GENOMIC SELECTION FOR GLUME BLOTCH RESISTANCE AND MILLING AND BAKING QUALITY TRAITS IN SOFT RED WINTER WHEAT

BY

OLIVIA CLARE JONES

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Crop Sciences
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2018

Urbana, Illinois

Master's Committee:

Professor Frederic L. Kolb, Chair
Assistant Professor Santiago X. Mideros
Dr. Carolyn J. Butts-Wilmsmeyer

ABSTRACT

Wheat (*Triticum aestivum* L.*)* is a major cereal crop with global importance, responsible for providing 20% of human calorie intake, commonly in the form of flour. Maintaining superior milling and baking quality while improving disease resistance are key objectives in a breeding program. Selection for milling and baking quality is critical for the acceptance of new wheat varieties to end users - millers and bakers. Fungal pathogens present a significant biotic threat to the quality and quantity of the wheat crop annually. Necrotrophic fungus *Parastagonospora nodorum* (syn. *Stagonospora nodorum, Septoria nodorum*) is a leading fungal threat to wheat production in humid regions. A *P. nodorum* infection results in leaf blotch and glume blotch in wheat and related grass species. The development of varieties possessing resistance to *P. nodorum* infections is essential to minimize the fungal threat. Glume blotch infections result in shriveled low weight kernel production, with losses as high as 30 to 50 percent under optimal conditions for *P. nodorum*.

Genomic selection (GS) offers a promising avenue for the improvement of quantitative traits, especially those difficult to improve through traditional breeding methods. GS is a statistical genomics tool that combines all molecular marker information for an individual to calculate genomic estimated breeding values (GEBVs) that can be used for advancement selections. GS models provide more comprehensive estimates of quantitative traits than marker-assisted selection, as it captures small and large effect loci contributing to the phenotype. The implementation of GS models permits the prediction of an individual's performance even before phenotyping has occurred. The utilization of GS models has the potential to accelerate the improvement of quantitative traits, including those that are difficult to phenotype, that are

measured on an irregular basis, or those that are not assessed until late stages of development, in the breeding of wheat varieties.

In this research, genotypic data already available for a panel of soft red winter wheat breeding lines representative of the University of Illinois' breeding program was leveraged by collecting phenotypic data on glume blotch resistance and several milling and baking quality traits. Glume blotch resistance and milling and baking quality traits are known to be quantitative in nature. The objective was to determine if genomic selection could be used to select for these quantitative traits. Glume blotch resistance is often difficult to phenotype, and milling and baking quality parameters usually are not evaluated until a breeding line has been assessed agronomically for several years. As such these traits are attractive targets for genomic selection.

ACKNOWLEDGEMENTS

Thank you to my advisor, Dr. Frederic Kolb, for taking a last minute chance on a girl from Kentucky. You have provided me with many opportunities. To Dr. David van Sanford for reaching out to a student who was interested in planting breeding but didn't know where to turn. To Dr. Sharyn Perry for building many bridges leading to my future. Dr. Santiago Mideros, thank you for allowing me to take over much of your lab and your assistance as needed. To Dr. Carrie Butts-Wilmsmeyer, I was more than lost but every step of the way your patience, pure love of statistics, and chats about Kentucky brought understanding and relief. Your help has been essential to the completion of this research. To the Small Grains team, thank you, for all the assistance in accomplishing this research, on the ground support makes the world of difference. Deanna and Melissa thank you for comedic relief, encouragement, insight into issues, and advice. To all of the amazing lab techs I worked with as an undergraduate, especially Joe Kupper, your friendship, guidance, and positive nature has meant so much to me. To Shari Dutton, thank you for sharing your love of horticulture, friendship, and advice. To my family, especially my parents, Tom and Annette Jones, thank you for all the encouragement and belief in me even when I couldn't believe in myself. Thank you to my Dad for instilling in me his passion for agriculture and taking care of me even 285 miles away. That change over pipe was critical to my success the first year. Thank you to my Mom who instilled the importance of education and passion for learning, always providing an ear and advice and reminding me why I moved so far from home. Lastly, to Elias for calming the storm of stress a few too many times, keeping my light ablaze, and your continued support as I move to the next chapter.

TABLE OF CONTENTS

CHAPTER 1: LITERATURE REVIEW

Wheat

Wheat (*Triticum aestivum* L.) is the most widely cultivated crop based on acreage globally. In 2014 730 million tons were harvested on 222 million hectares (IWGSC, 2018b). While wheat remains the third most cultivated crop in Illinois, total acreage has been in decline and has dropped below a tenth of the state corn acreage with 500,000 acres of wheat planted in Illinois for the 2017 season (Illinois Wheat Association, 2016; Nafziger, 2009; USDA, 2018b). As a globally important crop, wheat is a staple source of protein, vitamins, and minerals and supplies 20 percent of human caloric intake annually (Brenchley et al., 2012; IWGSC, 2018b). Although many classes of wheat are grown in the United States, each with unique characteristics and end uses, the primary class of wheat grown in Illinois is soft red winter wheat (SRWW). SRWW is characterized by lower gluten strength and overall protein, low starch damage and low water absorption as compared to hard wheat (Cabrera et al., 2015). Flour produced from SRWW typically is used to make cookies, pastries, cakes, cereals, biscuits, crackers, flat breads and is blended into all-purpose flour (Illinois Wheat Association, 2016).

Wheat is an allohexaploid with A, B, and D genomes. The diploid AA (*Triticum urartu*) and BB (relative of *Aegilops speltoides*) genomes first combined. Then, the newly formed tetraploid AABB genome species, known as modern durum wheat (*Triticum turgidum*), combined with the diploid DD (*Aegilops tauschii*) genome, giving rise to the hexaploid AABBDD modern wheat (*Triticum aestivum* L) (IWGSC, 2017). The wheat genome has 2n=6X=42 chromosomes, with 21 homeologous chromosome pairs and 7 pairs coming from each of the 3 progenitor species. Wheat has a total genome size of 17 gigabase pairs (Gb) and possesses 110, 000 to 150, 000 genes (IWGSC, 2018a). For comparison, the genome size of rice

is 389 megabase (Mb), or 0.389 Gb pairs (International Rice Genome Sequencing Project, 2005), and the corn genome is 2,500 Mb or 2.5 Gb pairs (Plant Genome and Systems Biology, 2018).

Importance of *Parastagonospora nodorum* glume blotch

Fungal pathogens are one of the largest threats to the quality and quantity of wheat production. Glume blotch is caused by the necrotrophic fungus *Parastagonospora nodorum* (syn. *Stagonospora nodorum*, *Septoria nodorum*) and can result in significant agronomic and economic losses to the wheat industry (Francki, 2013; Murray and Brennan, 2009). *P. nodorum* was first described in 1845 as a wheat pathogen and can infect all above ground tissues of its host plant (Baker, 1978; Ficke et al., 2018). Under optimal conditions for pathogen infection and disease progression, glume blotch can result in losses upwards of 30 to 50 percent with the most significant losses occurring when leaves and glumes are infected before grain fill, but plants are susceptible to *P. nodorum* at all growth stages (Bertucci et al., 2014; Bostwick et al., 1993; Fried and Meister, 1987; Kim et al., 2004; Kleczewski, 2015; Leath et al., 1993; Mehra et al., 2016). In 1990, a questionnaire based disease survey by Leath et al., (1993), described *P. nodorum*'s presence in Canada, the United States, Mexico, Central America, South America, China, Japan, Africa, Russia, Denmark, India, Australia, and New Zealand thereby, illustrating the global presence of the disease (Ficke et al., 2018). A 2009 review by Murray and Brennan projected current annual economic losses as high as $108 million, and potential cost as high as $230 million for *P. nodorum* in Australia alone (Ficke et al., 2018).

P. nodorum infections can result in leaf blotch and glume blotch. Leaf blotch reduces the photosynthetic leaf area, subsequently reducing grain fill. Glume blotch is also known to cause yield losses due to shriveled, low weight kernels. The reduction of photosynthetic leaf area has been proposed to be the leading factor in reduced yield quantity and quality, as the flag leaf is

known to provide much of the energy required for grain filling (Ficke et al., 2018; Krupinsky et al., 1973). However, glume blotch infections impact the kernel development as the glumes and awns are directly attacked (Ficke et al., 2018). Various studies on the photosynthetically active leaf area of the flag leaf and glume tissue after flowering have indicated the heads and flag leaves are the largest contributors to grain yield during the grain filling period (Ficke et al., 2018; Krupinsky et al., 1973). Glume infections occurring during the grain filling stage reduce the plant's ability to fill out the grain. The light weight shriveled kernels produced reduce the test weight and overall crop yield and according to Leath et al. (1993) is the component of yield most often affected by *P. nodorum*. The protection and increased resistance of the glume tissues are essential to preserving critical energy for grain filling.

Disease cycle

P. nodorum is a necrotrophic fungus with a polycyclic disease cycle in wheat and related grass species. After infecting a host plant, a necrotrophic fungus kills the infected tissue and feeds from its organic compounds to support the fungal life cycle (Selter et al., 2014). *P. nodorum* has a long latent period (10-20 days), but its ability to infect plants from seedling to maturity and to produce secondary inoculum gives rise to the polycyclic infection behavior (Mehra et al., 2015; Mehra et al., 2017). The pathogen overwinters on wheat residue and stubble for up to three years. Fruiting bodies form on the stubble, producing pycnidiospores. The pycnidia are burst open by rainfall and splash onto the tender plant surface, initiating seedling infection in the fall. The fungus then colonizes the leaf tissue directly through epidermal tissues or through openings such as the stomata (Francki, 2013), producing more pycnidia that burst and disperse the spores vertically up the plant. The spores progress up the canopy throughout the growing season. Optimal environmental conditions for the pathogen occur in temperate growing regions at 18-24 degrees Celsius along with 42-96 hours of high relative humidity (>75%) and

rainfall (Francki, 2013; Mehra et al., 2015; Wainshibaum and Lipps, 1991). Leaf blotch infections of *P. nodorum* cause brown lens shaped lesions with yellow halos on the leaves. Glume blotch infections of *P. nodorum* are characterized by brown-purple blotches beginning at the spikelet tips and moving inward to the base of the glume. When optimal conditions for the pathogen occur at heading and flowering, glume infections are more prevalent. A generalized disease cycle for *P. nodorum* can be found in Figure 1.1.

In a field experiment carried out by Griffiths and Ao (1976), the upward dispersal of spores was assessed using funnel traps placed 0.5 and 1.0 meter from artificially inoculated plants at varying heights from the ground (0, 40, 80, and 120 cm) and at varying angles around the plant. Their work demonstrated that the pathogen was able to splash 0.5 meters away from an infected source, but was not able to splash 1.0 meter from the source. Spores were only found in the traps placed at 0 and 40 cm above the ground, with equal amounts of spores represented in all traps at 0 cm regardless of the angle placement, and varying amounts of spores in the 40 cm traps depending on the angle placement around the infected source. The effect of angle placement at 40 cm demonstrates the influence of wind at higher altitudes in respect to spore dispersal. Griffiths and Ao (1976) further demonstrated how severe losses can occur given a sufficient number of infected plants, even though the spores tend to splash within a 0.5 meter radius.

P. nodorum management

Current methods of *P. nodorum* control include cultural practices, chemical controls, and host resistance. Cultural practices include tillage to incorporate debris in order to reduce primary inoculum, rotation to non-host crops such as corn, soybean, or vegetables, the avoidance of high plant populations, and the avoidance of overhead irrigation that increases the optimal conditions for the disease (Ficke et al., 2018; Francki, 2013; Kleczewski, 2015; Mehra et al., 2015; Murray and Brennan, 2009). Von Wechmar (1966) demonstrated that wheat residues buried 6 and 10

inches deep in the soil do not produce viable pycnidia after one month. Mehra et al. (2015) established correlations between residue coverage and *P. nodorum* severity with severity leveling off above 30 percent residue coverage (Ficke et al., 2018).

Fungicides are common chemical controls used especially to protect the head and flag leaf at the critical times from heading to maturity. Commonly used fungicides include Prosaro, Tilt, and Caramba, all of which are demethylation inhibitors (DMI). Prosaro is produced by Bayer Crop Science, and the active ingredient is tebuconazole. Tilt is a Syngenta product with propiconazole as the active ingredient. Caramba is produced by BASF Ag Products with metconazole as the active ingredient. DMI fungicides are upward systemic that inhibit ergosterol synthesis in fungi by affecting the cell wall and cell membrane. In regions or seasons with especially high disease pressure, chemical controls are also used throughout the season to offer increased whole plant protection. Various studies have tried to establish a disease threshold for recommended application of chemical controls against *P. nodorum* (Ficke et al., 2018). The threshold for which chemical controls become gainful depends on many factors including yield potential, the presence of other diseases, host resistance and wheat prices (Ficke et al., 2018).

Host resistance is the plant's natural ability to overcome or withstand the pathogen. The planting of wheat varieties with higher levels of host resistance to *P. nodorum* provide producers with an initial defense that may be enough to divert the need for more costly chemical controls or more time intensive cultural practices (Francki, 2013).

Wheat resistance to *P. nodorum*

Host resistance has been determined to be mostly quantitative, and no complete resistance mechanism is known in wheat (Bostwisk et al., 1993; Eyal et al., 1987; Ficke et al., 2018; Fried and Mesiter, 1987). Multiple studies have found no correlation between leaf resistance and glume resistance against *P. nodorum* in wheat (Baker, 1978; Bostwick et al.,

1993; Eyal et al., 1987; Fried and Meister, 1987). This suggests that resistance in leaves and glumes are controlled by different genes or differential expression of the same genes (Bostwick et al., 1993; Fried and Meister, 1987). Detached leaf assays have been commonly used to assess *P. nodorum* resistance. Detached leaf assays measure the ability of the plant to withstand the leaf blotch colonization of *P. nodorum*. Less progress has been made in the assessment of glume blotch resistance. To evaluate glume blotch resistance, the common practices of screening seedlings and using detached leaf assays cannot be employed due to the lack of leaf and glume resistance correlation. Some studies have shown an increase in host susceptibility at physiological maturity (Feekes 10.0 and on) for genotypes displaying resistance at the seedling stage (Francki, 2013; Wainshilbaum and Lipps, 1991). Numerous traits may contribute to a plant's ability to withstand *P. nodorum* infections including "initial infection, limiting lesion expansion, reducing sporulation in lesions, insensitivity to necrotrophic effectors, and lengthening the latent period" (Ficke et al., 2018).

Breeding for *P. nodorum*

Progress in the development of *P. nodorum* glume blotch resistant lines has been slow as breeders commonly depend on natural epidemics for selection of resistance (Cowger and Murphy, 2007; Francki, 2013; Fraser et al., 2003; Krupinsky et al., 1973). The lack of correlation between foliar and glume resistance has not only made the development of early screening techniques difficult but has subsequently increased the difficulty of breeding for resistance on a whole plant level (Ficke et al., 2018; Fried and Meister, 1987). According to Francki (2013), the selection for moderately resistant genotypes is possible when the disease is severe, but advances have been largely hindered by low disease pressure years. To overcome the inconsistency of natural epidemics, Cowger and Murphy (2007) performed a series of field trials to assess the efficacy of artificial inoculation methods for cultivar discrimination to glume blotch. The use of

artificial inoculations has the potential to accelerate the development of wheat lines resistant to *P. nodorum* glume blotch. According to Cowger and Murphy (2007), the most effective field artificial inoculation methods are: applying infected wheat straw between the plots in the late fall or early winter and applying a conidia spray in the spring at heading. Cowger and Silva-Rojas (2006) showed classification of wheat cultivars with resistance to *P. nodorum* would greatly benefit from natural infections within target growing regions or the use of artificial inoculation with a mixture of genetically diverse, but recent, pathogen isolates (Francki, 2013). The reliance on natural infection has also been called into question due to the increased risk of misclassification of escapes as the result of the absence of optimal disease environments for some lines (Francki, 2013).

Plants being used for glume blotch resistance evaluation must be allowed to reach full maturity. Allowing plants to reach full maturity increases the time required for evaluation and adds constraints to screening. Winter wheat must be planted to allow for early growth before entering into the cold winter period in a natural field setting or in a greenhouse setting with vernalization for eight to ten weeks. The cold period is required before winter wheat receives the biological signals to allow reproductive growth to begin. After completion of the required cold period, winter wheat requires ten more weeks to reach maturity. When completing field trials, planning must begin at least six months prior to evaluation. In the greenhouse, four months must be allowed for the plants to reach the desired growth stage, resulting in a limited ability to screen multiple cycles in a year. The lack of correlation in foliar and glume resistance suggest the combining of independent genes controlling the resistance of both would provide the highest degree of plant protection (Francki, 2013). In summary, the community agrees upon the need of reliable disease evaluations in target regions in a regular manner for continued breeding progress.

Milling and baking quality parameters and importance

Milling and baking quality parameters assess the acceptability of a wheat variety to end users – millers and bakers. For soft wheats, desirable milling and baking quality parameters include low to medium protein content, high break flour yield, fine flour granulation, and a high degree of kernel softness (Finney, 1990; McKendry et al., 1995). Flour yield is a milling quality parameter measuring the percent of flour recovery from an initial sample of grain tempered to 15 percent moisture (Schuler et al., 1995; USDA, 2018). Flour yield is calculated by dividing the combined recovery weight of mids and break flour by the initial tempered grain weight (USDA, 2018a). Mids are the medium sized flour particles caught above a 94-mesh screen when the milled product from the Quadrumat break roll unit is sifted (USDA, 2018a). Break flour consists of the finest flour particles that pass through the 94-mesh screen when the milled product is sifted. Several factors influence flour yield including grain size, shape, the thickness of the bran coat affecting the proportion of endosperm in the kernel, kernel hardness, cell wall thickness, and expertise of the miller (Schuler et al., 1995). Flour yield provides the millers with information on the potential profitability of the milled product if they purchase the variety from the grain elevator.

Softness equivalence is both a milling and baking quality parameter. The results of softness equivalence affects the way millers and bakers accept a wheat variety. Softness equivalence is a measure of break flour (finest flour particle size, passing through a 94-mesh screen) as a percentage of total flour (mids and break flour weight) (USDA, 2018a). The softness equivalence parameter provides an estimate of grain hardness, and flour particle size to the millers, who then market the flour to the bakers for use in specific end products.

Lactic acid solvent retention capacity (SRC) is a baking quality that provides bakers with an estimate of flour gluten strength for commercial baking performance (Souza et al., 2012; USDA, 2018a). Gluten strength directly impacts the baked products that can be created with the milled product. Higher or stronger gluten strength allows the bakers to make products with reduced spreading and increased rising, such as breads. Lower or weaker gluten strength (like that of soft wheats) allows bakers to make products with increased spread and reduced rising, such as cookies, pastries, and crackers. Flour protein concentration and the formation of viscoelastic networks are primary elements in the determination of gluten strength (Souza et al., 2012). In soft wheats, typical values for weak protein or gluten are below 85 percent, and strong protein or gluten values are above 110 percent (USDA, 2018a).

Breeding for milling and baking

In soft wheat, quality parameters are quantitatively controlled by both major genes and genes of small effect with moderate to high heritability (Cabrera et al., 2015; McKendry et al., 1995; Souza et al., 2012). Numerous QTL analyses have identified important chromosomal regions for milling and baking in wheat (Souza et al., 2012). Flour yield, softness equivalence, and the results of SRC evaluations display the "highly heritable, repeatable, and reliable" nature of milling and baking quality in soft wheat (Souza et al., 2012). Breeders must provide producers with agronomically improved varieties while maintaining superior milling and baking quality for the end users (Schuler et al., 1995). While selecting for improvement of agronomic traits, such as grain yield and disease resistance, wheat breeders have efficiently and simultaneously selected toward a quality standard (Souza et al., 2012). Due to large genetic variation for milling and baking traits and relatively small genotype-by-environment interaction for most traits, Souza et al. (2012) claim improvement selection in soft wheats should be highly effective. The known

high heritability, small genotype-by-environment interactions, and polygenic nature of milling and baking traits create the ideal scenario for GS for trait improvement (Heffner et al., 2011; Souza et al., 2012). Having the ability to assess milling and baking quality parameters earlier in a breeding program than during the traditional F_6 or F_7 stage would reduce the cost of evaluating additional breeding lines with unfavorable end use quality. Time spent on the development of wheat varieties not possessing superior end use quality detracts from potential forward progress. The development of GS models for milling and baking quality parameters holds the potential to eliminate inferior lines much earlier in the breeding program, when it is still unfeasible to perform milling and baking quality assessments due to cost, time constraints, the destructive nature of phenotyping, and the lack of sufficient amounts of seed (Heffner et al., 2011).

Genomic selection

GS is a statistical genomics tool first proposed in dairy livestock by Meuwissen et al. (2001). GS predicts the phenotypic response of individuals through the estimation of genomic estimated breeding values (GEBVs) calculated using genome-wide molecular effects (Meuwissen et al., 2001). GS is more effective than marker-assisted selection on quantitative traits as it captures small and large effect loci (Bernardo and Yu, 2007; Heffner et al., 2011; Huang et al., 2016; Meuwissen et al., 2001). Heffner et al. (2010) found a GEBV accuracy of 0.3 or greater in winter wheat provides a higher expected genetic gain per year than marker-assisted selection, even with relatively low heritability. GS breeding programs can achieve double the genetic gain per year experienced in a marker-assisted breeding program, even with moderate GEBV accuracies of 0.5 (Heffner et al. 2010). A 2011 study by Heffner et al. for grain quality traits in biparental wheat populations provided additional evidence that "genomic selection could produce greater genetic gains per unit time and cost than both phenotypic selection and

conventional marker-assisted selection in plant breeding" (Heffner et al., 2011). In GS, representative, randomly selected training populations containing known phenotypes and genotypes are used to train the model, and marker effect estimates are calculated (Heffner et al., 2009; Heffner et al., 2011; Meuwissen et al., 2001). The marker effect estimates from the representative training population are then used to calculate the GEBVs of individuals for which only genotypic information is available (Heffner et al., 2009; Heffner et al., 2011). Thus, GS is able to estimate the phenotypic performance of individuals before traditional phenotyping has occurred. GS can be used to aid in the improvement of quantitative traits with many small effect loci as it does not track specific quantitative trait loci (QTL) like other methods such as marker-assisted selection, but rather works to identify the best individuals in a population (Bernardo, 2016; Meuwissen et al., 2001). Although GS was first proposed in 2001 by Meuwissen et al. in dairy, the first GS study in plants was not published until 2007 by Bernardo and Yu. The first plant GS study was based on simulation data in maize, and Bernardo and Yu (2007) concluded GS was superior to marker-assisted recurrent selection. Since the first simulation study in plants, many studies have been published on the prediction accuracy for quantitative traits in numerous species (Bernardo, 2016). The work of Huang et al. (2016) supports the idea of the use of GS for improving values for quality and agronomic traits.

CHAPTER 2: GENOMIC SELECTION FOR GLUME BLOTCH IN SOFT RED WINTER WHEAT

Abstract

Glume blotch in wheat, caused by the necrotrophic fungus *Parastagonospora nodorum* (syn. *Stagonospora nodorum, Septoria nodorum*), is a leading fungal threat to wheat production in humid regions. Severe glume blotch infections reduce the quality and quantity of the grain through the production of shriveled low weight kernels. Breeding for resistant varieties offers an effective means to reduce the potentially devastating impact of *P. nodorum*. Genomic selection (GS) incorporates all marker information for an individual, regardless of the loci effect, to calculate genomic estimated breeding values (GEBVs) and to make advancement selections. GEBVs are the summed effects of all the loci and provide more comprehensive estimates of quantitative traits than marker-assisted selection. GS models allow for the prediction of breeding line performance before phenotyping. When utilized on quantitative traits, such as glume blotch resistance in wheat, GS has the potential to accelerate the improvement of wheat varieties with increased resistance. In this study, we investigated the potential of assessing the glume blotch resistance of germplasm at alternate times, the functionality of establishing and maintaining a glume blotch disease nursery, and the development of a GS model for improved resistance to glume blotch in the University of Illinois' breeding program. Results indicate the importance of taking disease evaluations at the proper time, a disease nursery for glume blotch can successfully be established, and supports the idea of using GS for improvement of quantitative traits, including glume blotch resistance.

Introduction

Glume blotch, caused by the necrotrophic fungus *Parastagonospora nodorum* (syn. *Stagonospora nodorum, Septoria nodorum*), results in significant agronomic and economic losses to wheat, especially in warm, humid environments. As a disease that can result in upwards of 30 to 50 percent yield losses under conditions favorable for disease progression, improving host resistance is a primary target for reducing its impact (Bertucci et al., 2014; Bostwick et al., 1993; Fried and Meister, 1987; Kim et al., 2004; Kleczewski, 2015; Mehra et al., 2016). Breeders have had limited ability to improve host resistance as evaluations are chiefly taken on an opportunistic basis when the disease naturally presents itself (Cowger and Murphy, 2007; Francki, 2013; Fraser et al., 2003; Krupinsky et al., 1973). In some areas of the soft red winter wheat growing regions of the United States, the disease does not naturally occur every growing season, therefore resulting in sporadic and incomplete assessments of the germplasm's resistance to *P. nodorum* (Cowger and Murphy, 2007). Inoculated nursery methods have been explored to allow for more consistent evaluation of germplasm resistance. Cowger and Murphy (2007) explored the options of conidia suspensions applied in early winter, conidia suspensions applied at boot stage in late spring, and infected straw applied in early winter. Cowger and Murphy (2007) found significantly more disease when conidia were applied at boot stage and when straw was applied in early winter. The establishment of a successful inoculated nursery could provide a breeder with regular evaluation of a program's germplasm, allowing for continual progress toward resistance. Many early studies using various methods to quantify resistance to *P. nodorum* determined resistance to be governed quantitatively with additive effect genes (Bostwick et al., 1993; Fried and Meister, 1987).

Genomic selection, first proposed by Meuwissen et al. (2001), could hold the key to improving resistance in breeding programs, especially those evaluating traits on an opportunistic basis. GS is a molecular breeding tool that uses genome-wide molecular marker effects to predict the phenotypic reaction of individuals through estimation of genomic estimated breeding values (GEBVs). GS models are trained using known genotypes and phenotypes in a training population (Meuwissen et al., 2001). Marker effects that have been calculated from the training population are then used to calculate the GEBVs of breeding lines where only genotypes are available (Heffner et al., 2009). GS has been found to be especially useful for improving quantitative traits characterized by many additive small effect loci (Meuwissen et al., 2001). Advantages for GS are larger for traits that are more difficult to improve through traditional breeding methods (Goddard and Hayes, 2007). Glume blotch resistance in wheat is a quantitative trait, and increasing resistance is a gradual process. The use of GS allows breeders to leverage the opportunistic evaluation of data in a manner that allows for more complete assessment of their program's germplasm. Opportunistic evaluations could be used to train GS models, thereby enabling the evaluation of breeding lines, even in years without disease through GEBVs. GS could also be implemented in a breeding program with a successful *P. nodorum* inoculated nursery. Instead of including all breeding lines in the inoculated nursery, a genetically representative subset could be evaluated, GEBVs calculated, and advancements selected (Heffner et al., 2009).

In this study, we compared ratings taken at alternate times to the "gold standard" peak progression rating, evaluated the usefulness of an inoculated irrigated disease nursery to accurately assess the resistance of breeding lines, and tested the capability of GS as a breeding tool in the University of Illinois' wheat breeding program to improve resistance to glume blotch.

Materials and Methods

Plant material and field design

In this experiment, 379 breeding lines were evaluated for resistance to *P. nodorum*, specifically to the glume blotch phase of infection. The original panel used in the study included 272 breeding lines selected by Arruda et al. (2015) to represent the genetic diversity of the University of Illinois' wheat breeding program. This panel is referred to as the MAM panel. The MAM panel consists of 185 lines from the University of Illinois' breeding program and 87 lines from 17 soft red winter wheat midwestern and eastern US breeding programs. The 272 line MAM panel represented the current breeding lines and lines commonly used as parents in the program. The arrangement of lines in the MAM panel captured the largest portion of genetic diversity in the University of Illinois' breeding program in a manner that was manageable for phenotyping. In addition to the MAM panel, 107 breeding lines chosen for advancement within the program for the 2016 season were included in the study. These lines were included to attempt to capitalize on collected genomic data within the program.

Wheat lines were planted in 1 meter long single rows at two locations for two years. Field experiments were grown at the University of Illinois' Maxwell Research Farm in Savoy, IL and at the University of Illinois' Brownstown Research Station in 2016, with planting dates of September 28 and October 2, respectively. In 2017, field experiments were grown at the University of Illinois' Research Farms South in Urbana, IL and Brown Seed Enterprises in Neoga, IL, with plating occurring September 23 and October 8, respectively. The 2016 field site at the University of Illinois' Maxwell Research Farm was a Drummer silty clay loam with 0-2 percent slope, while the 2017 field site at the University of Illinois' Research Farms South was an Elburn silt loam with 0-2 percent slope. In the natural pressure nurseries, different soil types

were represented in the experiments. The 2016 University of Illinois' Brownstown Research Station field site was a Cisne silt loam with 0-2 percent slope and the 2017 Brown Seed Enterprises' field site was a Bluford-Darmstadt silt loam with 0-2 percent slope. Each experiment was designed as a randomized complete block with two replications. All experiments received 0.0448 metric tons of liquid nitrogen per hectare pre-planting and did not receive a spring nitrogen application.

Disease nursery establishment

The rows grown at the Savoy and Urbana, IL locations were cultivated in a mist irrigated disease nursery, while the Brownstown and Neoga, IL locations were cultivated using natural growing conditions. In the mist irrigated nurseries, wheat straw was laid between the rows in early winter to provide a source of primary inoculum as discussed by Cowger and Murphy (2007). The wheat straw was laid between rows on December 17, 2015 and January 28, 2017 for the 2016 and 2017 experiments, respectively, and was allowed to naturally settle between rows. Pycnidiospore solution was used to inoculate the straw laid between the nursery rows in the early spring in an effort to increase the disease pressure. Straw inoculations occurred on April 20, 2016 using 8 liters of spore suspension and April 18, 2017 with 10 liters of spore suspension. A second round of spore solution was used to inoculate the wheat heads between Feekes 10.1 (heading) and Feekes 10.5 (flowering), to ensure fungal spores reached the glume tissue. The straw and the head inoculations were carried out using a CO_2 pressurized backpack sprayer using 40 psi in the solution tank on May 11 and May 16, 2016 and May 5 and May 11, 2017 for the two years respectively. Overhead mist irrigation ran three times daily for 30 minutes in 2016 and three times daily for 1 hour in 2017 for a period of 72 hours following inoculation of the wheat heads to encourage optimal infection. Inoculation of the wheat heads occurred approximately

one hour before the evening irrigation cycle at 6 pm. The irrigation ran at 9 am, just as dew was drying off, at 1 pm to maintain the humidity during the day, and at 6 pm to maintain moisture overnight. The mist irrigation system applied 0.3 centimeters of water per hour. Thus in 2016, 0.15 centimeters of water was applied three times per day and 0.3 centimeters of water was applied three times per day in 2017.

The two *P.* nodorum isolates used in the study were collected and isolated by Dr. Carl Bradley and his research team. Using the notation from the Bradley lab, the isolates used in this study were BF0994 and SN1464. Isolate BF0994 was collected and isolated in 2009 from Gallatin County in the southeast part of the state. Isolate SN1464 was collected and isolated in 2014 from Effingham County in the south-central part of the state. Thus, the two isolates were collected within the state's target wheat region and in different years. The mist irrigated nurseries were grown north of the target wheat areas for the state of Illinois resulting in lower natural disease pressure in the area. The wheat heads were inoculated to ensure ample disease pressure was available for glume infection. The Brownstown and Neoga, IL nurseries were grown within the target wheat growing region of the state, leading to higher natural disease pressure. It was reasonable to expect enough natural disease pressure for glume infections based on previous disease observations made by the small grains breeding program at the target region locations (Norman Smith and Frederic Kolb, Personal Communication, 2015).

Agar plugs containing the *P. nodorum* isolates, obtained from the -80 degree Celsius freezer, were placed on V8 agar plates and allowed to grow and sporulate under white light on a 12 hour light, 12 hour dark cycle at 20 degrees Celsius. After ample spore production occurred, usually between 14 to 18 days, spores were harvested from the V8 agar plates by flooding the plate with autoclaved distilled water and scraping the plate with a bent rod to release the spores

from the pycnidia, as described by Cowger and Murphy (2007), Bostwick et al. (1993), and Wainshilbaum et al. (1991). The water containing the first released spores was allowed to sit on the plate for 10 minutes to encourage more spore release before the plate was scraped a second time (Bostwick et al., 1993; Cowger and Murphy, 2007). After scraping the plate a second time, the water/spore mixture was pipetted up and down several times to ensure the spores were in solution. Then, the water and spore solution was pipetted from the plate into a labeled Falcon tube. All spore solutions were counted using a hemocytometer and adjusted to 10^6 spores per mL concentration (Cooke and Jones, 1970; Holmes and Colhoun, 1974; Jenkyn and King, 1977; Rooney and Hoad, 1989).

Phenotypic data

Disease ratings were taken on a 1 to 9 scale, with 1 being resistant or showing no symptoms and 9 being susceptible or highly infected. More specifically, breeding lines receiving a glume blotch rating of 1 did not display noticeable symptoms of glume blotch, and breeding lines receiving a rating of 9 had heads fully displaying glume blotch symptoms with the dark purple/brown markings covering all aspects of the glume Disease evaluations were performed by visually estimating the average severity of glume blotch infection throughout the row. Average severity estimates were performed by one evaluator to maintain consistency and reduce variability of potential scaling differences. Ratings were taken 27 and 34 days after Feekes 10.1 for each plot, wherein the inoculated nursery inoculations occurred between Feekes 10.1 and 10.5 (Wainshibaum et al., 1991). To obtain the best phenotypic rating of each line across environments, locations, and methods, best linear unbiased predictions (BLUPs) were calculated using PROC MIXED SAS version 9.4 (SAS Institute, 2017), according to the equation:

$$Y_{ijkl} = \mu + Y_i + L_{(i)j} + B_{(ij)k} + G_l + YG_{il} + LG_{(i)jl} + \varepsilon_{ijkl}$$

where Y_{ijkl} is the observed phenotype, μ is the overall grand mean, Y_i is the random effect of the ith year, $NID(0, \sigma_Y^2)$, $L_{(i)j}$ is the random effect of the jth location nested in the ith year, $NID(0, \sigma_L^2)$, $B_{(ij)k}$ is the random effect of the kth block nested in the jth location in the ith year, $NID(0, \sigma_B^2)$, G_l is the random effect of the lth genotype or wheat line, $NID(0, \sigma_G^2)$, YG_{il} is the random interaction of the ith year and the lth genotype, $NID(0, \sigma_{YG}^2)$, $LG_{(i)jl}$ is the random interaction of the jth location nested in the ith year and the lth genotype, $NID(0, \sigma_{LG}^2)$, and ε_{ijkl} is the random error term, $NID(0, \sigma_e^2)$. The overall mean ($\bar{y}_{\dots}$) of the peak progression rating was calculated using PROC MEANS in SAS 9.4 (SAS Institute, 2017). The overall mean and the solution for the lth genotype effect were summed to calculate the BLUP for each genotype in the study. The equation for the calculation of BLUPs is below.

$$BLUP_{\dots l} = \bar{y}_{\dots} + G_l$$

Genotypic data

The MAM panel of lines used in this study was originally designed by previous Doctoral candidate Marcio Arruda for his study on GS for *Fusarium* head blight in soft red winter wheat. Genotypic data on all the wheat lines used in this study were collected by the small grains research team using genotyping-by-sequencing (GBS) technology. The procedure used by the research team is outlined in Arruda et al. (2015). In brief, DNA was extracted from 5-day old leaves, GBS libraries were constructed, sequence data were obtained from a 96-plex Illumina HiSeq2000, SNPs were called, and the data were filtered for minor allele frequency. Missing data were imputed using the expectation maximization imputation (EMI) method due to its shorter computational time, higher accuracies for some traits, and its use in other wheat GS studies (Arruda et al., 2015). The genotype files used in this study contain 6,451 SNP markers that have been filtered for minor allele frequency and position.

Coefficient of variation and correlations

Coefficient of variation (CV) provides an assessment of data quality gathered within in a study. A CV is the ratio of the standard deviation (s) to the mean ($\bar{y}$), as shown in the equation:

$$CV = \frac{s}{\bar{y}} * 100$$

Lower CV's indicate higher quality data, as it indicates more of the variation is coming from genetic differences than from non-genetic differences such as human error or field placement. Higher CV's provide a warning that large portions of the variation occurred in non-genetic factors resulting in reduced ability to assess the differences among lines based on genetics. Lower CVs are favorable as it indicates genetic gains can be made. When working with disease ratings it is common for CV's to be between 20 and 30 percent, especially with ratings that are visually scored.

In this study, correlations were used to compare disease ratings taken early (27 days post Feekes 10.1), at peak progression (34 days post Feekes 10.1), and kernel quality (post-harvest). Correlations were calculated using PROC CORR of SAS version 9.4 (SAS Institute, 2017). The correlations assessed the ability of the early and post-harvest ratings to provide the same disease resistance information as the peak progression rating. It is known that peak progression ratings are ideal but is unknown if ratings at differing time points provide the same conclusions.

Genomic selection

PopVar is a statistical package created by Mohammadi et al. (2015) for R software (R Core Team, 2016) that predicts "standard statistics and correlated response in plant populations" and performs cross-validated GS with several model options (Tiede et al., 2015). In this study,

the ridge-regression best linear unbiased prediction (rrBLUP) model was used. Ridge-regression best linear unbiased prediction equally shrinks all markers toward zero in an infinitesimal model that sets markers as random effects with a common variance (Arruda, et al., 2015). PopVar allows the user to specify the number of folds and iterations to be used in the model creation.

Training and validation populations

A 5-fold cross validation was performed with the data using the PopVar package. The training and validation populations are not specified within the program, instead, the number of folds and the number of iterations are specified. Based on the number of folds specified, PopVar divides the phenotyped individuals. In the case of the 5-fold cross validation, PopVar divided the phenotyped lines into 5 groups. PopVar then used 4 of the 5 groups to train the model, reserving 1 group for validation. After the model was trained and validated based on the randomly assigned groups, PopVar randomly reassigned the phenotyped individuals to new groups and performed the procedure again. PopVar continued in this manner until the number of specified iterations was met. Upon completion of the specified iterations, PopVar assess the ability of the model to accurately predict the phenotypes of each validation set across each iteration and expressed the average accuracy (r) and a standard deviation (s) in the output file.

Heritability on a line mean basis

Heritability is assessed before creation of GS models. Traits with higher heritability have a higher success rate in GS studies, but lower heritable traits can also be improved by GS. Heritability on a line mean basis is the proportion of the total plant phenotypic variability contributed by the genotype. This type of heritability is similar to narrow sense heritability except it focuses on one line/genotype at a time. The equation for heritability on a line mean basis is as follows:

$$h^2 = \frac{\sigma_G^2}{\sigma_{\bar{p}}^2}$$

Where, h^2 = the proportion of variability contributed by the genotype

σ_G^2 = the mean genotypic variance

$\sigma_{\bar{p}}^2$ = the mean phenotypic variance

Genomic selection prediction accuracies

Prediction accuracies were measured by calculating the correlation between the GEBVs and PEBVs. PopVar performs this calculation in the background and provides the user with a "Results" file expressing model accuracies as r values with standard deviations. It has been shown r values greater than 0.3 in winter wheat, result in genetic gains when used in GS models (Heffner et al., 2010).

Results and Discussion

Phenotypic data

Disease ratings across years, locations, disease pressure type, and replications were combined using PROC MIXED in SAS 9.4 (SAS Institute, 2017) into one representative number referred to as a best linear unbiased prediction or BLUP. Each wheat line in the study possesses a unique BLUP indicating the best phenotypic estimate of the line.

Evaluation of rating time points

The CV for the three rating time points in the experiment can be found in Table 2.1. In disease studies, CVs between 20 and 30 percent are common for rating data. Data with a CV in this range are still useful for increasing disease resistance. The early disease rating had a CV of 23.6 percent, the peak progression rating a CV of 20.0 percent, and the kernel quality post-harvest rating a CV of 43.7 percent. Based on the CV of the early rating it may be possible to take disease ratings a week early to relieve the workload at more peak times in the program. The

peak progression ratings provide the best CV at 20.0 percent, but the early rating CV of 23.6 percent is within the range of acceptable CV values for disease studies. The kernel quality post-harvest rating CV of 43.7 percent indicates the post-harvest rating is influenced by more by environmental variation than the field ratings. The higher CV indicates the larger influence of human error in the given ratings. Based on the CVs alone, it is reasonable to conclude post-harvest ratings taken on the kernel quality of the grain are not as useful as the field ratings.

From the data quality assessment of CVs, it appears early ratings may offer an alternative to peak progression ratings because the early ratings and peak progression ratings are almost equally variable. Knowledge of the disease cycle and progression indicates the most appropriate time to take ratings is at peak progression. However, flexibility in regards to when ratings are taken could enable breeders to ensure assessments occur each year free from worry of Mother Nature's weather plans. The correlations between the two different rating time points were calculated to address this question. The correlations among all three time points are found in Table 2.2. Although the early rating CV indicates greater flexibility in rating time, the correlation of the early rating to the known peak progression rating is moderate at 0.52. Having only a 50 percent correlation to the known standard means the early rating is only providing 50 percent of the same resistance information. In order to comfortably use another rating in place of the known ideal, a correlation of 90 percent or higher would be required. The CV of the post-harvest kernel quality data of 43.7 indicated the data were not informative for increasing disease resistance and the correlation of 0.41 or 41 percent to the peak progression rating reinforces the idea that post-harvest rating is not a sufficient substitute. Thus, evaluation at peak progression should not be replaced with either proposed evaluation time point. The remaining analyses of the study were performed using the peak progression data.

Functionality of disease nursery

To assess the efficacy of the inoculated irrigated nursery against the natural pressure nursery, the covariance parameter estimates, as described in the materials and methods, phenotypic data section were investigated. The covariance parameter estimates are listed in Table 2.3. To assess the efficacy, the genotype-by-environment interaction has been broken into two main interactions, the genotype-by-year and genotype-by-location interactions. The examination of the genotype-by-year interaction provided a significant variance component estimate of 0.1193 (p-value = 0.0015). This indicates the genotypes changed ranks or that they had a different magnitude of response between years. Given that the two field seasons in the study, 2016 and 2017, were noticeability different, this is to be expected. The spring of 2016 was warm, humid, and rainy at the time of heading, creating optimal conditions for the pathogen. The spring of 2017 was cold, humid, and rainy at the time of heading, resulting in inadequate conditions for aggressive growth of the pathogen. *P. nodorum* has reduced infection capability at lower temperatures, as observed in 2017. The high significance of the genotype-by-year interaction reinforces the major role environmental conditions, such as temperature and humidity, play in disease progression.

The second interaction in the genotype-by-environment term, genotype-by-location, had a variance parameter estimate of 0 and a highly non-significant p-value to the point SAS does not provide a number (it's undefined but approaching 1). The estimate and p-value of the genotype-by-location interaction indicates the ranks of the genotypes did not change within the same year across the two locations. The major influencing element of genotype-by-environment term is the given year and the environmental conditions within that year at heading and flowering. The non-significance of the genotype-by-location interaction suggest a breeder can maintain a nursery at

either location and obtain equivalent disease evaluations. In the two years of this study, the inoculated irrigated disease nursery performed comparable to the target region natural disease pressure nursery. The genotype-by-location interaction showed the high efficacy of the irrigated nursery. Assuming that the same disease information can be obtained within the same year at either of the locations raises the question about the need for the extra inputs and man power to operate the inoculated irrigated nursery. Instead of investing additional resources to establish a *P. nodorum* nursery, disease evaluations could be taken at peak progression on plots already planted in the target wheat growing region of Illinois.

Heritability on a line mean basis

Heritability on a line mean basis is the proportion of variability contributed by the genotype. The variance parameter estimates are found in Table 2.3. The mean phenotypic variance ($\sigma_{\bar{p}}^2$) was calculated as 0.8936, and the genotypic variance (σ_G^2) was estimated to be 0.6256. Therefore, heritability on a line mean basis for glume blotch resistance is $h\text{^}2 = 0.6256/0.8936 = 0.7$ or 70 percent. This is a moderate heritability for a trait known to be highly quantitative with numerous small effect loci. This suggests GS could be a beneficial tool to aid in the improvement of increased glume blotch resistance in wheat.

Calculation of genomic estimated breeding values

PopVar provides a second output file titled *Genomic Estimated Breeding Values*, often shortened to GEBVs. GEBVs are estimated breeding values based on both genomic and phenotypic information as opposed to simple BLUPs. The GEBV output file contains the model's best estimate of the phenotypic value of each wheat line based on its sum of genetic marker effects as calculated from the training population. The model was trained on the provided phenotypes (phenotypically estimated breeding values or PEBVs) and the

accompanying genotypes of the phenotyped lines. GEBVs are the accumulation of all genetic effects within each genetically distinct individual which predicts the phenotypic behavior of that individual. GS and the calculation of GEBVs allow for prediction of phenotypes that have never been collected solely on the genetic marker effects of the line.

Breeding line comparison

Glume blotch resistance GEBVs were calculated for the 380 breeding lines evaluated in the study. The breeding lines were then ranked from resistant to susceptible based on GEBVs and PEBVs. When compared there is an overlap of 19 of 22 lines in the top 5 percent of resistant lines in the study. When comparing the top 10 percent 25 lines of 44 are found to overlap. The model is performing well enough to identify roughly the same lines based on GEBVs as we established by PEBVs in the field. On the opposite end of the scale comparing the bottom 5 percent of resistant lines, we find 17 of 22 lines overlap from the PEBVs to the GEBVs. Expanding to the bottom 10 percent of susceptible lines we find 31 of 46 overlap from the PEBVs to the GEBVs. Once again the model identified the least resistant breeding lines in a manner that is consistent with the PEBVs being obtained from the field disease evaluations. The comparisons of overlap for the top 5 percent, top 10 percent, bottom 5 percent, and bottom 10 percent can be found in Table 2.4, Table 2.5, Table 2.6, and Table 2.7, respectively.

Genomic selection prediction accuracies

The prediction accuracies and standard deviations were calculated by PopVar, and single values were output into a "Results" file. For the 5 runs shown the model accuracies ranged from 0.467 to 0.480 and the standard deviations ranged from 0.077 to 0.098, with an average accuracy and standard deviation of 0.471 and 0.087, respectively. The model accuracies and standard deviations can be found in Table 2.8. The prediction accuracies and standard deviations for all

runs are consistent, demonstrating the stability of the model no matter which lines are randomly assigned to the training and validation populations. In all cases, the prediction accuracies and the standard deviations are not in danger of dropping below the 0.3 prediction accuracy mark, which is illustrated by the central limit theorem in Figure 2.1. Other disease traits in the University of Illinois' wheat breeding program have been shown to have higher prediction accuracies by Arruda et al., 2015, but genetic gains are achievable using GS models with 47 percent accuracies. Based on the results reported here GS is an additional tool plant breeders can use to improve the resistance of wheat to *P. nodorum* glume blotch.

CHAPTER 3: GENOMIC SELECTION FOR MILLING AND BAKING QUALITY IN SOFT RED WINTER WHEAT

Abstract

Wheat (*Triticum aestivum L.*) is a globally important crop providing 20 percent of human caloric intake yearly, ordinarily in the form of flour. Milling and baking qualities are primary breeding objectives in wheat breeding programs as the end use quality directly impacts the acceptance of new wheat varieties. Milling and baking qualities are traditionally assessed at later stages in a breeding program. Genomic selection (GS) combines all marker effect information for an individual into genomic estimated breeding values (GEBVs) to provide an estimate of the individual's performance before phenotyping. GEBVs are the summed effects of all the loci and provide a more comprehensive estimate of quantitative traits than marker-assisted selection. GS has the potential to accelerate the improvement of wheat quantitative traits such as milling and baking quality parameters. In this study, GS is explored as a method to assess potential milling and baking quality earlier in a breeding program to reduce resources used to evaluate unfavorable lines, including personnel time, field resources, and program funds. The three milling and baking quality parameters, flour yield, softness equivalence, and lactic acid solvent retention capacity, were assessed for their use in GS models in the University of Illinois' breeding program. The newly built milling and baking GS models support the idea of using GS to accelerate breeding line improvement and variety development.

Introduction

Milling and baking quality directly impacts the acceptance of new wheat varieties with end use consumers. Superior milling and baking quality ensures the functionality of the milled product for the millers and bakers purposes. Wheat breeders have a tightrope to walk between satisfying end users and providing producers with agronomically improved agronomic varieties (Schuler et al., 1995). The assessment of milling and baking quality is laborious, expensive, and destructive. Furthermore, the timing of phenotypic data availability slows selection for milling and baking quality, because only one research lab processes the samples for every breeder in the eastern soft wheat growing region of the United States (Heffner et al., 2011). Due to the constraints of testing milling and baking quality, breeders do not typically assess these qualities until the later stages of breeding line evaluation when line numbers have been reduced based on various agronomic traits, genotypes are more stable due to increased homozygosity from more rounds of self-pollination, and larger seed quantities are obtained (Heffner et al., 2011). By not assessing milling and baking characteristics until F_6 or F_7 in a breeding program there is increased potential of expending resources on breeding lines that will not be accepted in the market as a result of undesirable qualities. The potential to eliminate breeding lines possessing unfavorable milling and baking quality earlier in a breeding program would allow for funds to be redirected to developing varieties with acceptable milling and baking quality since resources wasted on inferior lines would be reduced.

Genomic selection (GS) offers an exciting opportunity to successfully re-direct program resources to the production of improved wheat varieties. GS is a statistical genomics tool, first proposed by Meuwissen et al. (2001) in dairy livestock, that predicts the phenotypic response of individuals through the estimation of genomic estimated breeding values (GEBVs) calculated

using genome-wide molecular effects (Meuwissen et al., 2001). GS is more effective on quantitative traits than marker-assisted selection (MAS) as it captures small effect loci in addition to large effect loci (Heffner et al., 2011; Meuwissen et al., 2001). Training populations containing known phenotypes and genotypes are used to train GS models, and marker effect estimates are calculated from the representative training population (Heffner et al., 2009; Heffner et al., 2011; Meuwissen et al., 2001). The newly calculated marker effect estimates for the training population are used to calculate the GEBVs of breeding lines for which only genotypic information is available (Heffner et al., 2009; Heffner et al., 2011).

GS has the ability to allow selection before traditional phenotyping has occurred. In the past, obtaining genomic information was often more difficult, expensive, and time consuming than collecting phenotypic information. Now, the roles have reversed and genomic information is readily available while phenotypic information has become the limiting factor. The use of a GS model could allow a breeder to gather phenotypic data on fewer breeding lines (that are genetically representative of the germplasm in question), use this information to train a GS model, and predict the response of other members of the breeding population (Heffner et al., 2009). This scheme has the potential to save labor, time, and monetary resources.

In this study, three milling and baking quality parameters were evaluated for their potential in GS models as a breeding tool in the University of Illinois' wheat breeding program. Flour yield, softness equivalence, and lactic acid solvent retention capacity are critical milling and baking quality parameters for end users. These milling and baking quality characteristics are known to be highly heritable and quantitatively inherited (Heffner et al., 2011), making them ideal candidates for potential use in GS.

Materials and Methods

Plant material and field design

In this experiment, 272 breeding lines were evaluated for milling and baking quality. The breeding lines were originally assembled into a panel by Arruda et al. (2015) to represent the University of Illinois' wheat breeding program for his studies on GS and association mapping for *Fusarium* Head Blight. The panel was designed to capture the largest portion of genetic diversity in the University of Illinois' breeding program. The panel referred to as the MAM panel contains 185 lines from the University of Illinois' wheat breeding program and 87 lines from 17 soft red winter wheat midwestern and eastern US breeding programs. The breeding lines in the MAM panel consisted of the current breeding lines in the University of Illinois' program and lines used as parents in the breeding program.

Wheat lines were planted in 1 meter long single rows in a randomized complete block design. Lines were grown in 2015 and 2016 in field plots at the University of Illinois' Maxwell Research Farm in Savoy, IL. Both experimental fields, 2015 and 2016, were grown in a Drummer silty clay loam with 0-2% slopes. The 2015 experiment was planted September 26, 2014, and the 2016 experiment was planted September 28, 2015. The 2016 experiment was grown in a conidia inoculated misted irrigated *Parastagonospora nodorum* glume blotch disease nursery. Each year the experimental plots received 0.0448 metric tons of liquid nitrogen per hectare pre-planting and did not receive a spring nitrogen application. The entire 1 meter long row of each breeding line was hand harvested, processed through a Wintersteiger stationary thresher, and analyzed using the phenotyping protocols described below.

Phenotypic data

Harvested grain was analyzed by the USDA-ARS Soft Wheat Quality Lab (SWQL) at Ohio State University in Wooster, Ohio. Grain samples typically weighed between 100 and 200 grams, but a few samples had less than 100 grams of grain. Grain samples were extensively cleaned to remove diseased damaged kernels before analysis. The Soft Wheat Quality Lab uses a modified Quadrumat milling method for grain quality analysis. The moisture content of the grain was estimated using a Perten NIR DA7200 whole grain analyzer, and the grain was then tempered to 15 percent moisture before milling. During tempering, grain samples were placed in glass jars with distilled water, sealed, and tumbled on a chain driven conveyor for approximately 30 minutes or until all the water was absorbed (USDA, 2018a). After tempering, grain samples were kept in the sealed jars at room temperature for a minimum of 24 hours to allow for consistent moisture content throughout the kernels (USDA, 2018a).

Tempered grains were fed through a Quadrumat break roll unit and passed through three sets of milling rolls, each allowing smaller particles to pass through to move from grain to flour (USDA, 2018a). After milling, the product was sieved on a Great Western sifter box with sequential 40- and 94-mesh stainless steel bolting cloth screens. The 40-mesh screen has 471 micron openings, and the 94-mesh screen has 180 micron openings, used "to separate the milled product into three factions: bran, mids, and break flour" (USDA, 2018a). Bran is the largest particle size and is collected above the 40-mesh screen, mids are medium particle size and harvested above the 94-mesh screen, and break flour passes through the 94-mesh screen (USDA, 2018a). The calculation of total flour is the weight of the break flour (passed through the 94-mesh) and the mids (caught by the 94-mesh). The total flour is then used to calculate flour yield in the equation:

$$\frac{total\ flour}{grain\ weight}\ x\ 100\%$$

where grain weight is the weight of the tempered, whole grain sample. Higher values for flour yield are desirable as it is a direct measure of total recoverable flour from wheat kernels during the milling process. Higher flour yields translate to higher volumes of flour produced from the same size grain sample.

Softness equivalence, the second trait of interest, estimates grain softness and flour particle size. Softness equivalence is calculated by the equation:

$$\frac{break\ flour}{total\ flour}\ x\ 100\%$$

where break flour is the weight of the break flour or the milled product passing through the 94-mesh screen. Total flour is determined as outlined above. Higher values for softness equivalence indicate a larger percentage of the total flour yield is break flour or the smallest flour particle size which identifies higher quality flour for baking.

The third quality trait is lactic acid solvent retention capacity (SRC) which is a measure of the "percentage of solvent retained by a flour/solvent slurry after centrifugation and draining" (USDA, 2018a; Cabrera et al., 2015). To obtain all fractions used in SRC test the mids are passed through the Quadrumat reduction roll unit to produce shorts and reduction flour (USDA, 2018a). The shorts and reduction flour are separated by an 84-mesh screen (213 micron openings) (USDA, 2018a). The break flour, which passed through the 94-mesh screen, and the reduction flour are blended into straight grade flour for use in SRC test (USDA, 2018a). The lactic acid SRC test is a 1 gram test using the straight grade flour and provides a measure of gluten strength with higher lactic acid SRC values indicating greater flour gluten strength, which directly influences potential future uses. Higher flour gluten strength results in increased

elasticity of the milled product leading to preferential baking uses based on the flours ability to stretch. Breads are traditionally baked using hard wheat flour which contains much higher gluten content allowing the bread to rise and stretch. Cookies, pastries, and biscuits are traditionally baked using soft wheat flour because of its weaker gluten strength resulting in flatter baked goods. In soft wheat varieties values below 85 percent indicate weak protein and values above 110 percent indicate strong protein. Lactic acid SRC is calculated according to the equation:

$$\left(\left(\frac{residue\ weight}{flour\ weight} - 1\right)\ x\ \left(\frac{86}{100 - fmoist}\right)\right)\ x\ 100\%$$

where residue weight is equal to the weight of drained, saturated flour. Flour weight is the weight of the dry flour and fmoist is the percent moisture of wheat flour estimated by Unity NIR.

The Quadrumat break roll milling unit and its milling process are shown in Figure 3.1. The figure is from the USDA-ARS SWQL Materials and Methods 2017, Figure 1 (USDA, 2018a). To reduce confounding, controlled temperature and humidity was used for milling of all samples. The temperature was held at 19-21 degrees Celsius and relative humidity of 58-62 percent (USDA, 2018a). Prior to milling samples, the mill was allowed to run empty to equilibrate the mill temperature to 33± 1.0 degree Celsius (USDA, 2018a).

The raw data from the milling and baking test were provided to the University of Illinois' research group by the USDA-ARS SWQL, and the data were further analyzed by the small grains breeding group. To obtain the best phenotypic value of each breeding line for the three quality traits best linear unbiased predictions (BLUPs) were calculated using PROC MIXED in SAS version 9.4 (SAS Institute, 2017). BLUPs were used to summarize the quality data of each line to obtain unbiased predictions across years and environments. The overall mean ($\bar{y}_{..}$) of the trait of interest and the breeding line difference from the overall mean(g_j) were calculated in PROC MIXED according to the equation:

$$Y_{ij} = \mu + Y_i + G_j + \varepsilon_{ij}$$

where Y_{ij} is the observed phenotype, μ is the overall grand mean, Y_i is the random effect of the ith year, $NID(0, \sigma_Y^2)$, G_j is the random effect of the jth genotype or wheat line, $NID(0, \sigma_G^2)$, and ε_{ij} is the random error term, $NID(0, \sigma_e^2)$. The overall mean and the breeding line difference were combined to calculate the BLUP for each genotype. The BLUP calculation equation is below.

$$BLUP_{.j} = \bar{y}_{..} + G_j$$

Genotypic data

The panel of lines used in this study was originally assembled by previous Doctoral candidate Marcio Arruda, and the genotypic files were prepared as described by Arruda et al. (2015) for his study on GS for *Fusarium* head blight in wheat. Genotypic data were gathered using genotyping-by-sequencing (GBS) technology by the University of Illinois' small grains research team. The genotyping procedure is outlined in Arruda et al. (2015). In brief, DNA was extracted from 5-day old leaves, GBS libraries were constructed, sequence data were obtained from a 96-plex Illumina HiSeq2000, SNPs were called, and the data were filtered for minor allele frequency. Missing data were imputed using the expectation maximization imputation (EMI) method due to its shorter computational time, higher accuracies for some traits, and its use in other wheat GS studies (Arruda et al., 2015). The genotype files contain 6,451 SNP markers that have been filtered for minor allele frequency and position.

Coefficient of variation

Coefficient of variation (CV) is a measure of data quality within a study, commonly used by plant breeders to assess the proportion of the data variation from genetic and non-genetic sources. A CV is the ratio of the standard deviation (s) to the mean($\bar{y}$), as shown in the equation:

$$CV = \frac{s}{\bar{y}}\ x\ 100\%$$

Higher quality data are indicated by lower CVs, as it signals more variation is from genetic differences than from non-genetic differences such as field placement and human error. Higher CVs caution the reduced ability to assess genetic differences among lines due to larger variation in non-genetic factors. Lower CVs indicate genetic gains can be made when using the data. Milling and baking traits are considered to be stable with high heritability, meaning CV values should be less than 10.

Genomic selection

GS models were produced using two platforms for the three milling and baking traits. The GS model for flour yield was produced using the PopVar statistical package created by Mohammadi et al. (2015) for R software (R Core Team, 2016). To assess flour yield the ridge-regression best linear unbiased prediction (rrBLUP) model is used. In ridge-regression best linear unbiased prediction all markers are equally shrunken towards zero in an infinitesimal model that sets all markers as random effects with a common variance (Arruda et al., 2015; Heffner et al., 2011; Meuwissen et al., 2001). PopVar allows the user to determine the number of folds and iterations for the creation of the model.

GS models for softness equivalence and lactic acid solvent retention capacity were attempted using PopVar; however, PopVar's usefulness is sometimes limited due to pre-determined computational seeds and small variances resulting in faulty calculations of GEBVs. Instead, the two traits were assessed using rrBLUP (Endelman, 2011; Endelman and Jannik, 2012) as modified by Dr. Carolyn Butts-Wilmsmeyer and Dr. Alexander Lipka. The code used for evaluation of GS for softness equivalence and lactic acid solvent retention capacity (SRC) performs model cross validation as a five-fold scheme and calculates genomic estimated breeding values (GEBVs) on the full set of genotyped breeding lines. Based on our

understanding of PopVar, the modified rrBLUP code is performing in the same manner as PopVar except the computational seed is known to differ and a random permutation is used to select the lines used in the training and validation data sets during cross-validation. These slight computational amendments prohibit the small phenotypic variance of the traits from causing cause issues. The user maintains the ability to set the number of folds and desired iterations while performing the rrBLUP GS code.

Training and validation populations

A five-fold cross validation was performed on all three traits regardless of the platform used to perform GS. In PopVar the training and validation populations are not assigned by the researcher, rather the program uses a random number to divide breeding lines into groups based on the specified folds. PopVar does not notify the user which breeding lines are being used for training and which are being used for validation. In the case of a five-fold cross validation, PopVar divides the breeding lines into five groups. PopVar then uses four of the groups to train the GS model and reserves one for validation. Once the model is run based on the assigned groups, PopVar randomizes the phenotypes individuals again into five new groups and performs the procedure over until the desired number of iterations has been met.

In the rrBLUP code used to perform GS on softness equivalence and lactic acid SRC, a random number permutation assigns breeding lines to five groups, four groups for training and one group for validation. The permutation is designed to randomly re-sort the breeding lines from 1 to 272, and then they are divided into fifths based on the placement in the permuted order. The model is trained on the last four fifths of the newly sorted breeding lines, validated on the first fifth and model correlations calculated. Then, the model is trained on the first fifth and last three fifths, validated on the second fifth, and model correlations calculated. The model is cross-

validated in this manner until all fifths have been used as the validation population. The model performing in this manner is considered a five-fold cross-validation scheme. The random number permutation in the code can be accessed, enabling the training and validation populations to be viewed.

Calculation of genomic estimated breeding values

Genomic estimated breeding values (GEBVs) are the summation of the loci effects of each individual breeding line. GEBVs are the best estimate of the genetic value for each breeding line from the GS model, when the model has been trained using the provided population which has been theoretically selected due to its genetic representation of the breeding population in question. The training population provides phenotypes (phenotypically estimated breeding values or PEBVs) and genotypes for the GS model to learn the values of the genetic effects of the population. The validation population also has phenotypes and genotypes, but only genotypes are used in the model to predict the phenotypes to confirm the prediction accuracy of the model. After running the specified number of iterations PopVar calculates the GEBVs of all the breeding lines without regard to their presence in the training or validation populations. The GEBVs provided in the output file are calculated by PopVar as an average of the predicted GEBVs from all iterations of the GS model. GEBVs allow for phenotypes to be predicted that have never been observed based solely on the breeding lines genetic factors (Heffner et al., 2009; Heffner et al., 2011).

In the rrBLUP code used for GS of softness equivalence and lactic acid SRC, GEBVs were run in a separate step of the code than the model validation. In the GEBV calculation all breeding lines are considered at the same time in a single iteration of a ridge-regression model. A simple ridge-regression model is performed to calculate the GEBVs of all the breeding lines.

Ridge-regression provides a summed response for the genetic markers and their loci effects of each marker within a breeding line. In ridge-regression markers are treated as random effects with a common variance while concurrently estimating marker effects (Arruda et al., 2015; Heffner et al., 2011; Meuwissen et al., 2001).

Genomic selection prediction accuracies

Prediction accuracies were measured by calculating the correlation between the GEBVs and PEBVs. The correlation of the PEBVs and the GEBVs indicate the ability of the model to properly predict how the breeding lines perform and are a measure of model accuracy. After PopVar has met the desired iterations specified by the user, PopVar evaluates the ability of the model to accurately predict the phenotypes of each validation population in each iteration and provides the average correlation (r) and the standard deviation (s) in an output file. In the rrBLUP code, the ability of the model to accurately predict the phenotypes of each validation population (one-fifth at a time) is assessed at each fold. The correlation of the GEBVs and the PEBVs of all five turns are returned from the code and an average correlation (r) is calculated with its standard deviation (s).

Results and Discussion

Phenotypic data

The raw phenotypic values for the milling and baking qualities of interest were collected by the USDA-ARS SWQL in Wooster, Ohio and provided to the University of Illinois' small grains breeding program for analysis. The raw phenotypic values for the 2015 seed increase rows were received by the small grains program on May 26, 2016, and the raw values for the 2016 seed submission were received on May 15, 2017. The raw phenotypic values were calculated as described above in the materials and methods section.

The raw phenotypic values for each milling and baking quality parameter were combined using PROC MIXED in SAS 9.4 (SAS Institute, 2017) across the two years into one representative value referred to as a best linear unbiased prediction or BLUP. A unique BLUP was calculated for each breeding line and for each of the milling and baking quality parameters. Thus, one breeding line possesses three unique BLUPs in this study, one BLUP for flour yield, one BLUP for softness equivalence, and one BLUP for lactic acid SRC. The BLUPs of each trait are also considered phenotypic estimated breeding values (PEBVs) and are referred to as PEBVs throughout the rest of the chapter.

Coefficient of variation

The CV for the three milling and baking traits can be found in Table 3.1. Milling and baking traits are considered to be stable indicating a CV value should be less than 10 in order to use the phenotypic data comfortably. In this study, the CVs were calculated for the three milling and baking qualities across the two years. The CVs of flour yield, softness equivalence, and lactic acid SRC were 1.53, 4.02, and 5.96, respectively. All of the CVs were less than 10 indicating high quality data with the variation among breeding lines is primarily due to genetic differences and not non-genetic differences or human/machine error. The high quality CVs indicate the milling and baking traits in question could be ideal candidates for GS. The CVs are an indicator of the trait heritability and demonstrate the traits are highly heritable. Heritability calculations are often performed to provide an indication of GS's potential usefulness; however, in the case of this study heritability calculations are not appropriate as there is only one replication of the data in a single year. Thus, genotype-by-environment interactions cannot be assessed.

Calculation of genomic estimated breeding values

GEBVs were calculated for all breeding lines in the study for the three milling and baking traits of interest. The GEBVs for flour yield were calculated using PopVar, while the GEBVs for softness equivalence and lactic acid SRC were calculated using the rrBLUP set of code. The breeding lines were ranked from high to low response quality for each trait based on PEBVs and GEBVs. Breeders are interested in evaluating breeding lines at the top and bottom of the distribution of lines; therefore, comparisons of overlap between PEBVs and GEBVs for the top ten percent of flour yield, softness equivalence, and lactic acid SRC were made and can be found in Table 3.2, Table 3.3, and Table 3.4, respectively. When compared there is an overlap of 25 of 27 lines in the top ten percent of flour yield with change of rank occurring between the PEBVs and the GEBVs. Comparing the PEBVs and GEBVs for softness equivalence and lactic acid SRC, the top ten percent of lines are displayed perfect overlap with no change in rank for each trait. The phenotypic variability for softness equivalence and lactic acid SRC are being highly explained by the genotypic effects, resulting in only slight differences in the PEBV and GEBV trait values. The models are performing exceptionally well, identifying all or nearly all of the same lines that were identified for their superior quality based on phenotypic data. On the other end of the scale, the bottom ten percent of breeding lines for flour yield (Table 3.5), softness equivalence (Table 3.6), and lactic acid SRC (Table 3.7) we find 24 of 27 lines overlap, 27 of 27 lines overlap, and 27 of 27 lines overlap, respectively. In the bottom ten percent of all three traits, change in rank does occur to some degree. Comparisons of overlap between the bottom ten percent of PEBVs and GEBVs for flour yield, softness equivalence, and lactic acid SRC are found in Tables 3.5, 3.6, and 3.7, respectively.

Genomic selection prediction accuracies

After completion of the desired iterations, PopVar assesses the ability of the model to accurately predict the phenotypes of the validation set of all iterations and expresses the average accuracy (r) and the average standard deviation (s) in the output file. The accuracy of the model being expressed is a correlation of the GEBVs of the validation set with the PEBVs of the validation set. The correlations of model accuracy from five runs of PopVar based on flour yield are shown in Table 3.8. The average correlation from PopVar for flour yield based on five runs of the model is 0.547 with an average standard deviation of 0.091. The correlations of the model for the five runs range from 0.535 to 0.557, and the standard deviations ranged from 0.080 to 0.100. There are only small fluctuations in the accuracy and standard deviations demonstrating the stability of the model.

After completion of the five-fold cross validation, the second set of GS code gauges the accuracy of the model through a correlation of the GEBVs and PEBVs of the validation set in each fold. This code is calculating accuracies or correlations in the same manner as PopVar. The code is different from PopVar as it outputs the five correlation values (r_CV#), one for each fold, along with the average correlation (r) and standard deviation (s). By outputting all five correlation values the researcher can evaluate if the breeding lines are being broken up into representative groups through the random permutation. Table 3.9, Table 3.10, and Table 3.11 shows the model correlations as a measure of accuracy along with the standard deviation of each run for flour yield, softness equivalence, and lactic acid SRC, respectively. Five runs of the model were performed for each milling and baking trait. The values of each run and the average for each trait are shown. Although flour yield was successfully run through PopVar, it was also run through the second set of code. The average model accuracy and standard deviation for flour

yield from this set of code are 0.551 and 0.093, with a range of 0.515 to 0.593 for accuracy and a range of 0.041 to 0.179 for standard deviation. The model accuracy and standard deviation for flour yield coming from the second set of code are consistent with PopVar's results, demonstrating the two platforms are performing similarly. GS for softness equivalence was performed using the second set of code, providing an average accuracy of 0.512 with an average standard deviation of 0.093. The model correlations for softness equivalence for the five runs range from 0.499 to 0.540 with a range of standard deviation from 0.068 to 0.108. GS for lactic acid SRC resulted in an average model correlation of 0.594 with an average standard deviation of 0.066. The model correlation for the five runs of lactic acid SRC range from 0.579 to 0.617 with a range of standard deviations from 0.047 to 0.086. The prediction accuracies and standard deviations for flour yield, softness equivalence, and lactic acid solvent retention capacity for all iterations are consistent, establishing the stability of the model even with random assignment of lines to training and validation populations. The prediction accuracies and standard errors are not in danger of dropping below the 0.3 prediction mark needed to accomplish gains in winter wheat through GS (Heffner et al., 2010). This is illustrated by the central limit theorem in Figure 3.2 for flour yield, Figure 3.3 for softness equivalence, and Figure 3.4 for lactic acid SRC. Thus, based on these results the use of GS for flour yield, softness equivalence, and lactic acid solvent retention capacity is an addition to the breeder's tool belt for improving milling and baking qualities. Use of GS for these quality parameters should allow breeders to select for these traits at an earlier stage of variety development.

TABLES

Table 2.1. Coefficients of variation for the early, peak progression, and post-harvest rating time points for disease resistance on the glumes of 380 soft red winter wheat breeding lines.

Early Rating	**Peak Progression Rating**	**Post-Harvest Kernel Quality**
23.6	20.0	43.7

Table 2.2. Correlations of the three experimental rating time points for glume blotch resistance in soft red winter wheat to each other.

	Early Rating	Peak Progression Rating	Post-Harvest Kernel Quality
Early Rating	1	0.52	0.26
Peak Progression Rating		1	0.41
Post-Harvest Kernel Quality			1

Table 2.3. Variance parameter estimates for *P. nodorum* glume blotch resistance in soft red winter wheat.

Variance Parameter Estimates		
Variance Parameter	Estimate	P-value
Year	0.2245	0.2551
Location(Year)	0.0138	0.3849
Block(Year-by-Location)	0.03899	0.1012
Genotype	0.6256	<0.0001**
Year-by-Genotype	0.1193	0.0015**
Location-by-Genotype	0	NS
Error	1.6625	<0.0001**

** Highly significant at alpha 0.05
NS Highly not significant at alpha 0.05; no value given

Table 2.4. Top five percent of soft red winter wheat breeding lines for glume blotch resistance based on PEBVs and GEBVs. Breeding line names highlighted in yellow also appear on the other side of the table in the top five percent of lines based on PEBVs compared to GEBVs

Breeding line	PEBV	Breeding line	GEBV
IL13_28511	3.826	IL10_6876	4.450
MO050921	4.002	IL10_30320	4.548
IL13_7027	4.002	NY99045_3110	4.605
Truman	4.002	IL12_26448	4.610
IL12_26448	4.002	IL10_6855	4.611
IL06_9607	4.089	KY02C_3005_25	4.626
KY02C_3005_25	4.089	Truman	4.626
NY99045_3110	4.089	IL06_9607	4.658
IL10_30320	4.089	IL476_78S	4.693
IL10_6876	4.177	IL13_7027	4.713
IL13_25652	4.352	MO050921	4.730
IL12_8545	4.352	MO080864	4.732
IL06_7550	4.352	IL10_29373	4.764
IL476_78S	4.352	IL13_28511	4.777
IL06_16639	4.439	IL06_7550	4.804
IL13_27973	4.439	IL13_20918	4.817
IL10_32545	4.439	MO011174	4.840
Bromfield	4.439	IL10_8568	4.849
IL06_31053	4.439	IL10_6608	4.870
MO080864	4.439	IL10_19317	4.871
IL10_6855	4.439		

.

Table 2.5. Top ten percent PEBVs for soft red winter wheat breeding lines for glume blotch resistance compared to the top ten percent GEBVs. Breeding line names highlighted in yellow also appear on the other side of the table in the top ten percent of lines based on PEBVs compared to GEBVs.

Breeding line	PEBV	Breeding line	GEBV
IL13_28511	3.826	IL10_6876	4.450
MO050921	4.002	IL10_30320	4.548
IL13_7027	4.002	NY99045_3110	4.605
Truman	4.002	IL12_26448	4.610
IL12_26448	4.002	IL10_6855	4.611
IL06_9607	4.089	KY02C_3005_25	4.626
KY02C_3005_25	4.089	Truman	4.626
NY99045_3110	4.089	IL06_9607	4.658
IL10_30320	4.089	IL476_78S	4.693
IL10_6876	4.177	IL13_7027	4.713
IL13_25652	4.352	MO050921	4.730
IL12_8545	4.352	MO080864	4.732
IL06_7550	4.352	IL10_29373	4.764
IL476_78S	4.352	IL13_28511	4.777
IL06_16639	4.439	IL06_7550	4.804
IL13_27973	4.439	IL13_20918	4.817
IL10_32545	4.439	MO011174	4.840
Bromfield	4.439	IL10_8568	4.849
IL06_31053	4.439	IL10_6608	4.870
MO080864	4.439	IL10_19317	4.871
IL10_6855	4.439	IL10_29377	4.887
IL13_20918	4.527	IL14_26726	4.889
IL10_6608	4.527	IL12_28257	4.891
IL14_26726	4.527	IL10_6948	4.897
IL13_29257	4.527	IL12_36166	4.904
IL10_29373	4.615	IL12_8545	4.907
IL11_12437	4.615	IL11_12437	4.912
IL09_17057	4.615	IL11_5676	4.914
IL06_23571	4.615	IL13_29257	4.920
IL12_21235	4.615	IL13_6912	4.928
IL12_21624	4.615	IL06_31053	4.933
IL11_3466	4.702	Bromfield	4.934
IL08_8844	4.702	IL10_6912	4.938
Pio25R47	4.702	IL10_32545	4.948
IL10_24617	4.702	IL11_3466	4.949

Table 2.5. (Con't)

IL04_10741	4.702	IL13_7045	4.950
NC8170_45_2	4.702	IL08_8844	4.950
IL13_18570	4.702	IL09_17057	4.955
IL07_19062	4.702	IL11_963	4.957
MD04W249_11_7	4.702	IL13_2156	4.964
IL369_168R	4.702		
IL10_26814	4.702		
IL01_11445	4.702		

Table 2.6. The bottom five percent PEBVs for soft red winter wheat breeding lines for glume resistance compared to the bottom five percent GEBVs. Breeding line names highlighted in yellow also appear on the other side of the table in the bottom five percent of lines based on PEBVs compared to GEBVs.

Breeding line	PEBV	Breeding line	GEBV
ARS09_228	7.942	ARS09_228	7.163
IL07_14547	7.679	IL01_16170	6.987
Clark	7.504	P92829A1_1_1_3_3	6.969
P92829A1_1_1_3_3	7.329	Clark	6.933
NE06607	7.329	IL07_14547	6.817
IL01_16170	7.242	NE06607	6.672
IL10_28023	7.242	P0128A1_22_22	6.669
IL11_7978	7.154	IL11_7978	6.656
P0128A1_22_22	6.979	IL13_4504	6.532
IL13_4504	6.891	P0179A1_17	6.509
IL369_22S	6.891	IL10_28023	6.490
ARGE07_1374_17_8_5	6.891	P981359C1_4	6.391
IL97_6755	6.804	P03112A1_7_14	6.370
Rosette	6.804	IL369_22S	6.347
P0179A1_17	6.716	IL13_6459	6.346
P981359C1_4	6.716	Patterson	6.333
IL13_6459	6.716	IL97_6755	6.317
Patterson	6.716	IL10_14458	6.302
IL10_7871	6.716	GANC8248_12DH1	6.294
GANC8248_12DH1	6.629	IL11_15514	6.285
OK11311F	6.629		
Shirley	6.629		

Table 2.7. The bottom ten percent of PEBVs for soft red winter wheat breeding lines for glume blotch resistance compared to the bottom ten percent of GEBVs. Breeding line names highlighted in yellow also appear on the other side of the table in the bottom ten percent of lines based on PEBVs compared to GEBVs.

Breeding line	PEBV	Breeding line	GEBV
ARS09_228	7.942	ARS09_228	7.163
IL07_14547	7.679	IL01_16170	6.987
Clark	7.504	P92829A1_1_1_3_3	6.969
P92829A1_1_1_3_3	7.329	Clark	6.933
NE06607	7.329	IL07_14547	6.817
IL01_16170	7.242	NE06607	6.672
IL10_28023	7.242	P0128A1_22_22	6.669
IL11_7978	7.154	IL11_7978	6.656
P0128A1_22_22	6.979	IL13_4504	6.532
IL13_4504	6.891	P0179A1_17	6.509
IL369_22S	6.891	IL10_28023	6.490
ARGE07_1374_17_8_5	6.891	P981359C1_4	6.391
IL97_6755	6.804	P03112A1_7_14	6.370
Rosette	6.804	IL369_22S	6.347
P0179A1_17	6.716	IL13_6459	6.346
P981359C1_4	6.716	Patterson	6.333
IL13_6459	6.716	IL97_6755	6.317
Patterson	6.716	IL10_14458	6.302
IL10_7871	6.716	GANC8248_12DH1	6.294
GANC8248_12DH1	6.629	IL11_15514	6.285
OK11311F	6.629	IL476_191S	6.258
Shirley	6.629	Shirley	6.246
IL11_8144	6.541	P99840C4_8_3_1_12	6.230
IL08_32814_4	6.541	Caldwell	6.222
IL07_6861	6.541	Milton	6.203
Caldwell	6.541	IL476_11R	6.203
IL13_23870	6.453	G09418	6.162
IL13_19346	6.453	IL07_6861	6.161
IL10_17707	6.453	P0724B1_13	6.158
IL11_4620	6.453	Freedom	6.151
IL11_15514	6.453	IL99_26442	6.150
IL10_14458	6.453	IL11_4620	6.143
IL13_20375	6.366	P9975RA1_6_3_94	6.124
IL13_4164	6.366	Excel307	6.121
IL97_1828	6.366	Sisson	6.115

Table 2.7. (Con't)

IL99_26442	6.366	IL13_19330	6.094
Freedom	6.366	IL476_256R	6.094
P99840C4_8_3_1_12	6.366	IL08_32814_4	6.079
IL10_12079	6.278	KWS013	6.061
IL13_28906	6.278	LA07178C_44	6.055
IL13_5222	6.278		
LA07178C_44	6.278		
KWS013	6.278		
IL476_256R	6.278		
IL13_19330	6.278		
Excel307	6.278		

Table 2.8. Five-fold cross validated genomic selection prediction accuracies and standard deviations from five runs of the PopVar model for glume blotch resistance in soft red winter wheat.

Iteration	Accuracy (r)	Standard Deviation (s)
1	0.468	0.077
2	0.481	0.082
3	0.472	0.088
4	0.467	0.089
5	0.469	0.098
Average	0.471	0.087

Table 3.1. Coefficients of variation for flour yield, softness equivalence, and lactic acid solvent retention capacity of 272 soft red winter wheat breeding lines.

Flour Yield	Softness Equivalence	Lactic Acid Solvent Retention Capacity
1.53	4.02	5.96

Table 3.2. Top ten percent of soft red winter wheat breeding lines for flour yield based on PEBVs and GEBVs. Breeding lines appearing in both the top ten percent for flour yield based on PEBVs and GEBVs are highlighted in the same color to allow changes in rank to be easily identified. Breeding lines that are not highlighted (remaining white) do not appear in the top ten percent on the other side of the table.

Breeding line	FY_PEBV	Breeding line	FY_GEBV
IL08_28833	71.102	IL08_28833	70.317
IL11_20829	70.301	IL11_20829	69.750
IL00_8061	69.980	IL00_8641	69.677
IL00_8641	69.900	IL06_13721	69.535
IL06_13721	69.780	IL00_8061	69.466
NE06607	69.780	IL10_19203	69.352
IL06_7550	69.740	IL06_7550	69.345
OH08_206_69	69.660	IL10_8568	69.331
IL10_19203	69.620	IL00_8530	69.310
IL10_19317	69.580	IL10_19317	69.305
Excel307	69.539	IL10_33664w	69.283
IL10_8568	69.499	OH08_206_69	69.260
IL10_33664w	69.459	NE06607	69.255
IL00_8530	69.419	IL07_21847	69.214
IL06_18051	69.419	IL09_18352	69.162
VA09W_188WS	69.419	IL09_18372	69.140
IL09_18352	69.379	IL06_18051	69.124
Milton	69.339	Excel307	69.058
IL10_33378	69.259	IL10_19464	69.025
IL09_18372	69.219	IL10_33378	69.019
IL07_21847	69.179	VA09W_188WS	69.012
IL10_24617	69.179	IL10_24617	69.009
IL04_24668	69.139	IL10_35020	68.832
IL10_19464	69.019	IL04_24668	68.791
IL08_8844	68.898	IL08_8844	68.788
IL10_35020	68.858	IL10_7970	68.768
IL09_11630	68.858	IL10_29377	68.674

Table 3.3. Top ten percent of soft red winter wheat breeding line for softness equivalence based on PEBVs and GEBVs. Breeding lines appearing in both the top ten percent for softness equivalence based on PEBVs and GEBVs are highlighted in the same color to allow changes in rank to be easily identified. No changes in rank occurred for softness equivalence between the PEBVs and GEBVs. Only slight differences in calculated values are observed.

Breeding line	SE_PEBV	Breeding line	SE_GEBV
OH08_269_58	66.438	OH08_269_58	66.438
IL04_9942	65.905	IL04_9942	65.905
IL08_28833	65.417	IL08_28833	65.417
IL79_002T	64.928	IL79_002T	64.928
Excel307	64.928	Excel307	64.928
Caldwell	64.884	Caldwell	64.884
IL10_34967	64.884	IL10_34967	64.884
F0065	64.706	F0065	64.706
Pio25R47	64.484	Pio25R47	64.484
IL06_7653	64.218	IL06_7653	64.218
IL06_7550	64.129	IL06_7550	64.129
IL02_19463B	64.085	IL02_19463B	64.084
IL08_12206	64.040	IL08_12206	64.040
IL11_10272	63.952	IL11_10272	63.951
LCS19227	63.685	LCS19227	63.685
KY90C_054_39	63.685	KY90C_054_39	63.685
IL04_8445	63.596	IL04_8445	63.596
IL10_6948	63.508	IL10_6948	63.507
IL10_35020	63.374	IL10_35020	63.374
IL10_23808	63.374	IL10_23808	63.374
IL10_30364	63.330	IL10_30364	63.330
IL06_7034	63.152	IL06_7034	63.152
IL07_16075	62.975	IL07_16075	62.975
IL09_14063	62.975	IL09_14063	62.974
IL08_8844	62.886	IL08_8844	62.886
IL07_7527	62.886	IL07_7527	62.885
IL10_7970	62.886	IL10_7970	62.885

Table 3.4. Top ten percent of soft red winter wheat breeding lines for lactic acid solvent retention capacity based on PEBVs and GEBVs. Breeding lines appearing in both the top ten percent for lactic acid solvent retention capacity based on PEBVs and GEBVs are highlighted in the same color to allow changes in rank to be easily identified. No changes in rank occurred for lactic acid solvent retention capacity between the PEBVs and GEBVs. Only slight differences in calculated values are observed

Breeding line	LA_PEBV	Breeding line	LA_GEBV
ARGE07_1374_17_8_5	134.644	ARGE07_1374_17_8_5	134.642
IL10_841	131.145	IL10_841	131.143
IL10_32864	130.895	IL10_32864	130.893
IL11_5331	127.021	IL11_5331	127.020
IL11_7978	126.605	IL11_7978	126.604
ARS07_1214	126.188	ARS07_1214	126.188
VA10W_112	126.147	VA10W_112	126.146
IL11_5425	125.147	IL11_5425	125.146
IL09_3264	124.980	IL09_3264	124.979
P981359C1_4	124.731	P981359C1_4	124.730
GA051477_12ES27	124.689	GA051477_12ES27	124.688
GANCZ4_12DH21	124.397	GANCZ4_12DH21	124.397
Roane	124.231	Roane	124.229
KY05C_1369_13_7_3	123.856	KY05C_1369_13_7_3	123.855
Bakhsh24	123.856	Bakhsh24	123.855
NC8170_45_2	123.564	NC8170_45_2	123.564
OK11311F	123.439	OK11311F	123.438
NC09_20986	123.398	NC09_20986	123.397
Milton	123.356	Milton	123.356
NE06607	123.065	NE06607	123.064
IL10_825	122.856	IL10_825	122.857
IL10_6876	122.356	IL10_6876	122.354
IL10_9246	121.940	IL10_9246	121.939
IL10_17659	121.940	IL10_17659	121.939
IL10_6608	121.732	IL10_6608	121.732
MDNC8248_64	121.690	MDNC8248_64	121.689
IL00_8641	121.482	IL00_8641	121.481

.

Table 3.5. Bottom ten percent of soft red winter wheat breeding lines for flour yield based on PEBVs and GEBVs. Breeding lines appearing in both the bottom ten percent for flour yield based on PEBVs and GEBVs are highlighted in the same color to allow changes in rank to be easily identified. Breeding lines that are not highlighted (remaining white) do not appear in the bottom ten percent on the other side of the table.

Breeding line	FY_PEBV	Breeding line	FY_GEBV
IL168_54_2_0_0_0	63.529	IL168_54_2_0_0_0	64.039
VA11W_FHB75	63.930	VA11W_FHB75	64.413
Roane	64.411	KY04C_2004_1_1_3	65.003
IL11_5425	64.651	IL11_5425	65.064
ARGE07_1374_17_8_5	64.731	Roane	65.120
KY04C_2004_1_1_3	64.731	Bakhsh24	65.187
OH904	65.012	OH904	65.244
Bakhsh24	65.012	ARS07_1214	65.404
ARS07_1214	65.092	GANCZ4_12DH21	65.592
GANCZ4_12DH21	65.372	NC08_140	65.637
P92829A1_1_1_3_3	65.412	NC09_20768	65.693
NC08_140	65.452	P92829A1_1_1_3_3	65.697
IL11_2909	65.492	KY02C_3005_25	65.775
NC09_20768	65.492	GANC8248_12DH1	65.873
GANC8248_12DH1	65.532	MO081652	65.915
KY05C_1020_2_19_1	65.613	IL11_2909	65.926
KY02C_3005_25	65.613	KY05C_1020_2_19_1	65.945
IL11_30452	65.653	KY05C_1369_13_7_3	65.950
KY05C_1369_13_7_3	65.693	IL11_110	66.044
MO081652	65.693	IL11_30452	66.051
IL11_110	65.733	IL10_34041	66.058
G09418	65.853	IL10_15225	66.065
IL11_3678	65.853	IL11_3678	66.067
IL10_15225	65.853	GA051477_12ES27	66.074
IL10_34041	65.853	IL11_3466	66.145
GA051477_12ES27	66.013	VA10W_112	66.180
Pio25R37	66.053	IL10_14976	66.210

Table 3.6. Bottom ten percent of soft red winter wheat breeding lines for softness equivalence based on PEBVs and GEBVs. Breeding lines appearing in both the bottom ten percent for softness equivalence based on PEBVs and GEBVs are highlighted in the same color to allow changes in rank to be easily identified. Breeding lines that are not highlighted (remaining white) do not appear in the bottom ten percent on the other side of the table.

Breeding line	SE_PEBV	Breeding line	SE_GEBV
Bakhsh24	38.421	Bakhsh24	38.422
ARS09_228	38.998	ARS09_228	38.999
ARS07_1214	43.260	ARS07_1214	43.261
NE06607	47.257	NE06607	47.257
IL168_54_2_0_0_0	50.897	IL168_54_2_0_0_0	50.898
KY97C_0554_03_02	51.253	KY97C_0554_03_02	51.253
OK11311F	53.650	OK11311F	53.651
IL02_18228	54.361	IL02_18228	54.362
IL02_28322	54.538	IL02_28322	54.539
P03112A1_7_14	54.672	P03112A1_7_14	54.672
Rosette	54.938	Rosette	54.939
KY02C_3005_25	55.071	KY02C_3005_25	55.072
GANCZ4_12DH21	55.160	GANCZ4_12DH21	55.160
IL97_6755	55.426	IL97_6755	55.427
IL04_10741	55.737	IL04_10741	55.738
MDNC8248_64	55.959	MDNC8248_64	55.960
IL99_26442	56.004	IL99_26442	56.004
IL11_20829	56.226	NC09_20768	56.226
NC09_20768	56.226	IL11_20829	56.226
P0128A1_22_22	56.270	M10_1659	56.270
M10_1659	56.270	P0128A1_22_22	56.270
KY02C_3004_04	56.359	KY02C_3004_04	56.359
IL10_12041	56.403	Madison	56.403
Madison	56.403	IL10_12041	56.404
IL10_11050	56.492	Sisson	56.492
IL07_20743	56.492	IL07_20743	56.492
Sisson	56.492	IL10_11050	56.492

Table 3.7. Bottom ten percent of soft red winter wheat breeding lines for lactic acid solvent retention capacity based on PEBVs and GEBVs. Breeding lines appearing in both the bottom ten percent for lactic acid solvent retention capacity based on PEBVs and GEBVs are highlighted in the same color to allow changes in rank to be easily identified. Breeding lines that are not highlighted (remaining white) do not appear in the bottom ten percent on the other side of the table.

Breeding line	LA_PEBV	Breeding line	LA_GEBV
KY93C_1238_17_1	82.455	KY93C_1238_17_1	82.456
P03112A1_7_14	87.411	P03112A1_7_14	87.412
IL10_5629	89.119	IL10_5629	89.121
Freedom	90.118	Freedom	90.119
IL10_5630	90.910	IL10_5630	90.911
NY99045_3110	91.076	NY99045_3110	91.077
KY93C_0004_22_1	92.368	KY93C_0004_22_1	92.369
P05247A1_7_3_120	92.451	P05247A1_7_3_120	92.451
IL08_32162_2	92.951	IL08_32162_2	92.952
VA09W_188WS	93.242	VA09W_188WS	93.243
P0179A1_17	93.617	P0179A1_17	93.618
IL02_28322	93.659	IL02_28322	93.660
F0065	93.909	F0065	93.910
Pio25R35	93.992	Pio25R35	93.993
IL08_18342_5	94.034	IL08_18342_5	94.034
IL168_54_2_0_0_0	94.409	IL168_54_2_0_0_0	94.409
MO080864	95.366	MO080864	95.368
OH07_263_3	95.783	OH07_263_3	95.784
IL06_7550	96.283	IL06_7550	96.284
M03_3616	96.324	M03_3616	96.324
Excel307	96.449	Excel307	96.451
P0128A1_22_22	96.783	P0128A1_22_22	96.783
KY02C_3005_25	97.199	KY02C_3005_25	97.200
Shirley	97.491	Shirley	97.491
IL04_9942	97.532	IL04_9942	97.532
P9975RA1_6_3_94	97.907	P9975RA1_6_3_94	97.908
IL11_20829	98.199	IL11_20829	98.200

Table 3.8. Five-fold cross validation genomic selection prediction accuracies and standard deviations from five runs of the PopVar model for flour yield in soft red winter wheat.

Iteration	Accuracy (r)	Standard Deviation (s)
1	0.552	0.080
2	0.557	0.094
3	0.556	0.091
4	0.539	0.100
5	0.535	0.089
Average	0.547	0.091

Table 3.9. Five-fold cross validated genomic selection prediction correlations and standard deviations from the rrBLUP genomic selection code for flour yield after five iterations in soft red winter wheat.

Iteration	r_CV1	r_CV2	r_CV3	r_CV4	r_CV5	r_avg (r)	r_std (s)
1	0.529	0.498	0.468	0.646	0.722	0.573	0.108
2	0.566	0.572	0.628	0.553	0.644	0.593	0.041
3	0.518	0.646	0.563	0.612	0.457	0.559	0.075
4	0.453	0.506	0.502	0.506	0.621	0.517	0.062
5	0.315	0.386	0.518	0.773	0.581	0.515	0.179
Average						0.551	0.093

Table 3.10. Five-fold cross validated genomic selection prediction correlations and standard deviations from the rrBLUP genomic selection code for softness equivalence after five iterations in soft red winter wheat.

Iteration	r_CV1	r_CV2	r_CV3	r_CV4	r_CV5	r_avg (r)	r_std (s)
1	0.541	0.456	0.640	0.401	0.506	0.509	0.090
2	0.580	0.439	0.637	0.634	0.412	0.540	0.108
3	0.377	0.478	0.641	0.486	0.514	0.499	0.095
4	0.596	0.525	0.617	0.436	0.379	0.510	0.102
5	0.417	0.568	0.560	0.448	0.523	0.503	0.068
Average						0.512	0.093

Table 3.11. Five-fold cross validated genomic selection prediction correlations and standard deviations from the rrBLUP genomic selection code for lactic acid solvent retention capacity after five iterations in soft red winter wheat.

Iteration	r_CV1	r_CV2	r_CV3	r_CV4	r_CV5	r_avg (r)	r_std (s)
1	0.479	0.654	0.506	0.601	0.671	0.582	0.086
2	0.621	0.643	0.659	0.624	0.537	0.617	0.047
3	0.668	0.548	0.621	0.651	0.492	0.596	0.074
4	0.532	0.509	0.718	0.560	0.576	0.579	0.082
5	0.558	0.658	0.560	0.616	0.551	0.597	0.044
Average						0.594	0.066

FIGURES

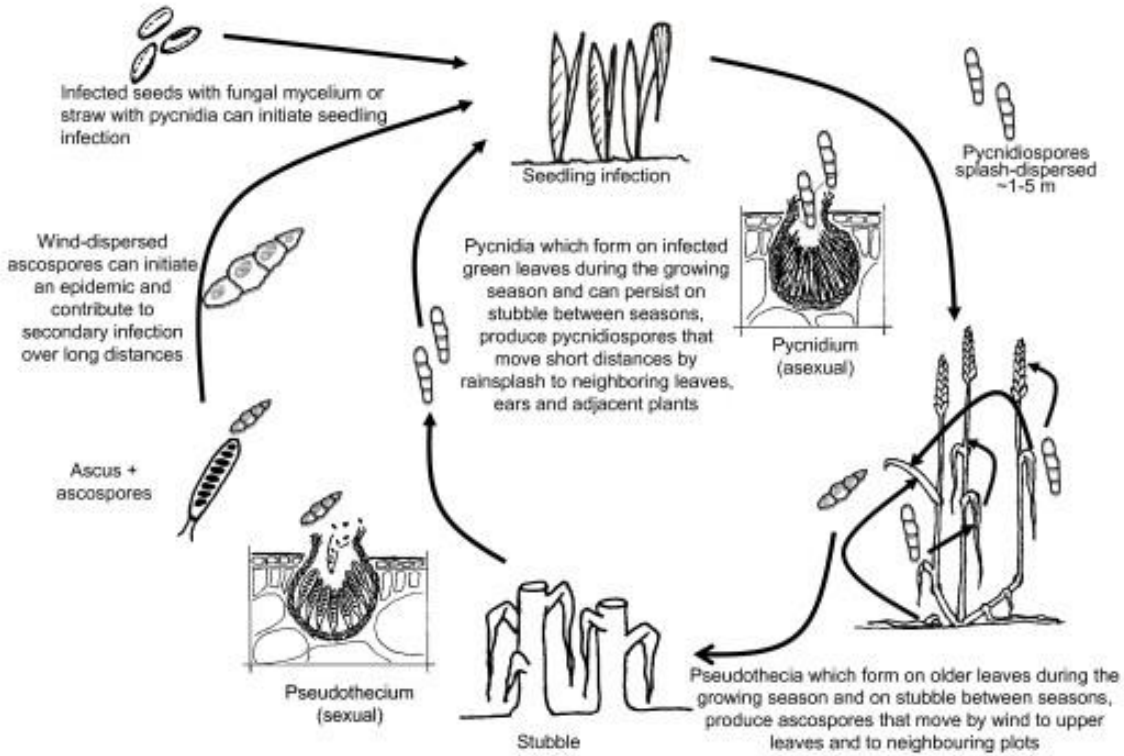

Figure 1.1. Generalized disease cycle for *Parastagonospora nodorum* from Sommerhalder et al., 2011.

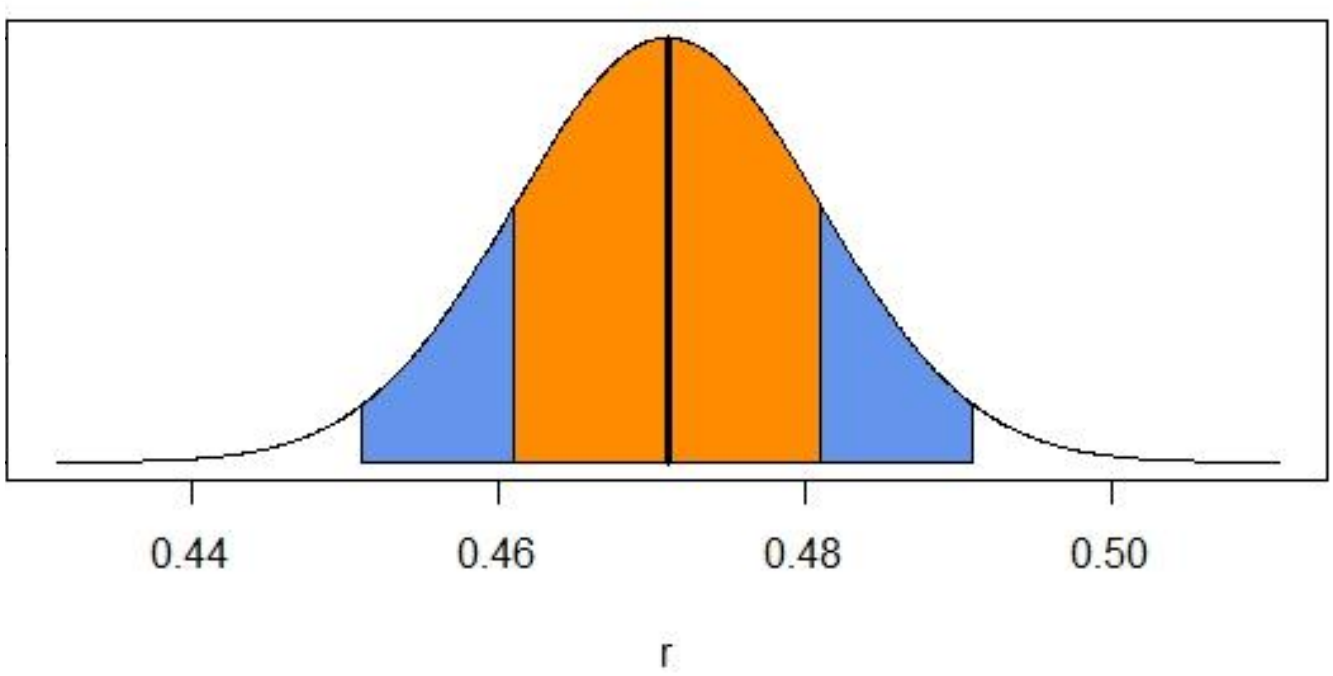

Figure 2.1. The average prediction accuracy of the five runs of the five-fold cross validation genomic selection model in PopVar for GEBVs for glume blotch resistance in soft red winter wheat breeding lines. From the central limit theorem, the average accuracy is not in danger of crossing below 0.3. It has been said as long as prediction accuracies remain above 0.3 genetic gains can be made through genomic selection. The trait of glume blotch resistance does not cross below 0.3 one standard error (orange) out from the average accuracy or even two standard errors (cornflower blue) out from the average accuracy.

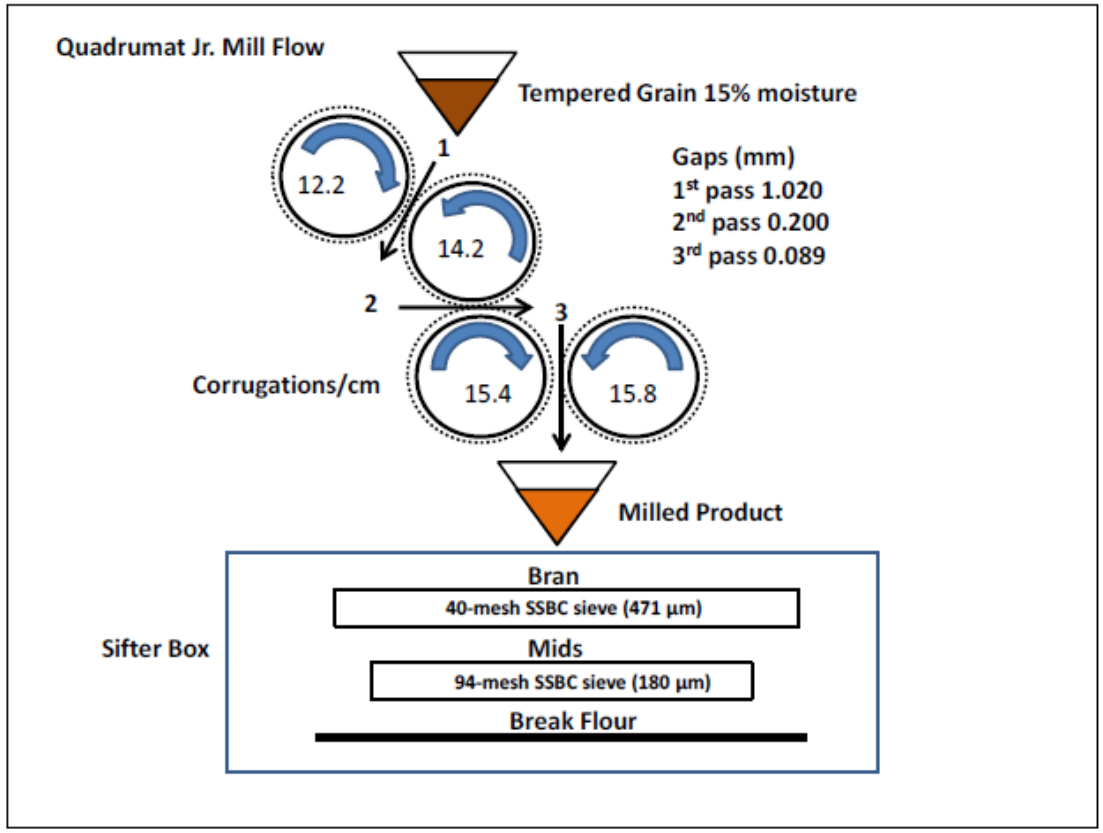

Figure 3.1. Illustration of the Quadrumat break roll milling unit and its milling process. The Quadrumat break roll milling unit is used by the Soft Wheat Quality Lab in Wooster, Ohio to test milling and baking qualities of soft wheats for the eastern United States growing region. The figure is from the USDA-ARS SWQL Materials and Methods 2017. Figure 1 (USDA, 2018a).

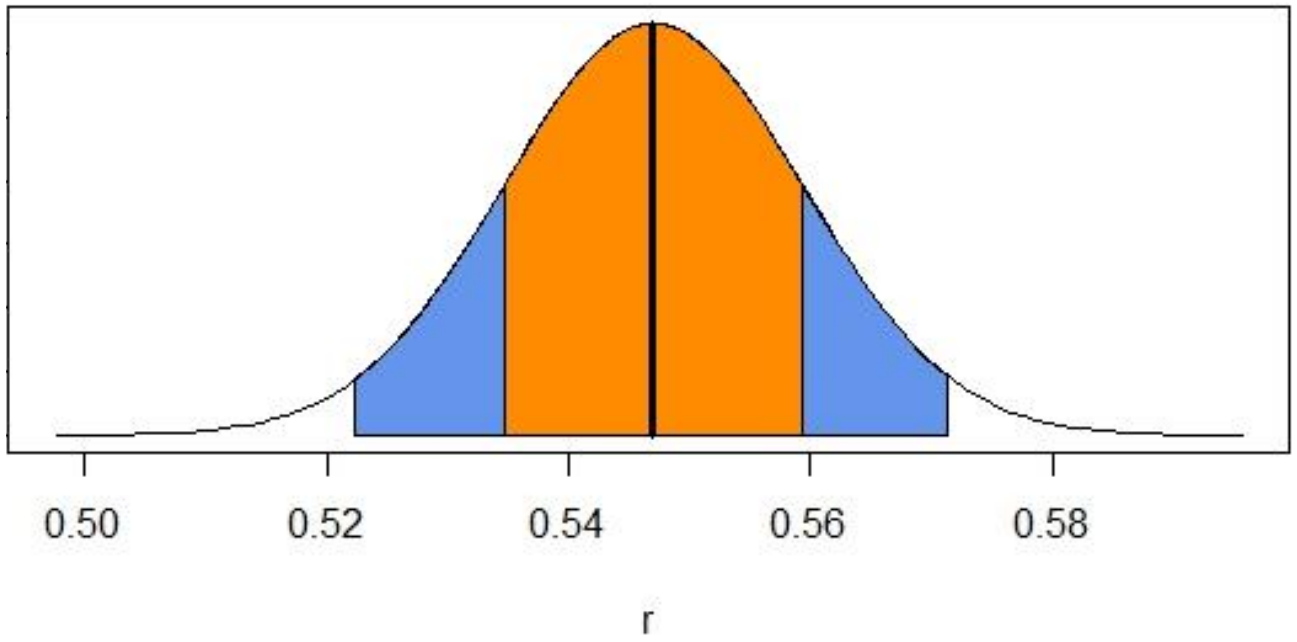

Figure 3.2. The average prediction accuracy of the five runs of the five-fold cross validation genomic selection model in PopVar for GEBVs for flour yield in soft red winter wheat breeding lines. From the central limit theorem, the average accuracy is not in danger of crossing below 0.3. It has been said as long as prediction accuracies remain above 0.3 genetic gains can be made through genomic selection. The trait of flour yield does not cross below 0.3 one standard error (orange) out from the average accuracy or even two standard errors (cornflower blue) out from the average accuracy.

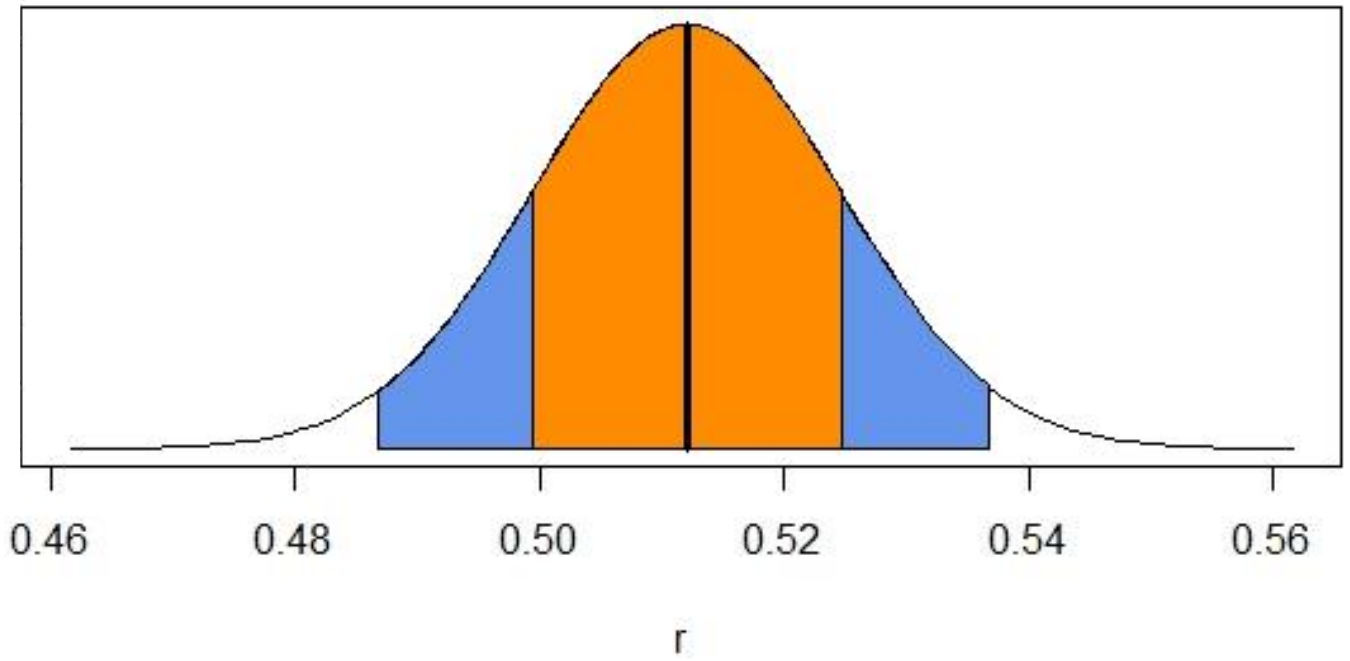

Figure 3.3. The average prediction accuracy of the five runs of the five-fold cross validation rrBLUP genomic selection model for GEBVs for softness equivalence in soft red winter wheat breeding lines. From the central limit theorem, the average accuracy is not in danger of crossing below 0.3. It has been said as long as prediction accuracies remain above 0.3 genetic gains can be made through genomic selection. The trait of flour yield does not cross below 0.3 one standard error (orange) out from the average accuracy or even two standard errors (cornflower blue) out from the average accuracy.

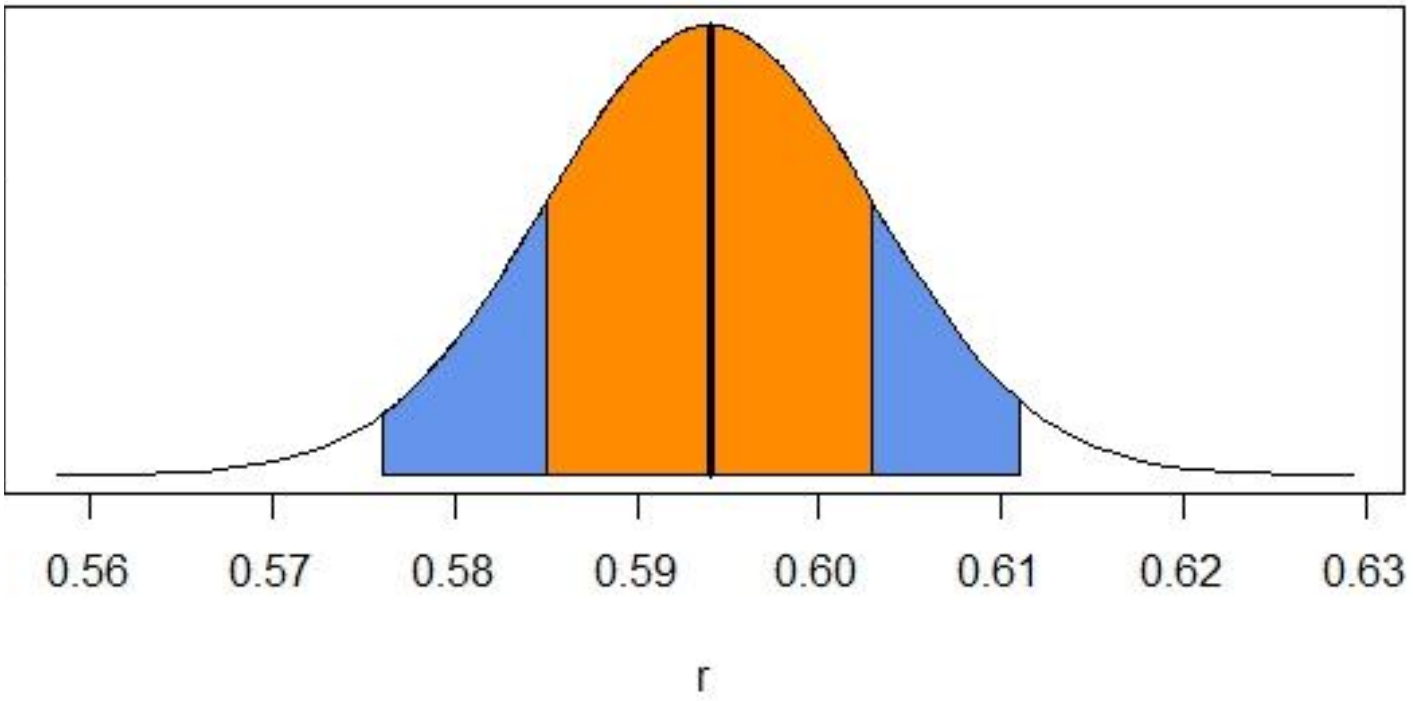

Figure 3.4. The average prediction accuracy of the five runs of the five-fold cross validation rrBLUP genomic selection model for GEBVs for lactic acid solvent retention capacity in soft red winter wheat breeding lines. From the central limit theorem, the average accuracy is not in danger of crossing below 0.3. It has been said as long as prediction accuracies remain above 0.3 genetic gains can be made through genomic selection. The trait of flour yield does not cross below 0.3 one standard error (orange) out from the average accuracy or even two standard errors (cornflower blue) out from the average accuracy.

REFERENCES

Arruda, M. P., Brown, P. J., Lipka, A. E., Krill, A. M., Thurber, C., and Kolb, F. L. (2015). Genomic selection for predicting Fusarium head blight resistance in a wheat breeding program. The Plant Genome, 8(3). doi:10.3835/plantgenome2015.01.0003,

Baker, E. A. (1978). Septoria – the lurking threat to wheat yields. EPPO Bull, 8 (1): 9-20.

Bernardo, R. (2016). Bandwagons I, too, have known. Theor Appl Genet. 129: 2323-2332. doi: 10.1007/s00122-016-2772-5.

Bernardo, R. and Yu, J. (2007). Prospects for genomewide selection for quantitative traits in maize. Crop Sci. 47:1082-1090. doi: 10.2135/cropsci2006.11.0690.

Bertucci, M., Brown-Guedira, G., Murphy, J. P., and Cowger, C. (2014). Genes conferring sensitivity to Stagnosopora nodorum necrotrophic effectors in Stagnosopora nodorum blotch-susceptible U.S. wheat cultivars. Plant Dis. 98:746-753.

Brenchley, R., Spannagl, M., Pfeifer, M., Barker, G. L., D'Amore, R., Allen, A. M., . . . Hall, N. (2012). Analysis of the bread wheat genome using whole-genome shotgun sequencing. Nature, 491(7426), 705-710. doi: 10.1038/nature11650

Bostwick, D. E., Ohm, H. W., and Shaner, G. (1993). Inheritance of Septoria glume blotch resistance in wheat. Crop Science, 33, 439-443

Cabrera, A., Guttieri, M., Smith, N., Souza, E., Sturbaum, A., …Sneller, C. (2015). Identification of milling and baking quality QTL in multiple soft wheat mapping populations. Theor Appl Genet. doi: 10.1007/s00122-015-2580-3

Cooke, B. M., and Jones, D. G. (1970). A field inoculation method for *Septoria tritici* and *Septoria nodorum*. Pl. Path., 19, 72-74.

Cowger, C., and Murphy, J. P. (2007). Artificial inoculation of wheat for selecting resistance to Stagonospora Nodorum blotch. Plant Disease, 91, 539-545. doi: 10.1094/PDIS-91-5-0539.

Cowger, C. and Silva-Rojas, H. V. (2006). Frequency of *Phaesophaeria nodorum*, the sexual stage of *Stagonospora nodorum*, on winter wheat of North Carolina. Phytopathology, 96:860-866. doi: 10.1094/PHYTO-96-0860.

Endelman, J. B. (2011). Ridge regression and other kernels for genomic selection with R package rrBLUP. Plant Genome, 4(3):250-255. doi: 10.3835/plantgenome2011.08.0024.

Endalman, J. B., and Jannink, J. L. (2012). Shrinkage estimation of the realized relationship matrix. G3, 2(11): 1405-1413. doi: 10.1534/g3.112.004259.

Eyal, Z., Scharen, A. L., Prescott, J. M. and van Ginkel, M. (1987). The Septoria diseases of wheat: concepts and methods of disease management. CIMMYT, Mexico, DF, Mexico.

Ficke, A., Cowger, C., Bergstrom, G., and Brodal, G. (2018). Understanding yield loss and pathogen biology to improve disease management: Septoria nodorum blotch- a case study in wheat. Plant Disease. 102:696-707. https://doi.org/10.1094/PDIS-09-17-1375-FE

Finney, P. L. (1990). Flour, milling and end-use quality of Ohio wheat. Profitable Wheat Management. J. E. Beuerlein, ed. Ohio State Univ. Ext. Bull. 811:79-87.

Francki, M. G. (2013). Improving *Stagonospora nodorum* resistance in wheat: a review. Crop Sci. 53: 355-365. doi: 10.2135/cropsci2012.06.0347.

Fraser, D. E., Murphy, J. P., Leath, S., and Van Sanford, D. A. (2003). Effect of inoculation with selected isolates of Stagonospora nodorum on field evaluations of host resistance in winter wheat. Plant Disease. 87:1213-1220.

Fried, P. M., and Meister, E. (1987). Inheritance of leaf and head resistance of winter wheat to Septoria nodorum in a diallel cross. Phytopathology 77:1371-1375.

Goddard, M. E., and Hayes, B. J. (2007). Genomic selection. J. Anim. Breed. Genet. 124:323-330.

Griffiths, E., and Ao, H. C. (1976). Dispersal of Septoria nodorum spores and spread of glume blotch of wheat in the field. Trans. Br. Mycol. Soc. 67: 413-418.

Heffner, E. L., Sorrells, M. E., and Jannink, J. L. (2009). Genomic selection for crop improvement. Crop Sci. 49:1-12. doi: 10.2135/cropsci2008.08.0512

Heffner, E. L., Lorenz, A. J., Jannink, J. L., and Sorrells, M. E. (2010). Plant breeding with genomic selection: gain per unit time and cost. Crop Sci. 50:1681-1690. doi: 10.2135/cropsci2009.11.0662

Heffner, E. L., Jannink, J. L., Iwata, H., Souza, E., and Sorrells, M. E. (2011). Genomic selection accuracy for grain quality traits in biparental wheat populations. Crop Sci. 51: 2597-2606. doi: 10.2135/cropsci2011.05.0253.

Holmes, S. J. I., and Colhoun, J. (1974). Infection of wheat by *Septoria nodorum* and *S. tritici* in relation to plant age, air temperature, and relative humidity. Trans. Br. mycol. Soc. 63 (2), 329-338.

Huang, M., Cabrera, A., Hoffstetter, A., Griffey, C., Van Sanford, D., Costa, J., …Sneller, C. (2016). Genomic selection for wheat traits and trait stability. Theor Appl Genet. doi: 10.1007/s00122-016-2733-z.

Illinois Wheat Association. (2016). About Illinois Wheat. http://www.illinoiswheat.org/about-illinois-wheat.html

International Rice Genome Sequencing Project. (2005). The map-based sequence of the rice genome. Nature 436, 793-800. doi: 10.1038/nature03895

International Wheat Genome Sequencing Consortium. (2018a). Infographic: the bread wheat genome. http://www.wheatgenome.org/News/Media-resources/Fact-Sheets-Infographics/Wheat-genome

International Wheat Genome Sequencing Consortium. (2018b). Infographic: facts about wheat. http://www.wheatgenome.org/News/Media-resources/Fact-Sheets-Infographics/Facts-about-wheat

International Wheat Genome Sequencing Consortium. (2017). Infographic: the origin of wheat. http://www.wheatgenome.org/News/Media-resources/Fact-Sheets-Infographics/Origin-of-Wheat

Jenkyn, J. F., and King, J. E. (1977). Observations on the origins of *Septoria nodorum* infection of winter wheat. Pl. Path., 26, 153-160.

Kim, Y., Brown-Guedira, G. L., Cox, T. S., and Bockus, W. W. (2004). Inheritance of resistance to Stagnosopora nodorum leaf blotch in Kansas winter wheat cultivars. Plant Disease, 88 (5), 530-536

Kleczewski, N. (2015). Stagonospora leaf and glume blotch of wheat. University of Delaware. Cooperative Extension.

Krupinsky, J. M., Scharen, A. L., and Schillinger, J. A. (1973). Pathogenic variation in *Septoria nodorum* (Berk.) Berk. in relation to organ specificity, apparent photosynthetic rate and yield of wheat. Physiological Plant Pathology, 3, 187-194.

Leath, S., Scharen, A. L., Lund, R. E. and Dietz-Holmes, M. E. (1993). Factors associated with global occurrences of Septoria nodorum blotch and Septoria tritici blotch of wheat. Plant Dis, 77: 1266-1270.

Mehra, L. K., Cowger, C., Weisz, R., and Ojiambo, P. S. (2015). Quantifying the effects of wheat residue on severity of Stagonospora nodorum blotch and yield in winter wheat. Phytopathology. 105: 1417-1426.

Mehra, L. K., Cowger, C., Gross, K., and Ojiambo, P. S. (2016). Predicting preplanting risk of Stagonospora nodorum blotch in winter wheat using machine learning models. Front Plant Sci. 7: 390. doi: 10.3389/fpls.2016.00390.

Mehra, L. K., Cowger, C., and Ojiambo, P. S. (2017). A model for predicting onset of Stagonospora nodorum blotch in winter wheat based on preplanting and weather factors. Phytopathology. 107: 635-644.

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

McKendry, A. L., Henke, G. E., and Finney, P. L. (1995). Effects of septoria leaf blotch on soft red winter wheat milling and baking quality. Cereal Chem. 72(2):142-146.

Mohammadi, M., Tiede, T., and Smith, K. P. (2015). PopVar: a genome-wide procedure for predicting genetic variance and correlated response in biparental breeding populations. Crop Science, 55, 1-10. doi: 10.2135/cropsci2015.01.0030

Murray, G. M., and Brennan, J. P. (2009). Estimating disease losses to the Australian wheat industry. Australasian Plant Pathology, 38, 558-570.

Nafziger, E. (2009). Illinois Agronomy Handbook. 24th ed. Department of Crop Sciences, University of Illinois Urbana-Champaign. Chapter 4, Small grains and grain sorghum; p. 37-47. http://extension.cropsciences.illinois.edu/handbook/pdfs/chapter04.pdf

Plant Genome and Systems Biology. (2018). Welcome to the maize genome database. http://pgsb.helmholtz-muenchen.de/plant/maize/index.jsp

R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Rooney, J. M., and Hoad, G. V. (1989). Compensation in growth and photosynthesis of wheat (*Triticum aestivum* L.) following early inoculations with *Septoria nodorum* (Berk.) Berk. New Phytol., 113, 513-521.

SAS Institute. (2017). SAS system for Windows. v. 9.4. SAS Inst. Inc., Cary, NC.

Schuler, S. F., Bacon, R. K., Finney, P. L., and Gbur, E. E. (1995). Relationship of test weight and kernel properties to milling and baking quality in soft red winter wheat. Crop Sci. 35:949-953.

Selter, L. L., Shatalina, M., Singla, J., and Keller, B. (2014). Identification and implementation of resistance: genomics-assisted use of genetic resources for breeding against powdery mildew and Stagonospora nodorum blotch in wheat. Genomics of Plant Genetic Resources, 359-383. doi: 10.1007/978-94-007-7575-6_15

Sommerhalder, R.J., McDonald, B.A., Mascher, F., and Zhan, J. (2011). Effect of hosts on competition among clones and evidence of differential selection between pathogenic and saprophytic phases in experimental populations of the wheat pathogen Phaeosphaeria nodorum. BMC Evol Biol. doi: 10.1186/1471-2148-11-188

Souza, E. J., Sneller, C., Guttieri, M. J., Sturbaum, A…Van Sanford, D. (2012). Basis for selecting soft wheat for end-use quality. Crop Sci. 52:21-31. doi: 10.2135/cropsci2011.02.0090.

Tiede, T. and Mohammadi, M. (2015). Package 'PopVar'. https://cran.r-project.org/web/packages/PopVar/PopVar.pdf

United States Department of Agriculture. Agricultural Research Service. (2018a). Soft Wheat Quality Lab, SWQL Materials and Methods. https://www.ars.usda.gov/ARSUserFiles/50821000/36070500/Materials%20and%20Methods%202017.pdf

United States Department of Agriculture. (2018b). 2017 State Agriculture Overview – Illinois. https://www.nass.usda.gov/Quick_Stats/Ag_Overview/stateOverview.php?state=ILLINOIS

Von Wechmar, M. B. (1966). Investigation on the survival of Septoria nodorum Berk. on crop residues. S. Afr. J. Agric. Sci., 9, 93-100.

Wainshilbaum, S. J., and Lipps, P. E. (1991). Effect of temperature and growth stage of wheat on development of leaf and glume blotch by Septoria tritici and S. nodorum. Plant Disease, 75, 993-998.

APPENDIX A

Table A.1. Variance parameter estimates for flour yield, softness equivalence, and lactic acid solvent retention capacity in a 272 wheat breeding line panel. Variance parameter estimates obtained through PROC MIXED in SAS 9.4. Each parameter was run in a separate RCBD model, but are shown side by side.

Variance Parameter Estimates						
Variance Parameter	**Flour Yield Estimate**	**Flour Yield P-value**	**Softness Equivalence Estimate**	**Softness Equivalence P-value**	**Lactic Acid SRC Estimate**	**Lactic Acid SRC P-value**
Year	0.864	0.2405	4.996	0.2402	24.888	0.2409
Genotype	1.715	<0.0001**	12.024	<0.0001**	88.291	<0.0001**
Residual	0.850	<0.0001**	3.032	<0.0001**	35.395	<0.0001**

**Highly significant at alpha 0.05

APPENDIX B

PopVar R code used for the glume blotch resistance (Ch. 2) and milling and baking, flour yield trait (Ch. 3) genomic selection models (Mohammadi et al., 2015). Special thanks to Marcio Arruda and Allison Krill for slight code modifications to better adapt for the University of Illinois' small grains program.

```
install.packages("PopVar")
library(PopVar)


##Prepare/load files for genomic selection
#SNP genotype file
geno=read.csv("genotype.csv",strings=F, header = TRUE)
geno[1:10,1:10] #check to make sure file was entered properly


#your phenotype file, taxa must match with SNP file
pheno = read.csv("pheno_alltraits.csv", header=TRUE)
pheno=pheno[order(pheno[,1]),]
pheno[1:10,1:2] #check to make sure file was entered properly


#make sure taxa match in both pheno and geno files
table(pheno[,1]==geno[,1]) #make sure line/taxa names are in the same order in both files


names = geno[,1]
names = as.matrix(names)
write.csv(names, file="names.csv") #save names


#run one pheno value at a time, name and phenotype
pheno_trait1  = pheno[,c(1,2)] #first trait is in column 2 with names in first column
pheno_trait2 = pheno[,c(1,3)] #second trait is in column 3 with names in first column
```

```
#Run Genomic Selection code from PopVar
#GEBV values
GS = x.val(G.in = geno, #reads in your geno file with names removed
        y.in = pheno_trait1, #reads in your pheno file for desired trait
        min.maf = 0, #no filtering, we did already did this
        mkr.cutoff=1, #no filtering
        entry.cutoff = 1, #no filtering
        remove.dups = FALSE,
        impute = "pass",
        nFold = 5, # number of folds we usually use 5 unless it is a small data set
        nFold.reps = 10,
        return.estimates = T,
        models = c("rrBLUP"))
#model accuracy
results = GS$CVs
model = GS$models.used
effects = GS$mkr.effects
beta = GS$betas


effects = as.matrix(effects)
geno = geno[,-1]
geno = as.matrix(geno)
geno=apply(geno[,1:ncol(geno)],2,as.numeric)


#Genomic estimated breeding values
GEBV = geno%*%effects
GEBV = GEBV + beta
GEBV = cbind(names, GEBV)

#this is the output file for GEBV #change the name for each trait
write.csv(GEBV, file="GEBV_trait1.csv")
#this is the output file for model accuracy
write.csv(results, file ="results_trait1.csv")
```

APPENDIX C

rrBLUP R code used for milling and baking, softness equivalence and lactic acid solvent retention capacity genomic selection (Endelman, 2011; Endelman and Jannik, 2012). Special thank you to Dr. Alex Lipka and Dr. Carrie Butts-Wilmsmeyer for general guidance in writing this code.

```
##Prepare/load files for genomic selection
#SNP genotype file
geno=read.csv("genotype.csv",strings=F, header = TRUE)
geno[1:10,1:10] #check to make sure file was entered properly


#your phenotype file, taxa must match with SNP file
pheno = read.csv("pheno_alltraits.csv", header=TRUE)
pheno=pheno[order(pheno[,1]),]
pheno[1:10,1:2] #check to make sure file was entered properly


#make sure taxa match in both pheno and geno files
table(pheno[,1]==geno[,1]) #make sure line/taxa names are in the same order in both files


names = geno[,1]
names = as.matrix(names)
write.csv(names, file="names.csv") #save names


#run one pheno value at a time, name and phenotype
pheno_trait1  = pheno[,c(1,2)] #first trait is in column 2 with names in first column
pheno_trait2 = pheno[,c(1,3)] #second trait is in column 3 with names in first column

#rrBLUP
library(rrBLUP)
library('MASS')
library(multtest)
library(gplots)
library("compiler")
source("http://zzlab.net/GAPIT/gapit_functions.txt")
```

```
#Setting Up Random Selection and Number of Folds, Formatting, etc.
seed.number <- sample(-1000000:1000000,1)
number.of.folds=5
sample.size=272
set.seed(seed.number)
sequence.sample <- rep(1:sample.size)
random.sample <- sample(1:sample.size, replace = FALSE)
increment <- ceiling(length(random.sample)/number.of.folds)
r.gy <- NULL
y=pheno_trait1 #phenotype file for desired trait
GEBV=matrix(NA,nrow=272,ncol=6)
GEBV[,1]=y$line
k <- number.of.folds - 1
G=geno1[,-1]

#Validating Model
for(i in 0:k){
  pred <- random.sample[((increment*i)+1):min(((increment*i)+increment) , sample.size)]
  train <- random.sample[-(((increment*i)+1):min(((increment*i)+increment) , sample.size))]

  ans <-
kinship.BLUP(y=as.numeric(y[train,2]),G.train=G[train,],G.pred=G[pred,],K.method="RR")

  r.gy <- c(r.gy, cor(ans$g.pred,as.numeric(y[pred,2])))
}
r.gy <- c(r.gy, mean(r.gy), sd(r.gy))
colname.r.gy <- NULL
for(i in 1:number.of.folds) colname.r.gy <- c(colname.r.gy, paste("r_CV",i,sep = ""))
r.gy.output <- t(as.matrix(r.gy))
colnames(r.gy.output) <- c(colname.r.gy, "r_avg", "r_std")
r.gy.output #model accuracy output

##write csv for correlations of GEBV and PEBV for accuracy
write.csv(r.gy.output, file ="model accuracy trait1.csv")

#GEBV calculations
GEBV.ans=kinship.BLUP(y=y[,2],G.train=G,K.method="RR")
GEBVs=as.matrix(GEBV.ans$g.train+as.numeric(GEBV.ans$beta))
names=as.matrix(y[,1])
GEBVs=cbind(names,GEBVs)
View(GEBVs)
write.csv(GEBVs, "GEBV_ trait1.csv") #save GEBV file for trait
```