMESTOWN
VA13W-74	Virginia	Tribute/GA961176-3A48//USG3315
VA13W-75	Virginia	Tribute/GA961176-3A48//USG3315
VA13W-8	Virginia	FG95195/VA02W-370
VA13W-83	Virginia	JAMESTOWN/AGS2060//NC02-1957

GENOTYPE	ORIGIN	PEDIGREE
VA13W-84	Virginia	JAMESTOWN/AGS2060//NC02-1957
VA13W-91	Virginia	JAMESTOWN/AGS2020
VA13W-92	Virginia	JAMESTOWN/AGS2020
VA13W-99	Virginia	SHIRLEY/GA98249G1-G1-2

APPENDIX H: Summary of trial environments included in chapter III

The table below lists all environments included in both the elite panel (EP) and yield validation panel (YVP) trials described in chapter III, along with their abbreviation, testing year, testing location, number of replications, nitrogen application rate, and the traits assessed.

Table H.1: Summary of trial environments included in chapter III

Panel †	Abbv ‡	Year	Location	nReps	N rate	Traits Assessed §
EP	12KYM	2012	Lexington, KY	1	Moderate	HD, HGT, TW, YLD
	12MDM	2012	Clarkston, MD	1	Moderate	
	12MOM	2012	Columbia, MO	1	Moderate	
	12OWL	2012	Wooster, OH	2	Low	
	12OWM	2012	Wooster, OH	1	Moderate	
	12VAL	2012	Warsaw, VA	2	Low	
	12VAM	2012	Warsaw, VA	1	Moderate	
	13MOM	2013	Columbia, MO	1	Moderate	
	13ONM	2013	Custar, OH	1	Moderate	HD
	13OVM	2013	Fremont, OH	1	Moderate	
	13OWL	2013	Wooster, OH	2	Low	HD, HGT, TW, YLD
	13OWM	2013	Wooster, OH	1	Moderate	
	13VAL	2013	Warsaw, VA	2	Low	
	13VAM	2013	Warsaw, VA	1	Moderate	
YVP	14KYM	2014	Woodford, KY	1	Moderate	HD, HGT, TW
	14MOL	2014	Columbia, MO	2	Low	HGT, YLD
	14MOM	2014	Columbia, MO	2	Moderate	
	14OWL	2014	Wooster, OH	2	Low	HD, HGT, TW, YLD
	14OWM	2014	Wooster, OH	1	Moderate	HD, HGT, YLD
	14VAM	2014	Warsaw, VA	1	Moderate	HD, HGT, TW, YLD
	15MOL	2015	Columbia, MO	2	Low	HD, HGT, YLD
	15MOM	2015	Columbia, MO	2	Moderate	
	15ONM	2015	Custar, OH	1	Moderate	TW, YLD
	15OVM	2015	Fremont, OH	1	Moderate	
	15OWL	2015	Wooster, OH	2	Low	HD, HGT, TW, YLD
	15OWM	2015	Wooster, OH	1	Moderate	
	15VAM	2015	Warsaw, VA	1	Moderate	

† EP elite panel; YVP yield validation panel

‡ Abbreviation is a combination of year, location, and rate of nitrogen treatment

§ HD heading date; HGT plant height; TW test weight; YLD grain yield

APPENDIX I: Genotype + genotype-environment (GGE) biplots for the elite panel (EP) tested in chapter III

Description

The package 'gge' (Wright and Laffont 2016) running in R (R Core Team 2016) was used to generate genotype + genotype-by-environment interaction plots (Yan et al. 2000, Yan (2001), Yan and Kang (2003)) for all traits for which data was collected in the Elite Panel during the 2011-2012 and 2012-2013 winter wheat growing seasons. Within-environment arithmetic means were used for the analysis. Genotype IDs have been suppressed for the sake of clarity.

Environment Legend

Environments were assigned a code consisting of the testing year, a two-letter abbreviation of the location, and the nitrogen treatment level. The table below gives the abbreviation for each environment, along with its year, location, number of replications, nitrogen application rate, and traits assessed

Table I.1: Environment codes for elite panel (EP) GGE biplots

Abbreviation	Year	Location	nReps	N rate	Traits Assessed
12KYM	2012	Lexington, KY	1	Moderate	HD; HGT; TW; YLD
12MDM	2012	Clarkston, MD	1	Moderate	HD; HGT; TW; YLD
12MOM	2012	Columbia, MO	1	Moderate	HD; HGT; TW; YLD
12OWL	2012	Wooster, OH	2	Low	HD; HGT; TW; YLD
12OWM	2012	Wooster, OH	1	Moderate	HD; HGT; TW; YLD
12VAL	2012	Warsaw, VA	2	Low	HD; HGT; TW; YLD
12VAM	2012	Warsaw, VA	1	Moderate	HD; HGT; TW; YLD
13MOM	2013	Columbia, MO	1	Moderate	HD; HGT; TW; YLD
13ONM	2013	Custar, OH	1	Moderate	HD
13OVM	2013	Fremont, OH	1	Moderate	HD
13OWL	2013	Wooster, OH	2	Low	HD; HGT; TW; YLD
13OWM	2013	Wooster, OH	1	Moderate	HD; HGT; TW; YLD
13VAL	2013	Warsaw, VA	2	Low	HD; HGT; TW; YLD
13VAM	2013	Warsaw, VA	1	Moderate	HD; HGT; TW; YLD

Biplot Generation

GGE biplots for each trait are shown below

Figure I.1: GGE biplot for heading date (HD) in the elite panel

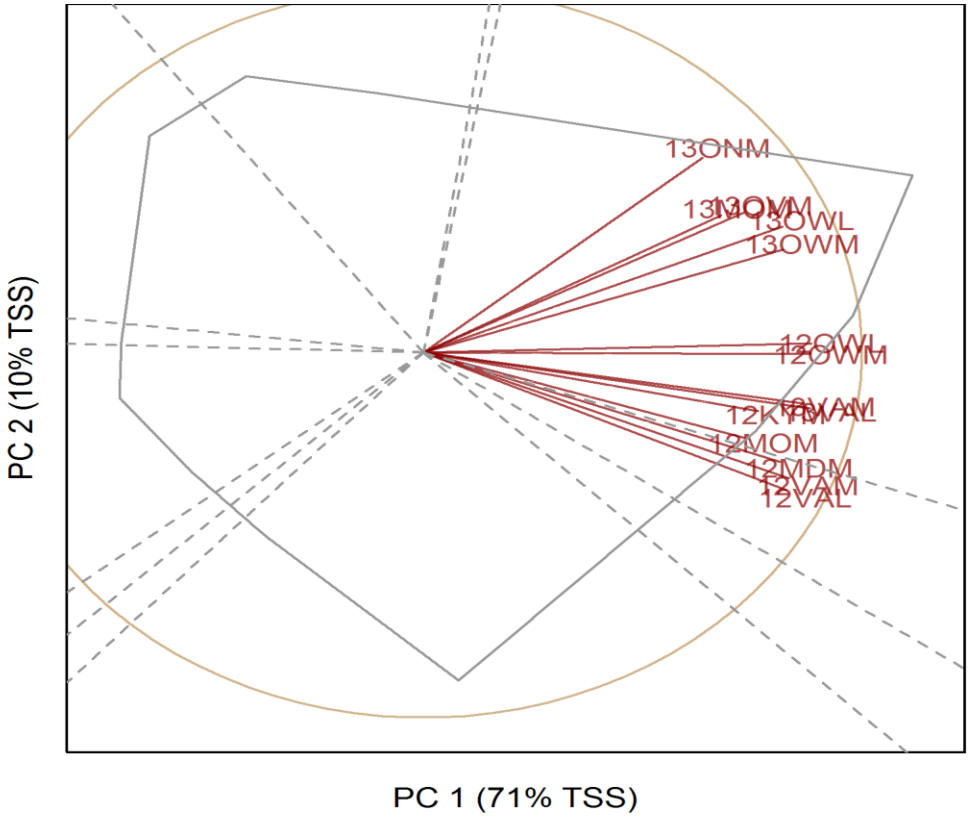


Figure I.2: GGE biplot for plant height (HGT) in the elite panel

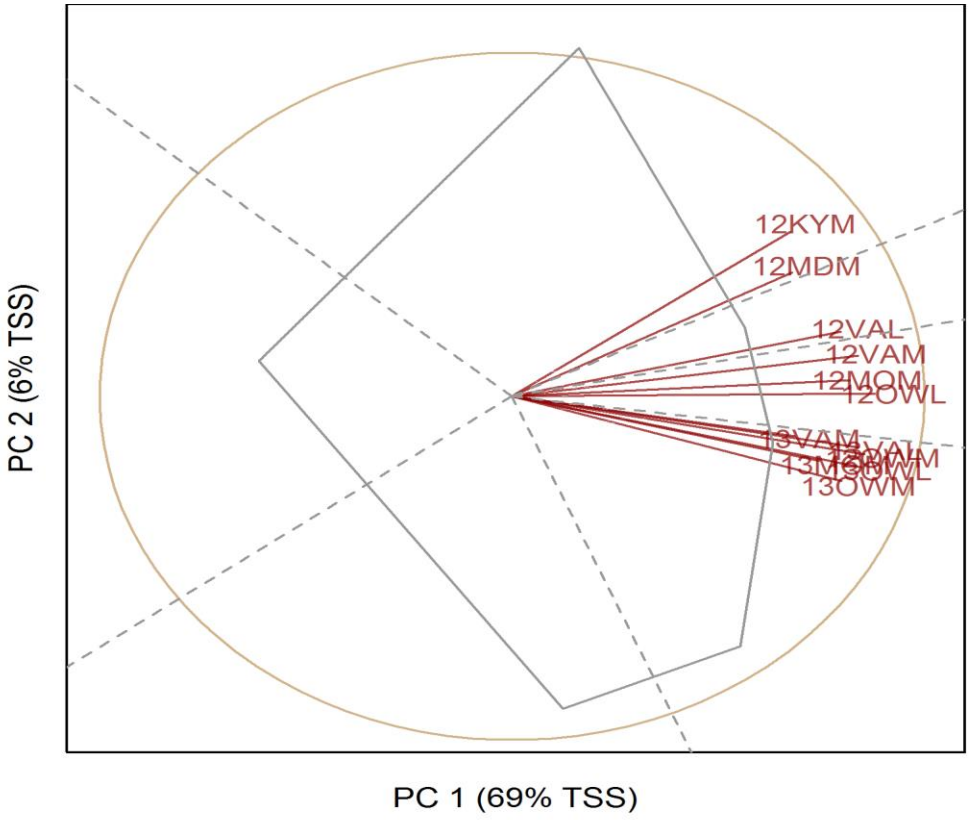


Figure I.3: GGE biplot for test weight (TW) in the elite panel

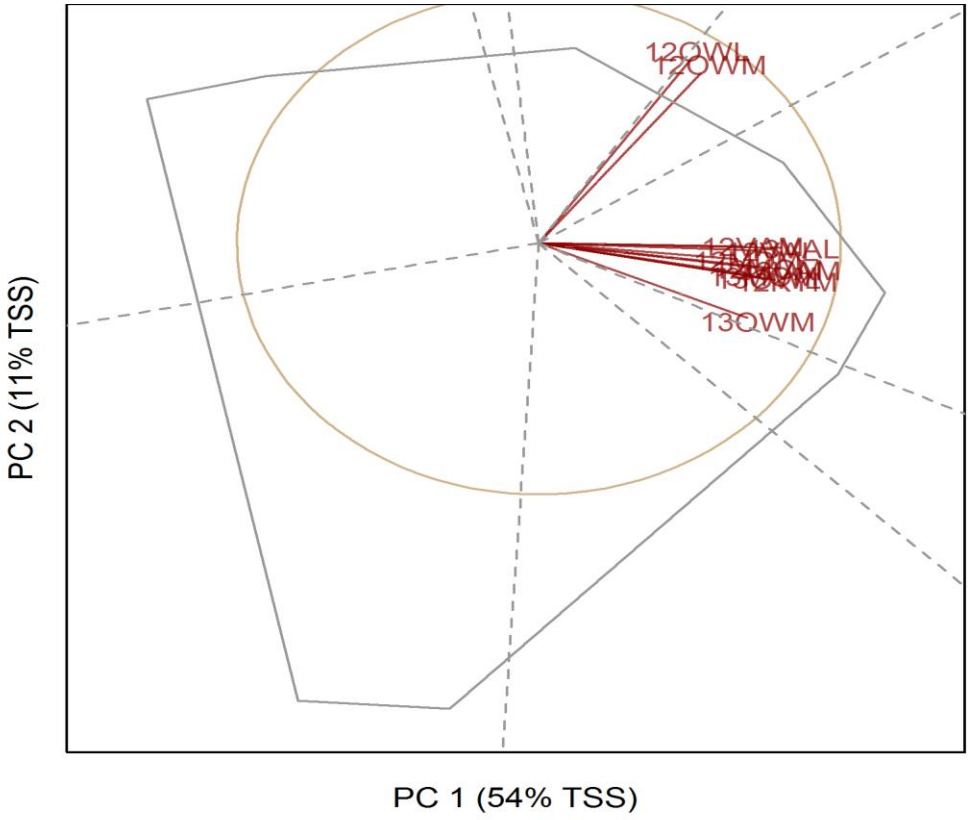
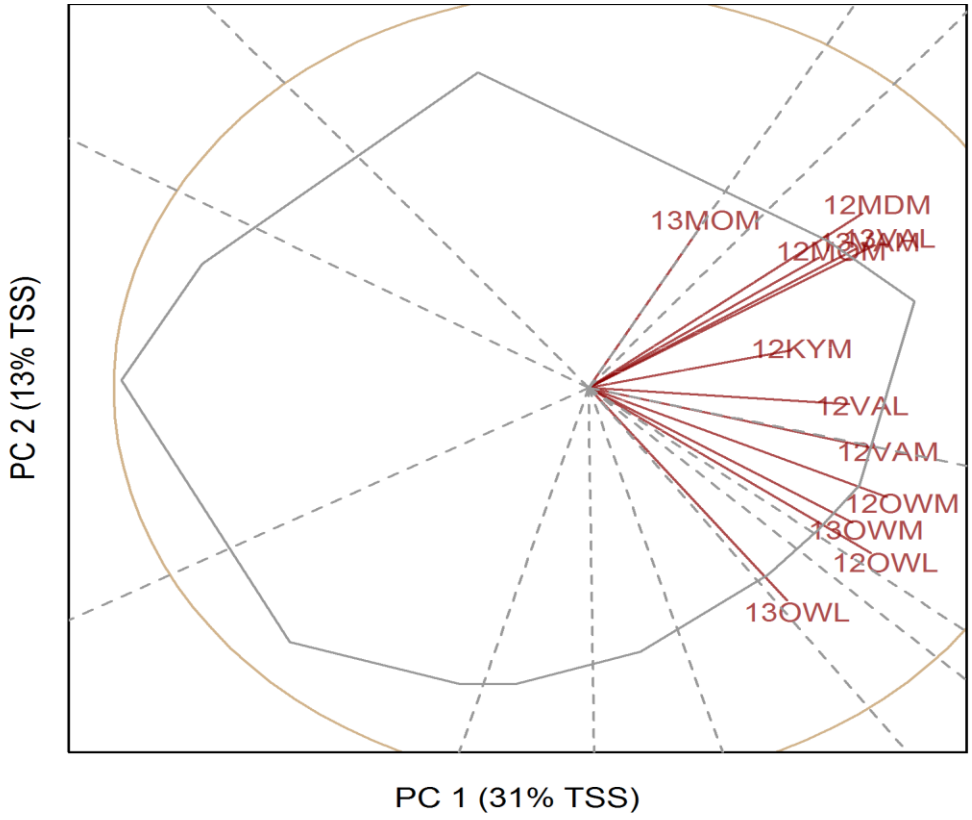


Figure I.4: GGE biplot for grain yield (YLD) in the elite panel



References

R Core Team. 2016. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Wright, Kevin, and Jean-Louis Laffont. 2016. *Gge: Genotype Plus Genotype-by-Environment Biplots*. https://CRAN.R-project.org/package=gge.

Yan, Weikai. 2001. "GGEbiplot—A Windows Application for Graphical Analysis of Multienvironment Trial Data and Other Types of Two-Way Data." *Agronomy Journal* 93 (5): 1111–8. doi:10.2134/agronj2001.9351111x.

Yan, Weikai, and Manjit S Kang. 2003. *GGE Biplot Analysis a Graphical Tool for Breeders, Geneticists, and Agronomists*. Boca Raton, Fla: CRC Press.

Yan, Weikai, L. A. Hunt, Qinglai Sheng, and Zorka Szlavnics. 2000. "Cultivar Evaluation and Mega-Environment Investigation Based on the GGE Biplot." *Crop Science* 40 (3): 597–605. doi:10.2135/cropsci2000.403597x.

APPENDIX J: Genotype + genotype-environment (GGE) biplots for the yield validation panel (YVP) tested in chapter III

Description

The package 'gge' (Wright and Laffont 2016) running in R (R Core Team 2016) was used to generate genotype + genotype-by-environment interaction plots (Yan et al. 2000, Yan (2001), Yan and Kang (2003)) for all traits for which data was collected in the Yield Validation Panel during the 2013-2014 and 2014-2015 winter wheat growing seasons. Within-environment arithmetic means were used for the analysis. Genotype IDs have been suppressed for the sake of clarity.

Environment Legend

Environments were assigned a code consisting of the testing year, a two-letter abbreviation of the location, and the nitrogen treatment level. The table below gives the abbreviation for each environment, along with its year, location, number of replications, nitrogen application rate, and traits assessed:

Table J.1: Environment codes for yield validation panel (YVP) GGE biplots

Abbreviation	Year	Location	nReps	N rate	Traits Assessed
14KYM	2014	Woodford, KY	1	Moderate	HD; HGT; TW
14MOL	2014	Columbia, MO	2	Low	HGT; YLD
14MOM	2014	Columbia, MO	2	Moderate	HGT; YLD
14OWL	2014	Wooster, OH	2	Low	HD; HGT; TW; YLD
14OWM	2014	Wooster, OH	1	Moderate	HD; HGT; YLD
14VAM	2014	Warsaw, VA	1	Moderate	HD; HGT; TW; YLD
15MOL	2015	Columbia, MO	2	Low	HD; HGT; YLD
15MOM	2015	Columbia, MO	2	Moderate	HD; HGT; YLD
15ONM	2015	Custar, OH	1	Moderate	TW; YLD
15OVM	2015	Fremont, OH	1	Moderate	TW; YLD
15OWL	2015	Wooster, OH	2	Low	HD; HGT; TW; YLD
15OWM	2015	Wooster, OH	1	Moderate	HD; HGT; TW; YLD
15VAM	2015	Warsaw, VA	1	Moderate	HD; HGT; TW; YLD

Biplot Generation

GGE biplots for each trait are shown below

Figure J.1: GGE biplot for heading date (HD) in the yield validation panel

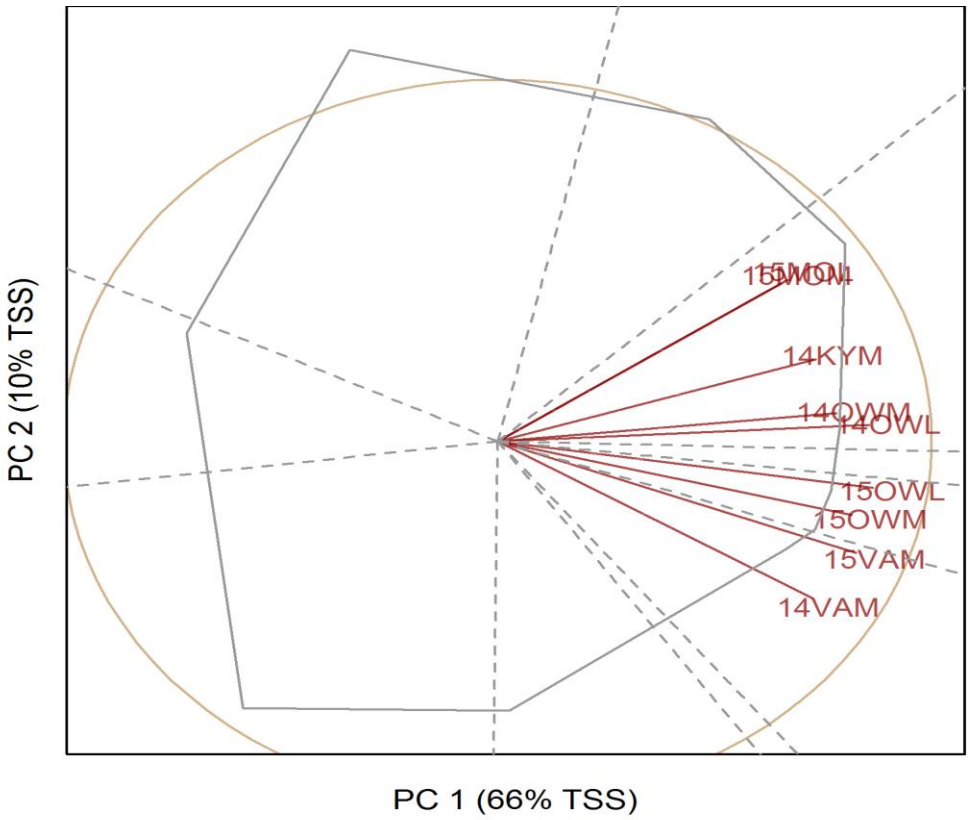


Figure J.2: GGE biplot for plant height (HGT) in the yield validation panel

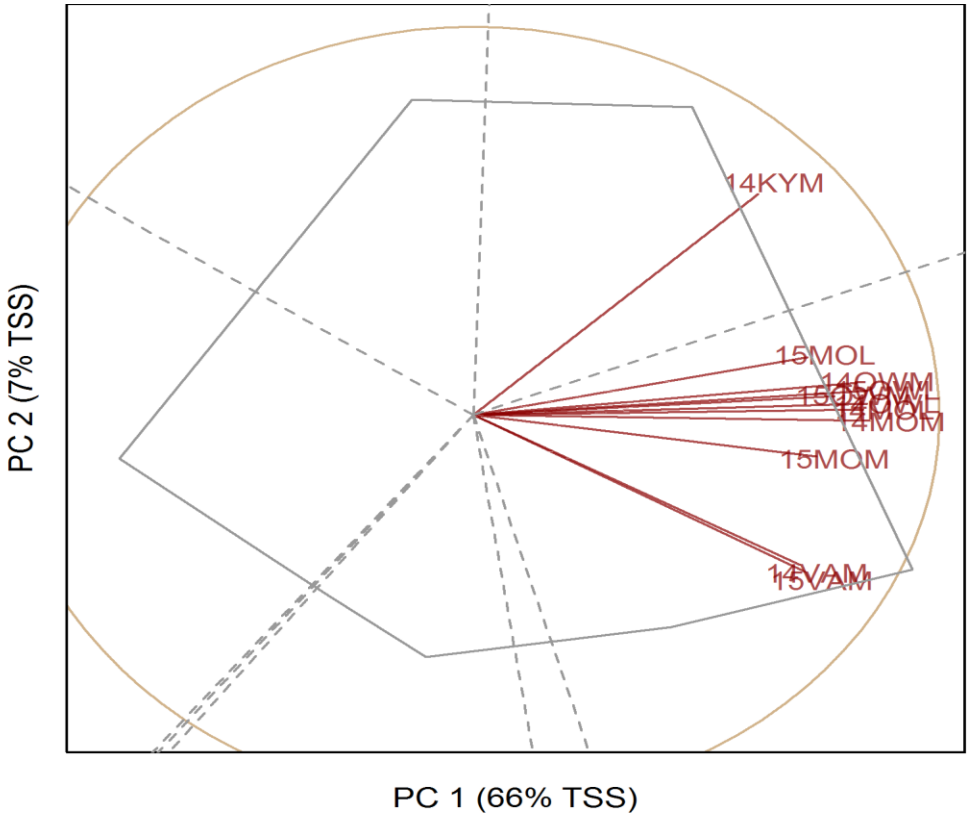


Figure J.3: GGE biplot for test weight (TW) in the yield validation panel

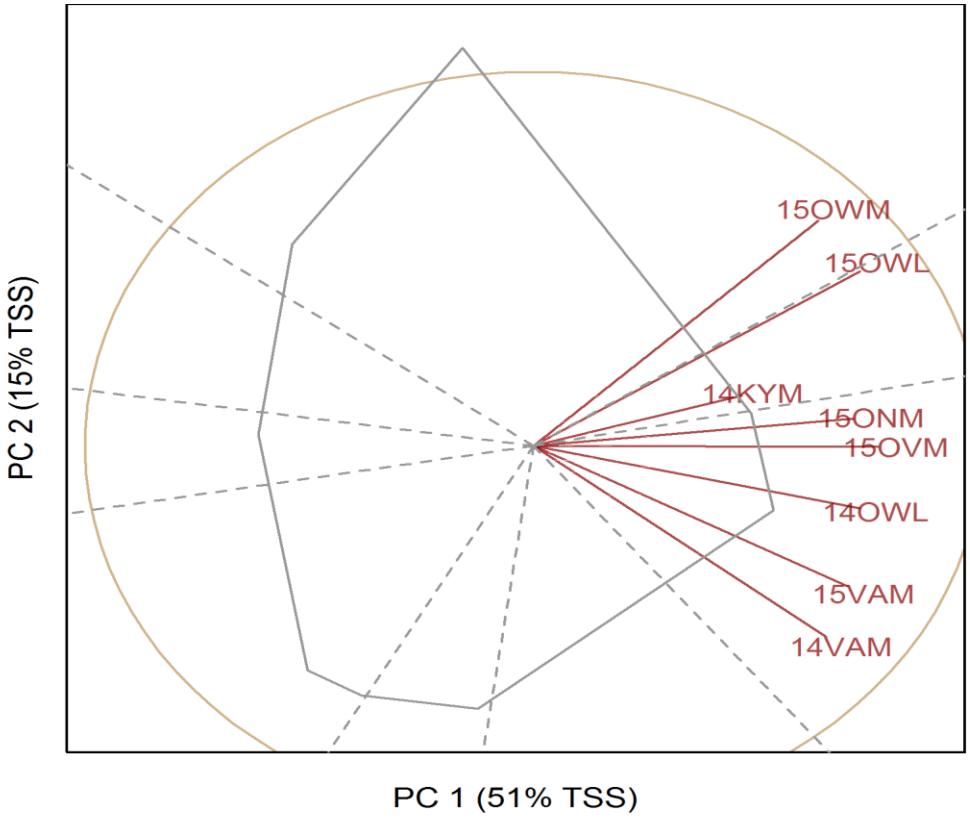
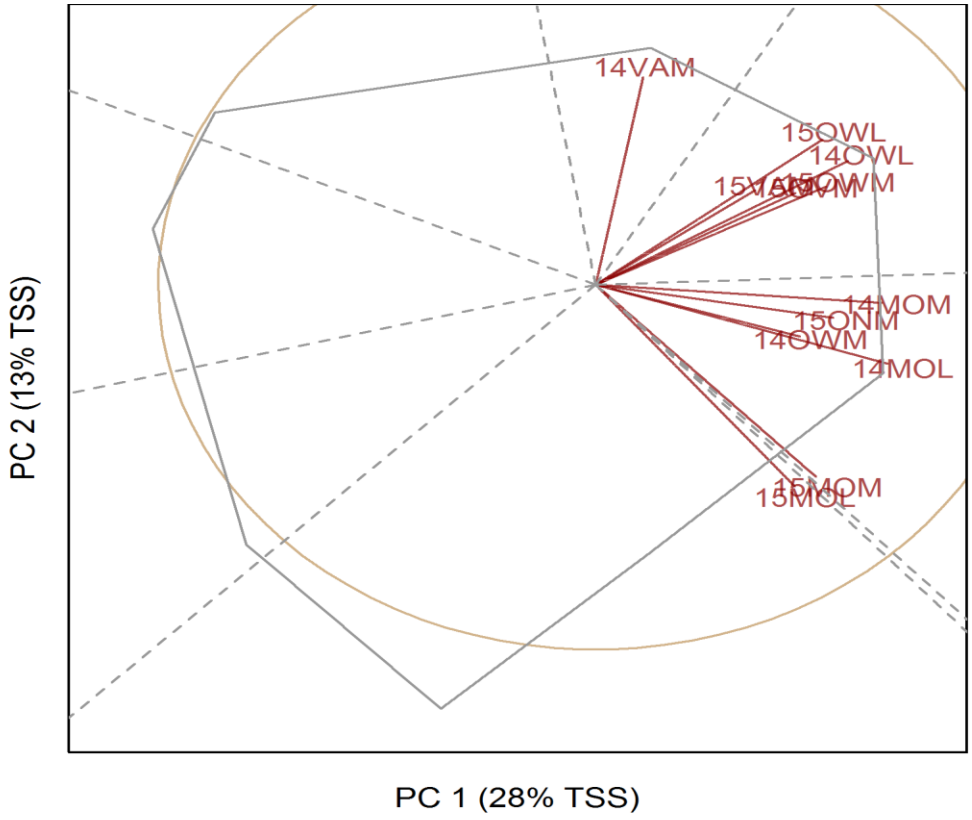


Figure J.4: GGE biplot for grain yield (YLD) in the yield validation panel



References

R Core Team. 2016. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Wright, Kevin, and Jean-Louis Laffont. 2016. *Gge: Genotype Plus Genotype-by-Environment Biplots*. https://CRAN.R-project.org/package=gge.

Yan, Weikai. 2001. "GGEbiplot—A Windows Application for Graphical Analysis of Multienvironment Trial Data and Other Types of Two-Way Data." *Agronomy Journal* 93 (5): 1111–8. doi:10.2134/agronj2001.9351111x.

Yan, Weikai, and Manjit S Kang. 2003. *GGE Biplot Analysis a Graphical Tool for Breeders, Geneticists, and Agronomists*. Boca Raton, Fla: CRC Press.

Yan, Weikai, L. A. Hunt, Qinglai Sheng, and Zorka Szlavnics. 2000. "Cultivar Evaluation and Mega-Environment Investigation Based on the GGE Biplot." *Crop Science* 40 (3): 597–605. doi:10.2135/cropsci2000.403597x.