2018

Application of Genomic Approaches to Improve Yield and Bacterial Leaf Streak Resistance in Winter Wheat

Sai Mukund Ramakrishnan
South Dakota State University

APPLICATION OF GENOMIC APPROACHES TO IMPROVE YIELD AND BACTERIAL LEAF STREAK RESISTANCE IN WINTER WHEAT

BY

SAI MUKUND RAMAKRISHNAN

A thesis submitted in partial fulfillment of the requirements for the

Master of Science

Major in Plant Science

South Dakota State University

2018

APPLICATION OF GENOMIC APPROACHES TO IMPROVE YIELD AND BACTERIAL LEAF STREAK RESISTANCE IN WINTER WHEAT

SAI MUKUND RAMAKRISHNAN

This thesis is approved as a creditable and independent investigation by a candidate for the Master in Plant Science degree and is acceptable for meeting the thesis requirements for this degree. Acceptance of this does not imply that the conclusions reached by the candidate are necessarily the conclusions of the major department.

Sunish Kumar Sehgal, Ph.D. Date
Thesis Advisor

David Wright, Ph.D. Date
Head, Department of Agronomy,
Horticulture and Plant Science

Dean, Graduate School Date

ACKNOWLEDGEMENTS

I would like to thank my major advisor Dr. Sunish Kumar Sehgal, for providing me this great opportunity to pursue my Master's degree in the Department of Agronomy, Horticulture and Plant Science of this university. I really appreciate his knowledge, excellence, patience and productive advice throughout my degree program. He always inspired me to become an independent researcher and helped me a lot in different aspects. I gained immense knowledge from him and learned a lot about different techniques.

I would like to extend my sincere gratitude to my thesis advisory committee members Dr. Shaukat Ali and Dr. Jixiang Wu and Graduate School Representative Trevor C. Roiger. They generously gave their time to provide me valuable suggestions and encouragement towards improving my research work. In particular, Dr. Shaukat Ali and Dr. Jixiang Wu provided me the opportunity to learn about the phenotyping techniques and design experiments for my study. Their knowledge, experience, and advice really strengthened my knowledge in my research area. I would also like to thank Dr. Karl Glover for providing the bacterial isolates. I am grateful to South Dakota Wheat Commission who provided the funding for this project and the South Dakota State University Agriculture Experimental Station who provided the field plots for me to test my samples.

I would like to thank my lab mates Jagdeep Singh Sidhu, Jyotirmoy Halder and Girma Tedese Ayana for their help and encouragement. Also, I would like to extend my thanks to the graduate and undergraduate students Christopher Lee, Cody, Sam who helped me with greenhouse work and other research chores.

I am definitely fortunate to have friends like Shrija Srinivasan, Navjot Kaur, Shikha Surendra Rathore, Abhinav Sharma, Mark Kirschenman, Jasdeep Singh Bhullar, Dilkaran Singh Dhillon and Navdeep Singh in my life who gave me constant moral support and their encouragement throughout my study. Thank you all for direct and indirect support and emotional support with endless patience.

Finally, I would like to thank my mother Mrs. Aravinda Ramakrishnan, my father Mr. T. V. Ramakrishnan back home in India without whom I could not have imagined this journey to get accomplished on time. It's their love, patience, and sacrifice which embraced me to achieve this degree.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

AM Association Mapping

BLS Bacterial Leaf Streak

GBLUP Genomic Best Linear Unbiased Prediction

GS Genomic Selection

HWWAMP Hard Winter Wheat Association Mapping Panel

LD Linkage Disequilibrium

MAS Marker Assisted Selection

NIL Near Isogenic Lines

PLS Partial Least Square

PS Phenotypic Selection

QTL Quantitative Trait Loci

RIL Recombinant Inbred Lines

rrBLUP ridge regression Best Linear Unbiased Prediction

SNP Single Nucleotide Polymorphism

WW Winter Wheat

Xct Xanthomonas campestris pv. translucens.

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

APPLICATION OF GENOMIC APPROACHES TO IMPROVE YIELD AND BACTERIAL LEAF STREAK RESISTANCE IN WINTER WHEAT

SAI MUKUND RAMAKRISHNAN

2018

Global wheat production is threatened by the change in climate thus leading lead to the increase in the biotic and abiotic stresses. We need to increase wheat productivity at a faster pace and manage these challenges to meet the growing demand. Development of cultivars with durable disease resistance and enhancing the rate of genetic gain in wheat are the major goals in wheat breeding programs. Bacterial Leaf Streak (BLS) is one of the most threatening bacterial diseases to wheat in the US Northern Great Plains. Unlike fungal diseases, bacterial diseases cannot be effectively managed using chemicals and thus developing disease resistant cultivars would be the most economical control for BLS. Identification and characterization of genomic regions in wheat that confer resistance to BLS can be an effective way to mobilize resistance genes in wheat breeding. Here we performed Genome – wide association mapping on a Hard Winter Wheat Association Panel (HWWMP) to identify genomic regions that confer resistance to BLS. The genotyped data for this panel of 300 winter wheat lines from the major breeding programs across the Midwestern region of the US was obtained from T3 Triticale Toolbox (under the GPL license). The responses of all these lines against *Xanthomonas*

campestris pv. translucens in the greenhouse and field conditions were evaluated. Association Mapping (AM) was used to detect marker – trait associations using ECMLM, and we identified five QTL regions (*Q.bls.sdsu.1AL*, *Q.bls.sdsu.1BS*, *Q.bls.sdsu.3AL*, *Q.bls.sdsu.4AL* and *Q.bls.sdsu.7AS*) conferring BLS resistance. In total, these five QTLs explained 42% of the variation. Eleven genotypes were identified, which could be used as a source of resistance against BLS. Comparative analysis of three of the identified QTLs (*Q.bls.sdsu.1AL*, *Q.bls.sdsu.3AL* and Q.*bls.sdsu.4AL*) with rice showed BLS resistance genes in rice (*qBLSr5d*, *qBLSr1*, and *qBLSr3d*) located on syntenic regions in rice chromosomes 5R, 1R and 3R respectively. The 11 BLS resistant genotypes and SNP markers linked to QTLs identified in our study could facilitate breeding BLS resistance in wheat. For grain yield improvement, we assessed the robustness for genomic selection (GS) in the South Dakota State Winter Wheat Breeding program (SDSWWBP). We performed GS with a set of 434 advanced breeding lines (AYT and PYT nurseries) between the years 2014 – 2017. These lines were genotyped by sequencing GBS and the yield data from 34 years × location combinations were used as a phenotype. We developed training and validation datasets for testing the genomic prediction accuracies. Single and multiyear analysis were done using several GS models (rrBLUP, PLSR, ELNET and Random Forest). The average predictions accuracies within a single year across locations were 0.62. However, with the multi-year-location analysis, the average genomic prediction accuracies were 0.26 for two-year combination, 0.32 for three-year combination and 0.36 for the four-year combination. Our results suggested several years of data is required to develop better genome-wide selection models.

CHAPTER – 1

GENERAL INTRODUCTION

1. Introduction

Wheat was domesticated around 8000 years ago and ever since has been the basic staple food of the major civilizations and the most important food grain source for humans (Peng et al. 2011). As the most widely planted cereal crop around the globe, wheat has the greatest world trade among all crops and its production leads most crops, including rice, maize, and potatoes (Lev-Yadun et al. 2002).

A traditional winter wheat breeding program normally requires at least 10 or 12 years before a cultivar is ready for commercial release (Kuchel et al. 2005; Reynolds et al. 2011; Kirigwi et al. 2004; Wrigley 1994). Majority of the wheat breeding programs around the world focus on increasing the yield of wheat cultivars while maintaining disease resistance (Kadar and Moldovan 2003). As the demand for wheat consumption is exceeding the current supply, an estimated 1.6% annual increase in wheat production is required to fulfill the projected demand in 2020 of 760 million tons (Reynolds et al. 2012). Given the present average increase rate of 1.1%, the mismatch between the projected supply and demand is an obvious global challenge (Lupton 2005). As a result, it is imperative to incorporate emerging technologies into wheat breeding programs to meet these challenges (Joosen et al. 2009).

With the availability of thousands of single nucleotide polymorphism (SNP) markers across the entire genome, GS can predict an individual's performance for quantitative traits and this has been demonstrated in animal breeding (Wu et al. 2001). GS holds promise in accelerating rate the genetic gain thereby shorting the breeding cycle (Varshney et al. 2007a). While the traditional phenotypic selection is cumbersome and inefficient, the use of GS with its genotypic information, made it possible to predict adult plants' performance from information generated at the early seedlings stage. This advancement can be used to predict phenotypes to substitute the phenotype – dependent field evaluation, thus the effort and investment on field assessment for phenotypes are substantially reduced (Kuti et al. 2012).

In addition to limiting yield, changing climate is leading to the emergence of new race and pathogens causing diseases (de Souza et al. 2014). With the availability of resistant cultivars for fungal diseases, there has been a significant increase in the incidence of bacterial diseases in wheat in recent times. In the US Northern Great Plains (NGP) Bacterial Leaf Streak (BLS) has become one of the most threatening bacterial diseases to wheat (Tillman et al. 1999). Unlike fungal diseases, bacterial diseases cannot be countered or controlled using bactericides or an antibacterial for cost-effective measures (Kumar and Sakthivel 2001). Identification of diseases resistant cultivars against BLS has become a goal of major breeding programs in the US NGP (Tillman et al. 1999). Tackling such complex traits and diseases require the incorporation of high computational methods along with the traditional phenotyping and breeding methodology (Tang et al. 2000). With the increase in genomic data and

bioinformatics techniques, several high-throughput techniques have emerged to tackle these bottlenecks in wheat breeding (Kuti et al. 2012).

Genome-wide Association studies (GWAS) is a strategy to identify marker – trait associations and has been used extensively in human and animal genetic experiments where large segregating populations are not available (Varshney et al. 2007b). GWAS has a number of advantages over other linkage mapping techniques including the potential for increased QTL resolution, and an increased sampling of molecular variation, both factors associated with the use of unrelated populations which is possible with GWAS (Christopher et al. 2007). Using GWAS, we can identify potential markers/QTLs that can be used in MAS to characterize disease resistant genotypes which can be used as a source of resistance to several breeding programs (Arora et al. 2017).

Implementing these techniques in the South Dakota winter wheat breeding program, the objectives of this study are as follows:

i. To identify genomic regions conferring bacterial leaf streak (BLS) resistance in hard winter wheat association mapping panel (HWWAMP) and develop SNP markers for marker-assisted selection.

ii. To evaluate the relative efficiency of genomic selection versus phenotypic selection for grain yield in the South Dakota winter wheat breeding program.

2. Literature cited

Arora S, Singh N, Kaur S, Bains NS, Uauy C, Poland J, Chhuneja P (2017) Genome-Wide Association Study of Grain Architecture in Wild Wheat Aegilops tauschii. Frontiers in plant science 8:886. doi:10.3389/fpls.2017.00886

10.3389/fpls.2017.00886. eCollection 2017.

Christopher M, Mace E, Jordan D, Rodgers D, McGowan P, Delacy I, Banks P, Sheppard J, Butler D, Poulsen D (2007) Applications of pedigree-based genome mapping in wheat and barley breeding programs. Euphytica 154 (3):307-316. doi:10.1007/s10681-006-9199-z

de Souza BJR, Perez PH, Bauer FC, Raetano CG, Neto PHW, Garcia LC (2014) Adjuvants for spraying of fungicides in wheat. Cienc Rural 44 (8):1398-1403

Joosen RVL, Ligterink W, Hilhorst HWM, Keurentjes JJB (2009) Advances in Genetical Genomics of Plants. Current Genomics 10 (8):540-549. doi:10.2174/138920209789503914

Kadar R, Moldovan V (2003) Achievement by breeding of winter wheat varieties with improved bread-making quality. Cereal Res Commun 31 (1-2):89-95

Kirigwi FM, van Ginkel M, Trethowan R, Sears RG, Rajaram S, Paulsen GM (2004) Evaluation of selection strategies for wheat adaptation across water regimes. Euphytica 135 (3):361-371. doi:Doi 10.1023/B:Euph.0000013375.66104.04

Kuchel H, Ye GY, Fox R, Jefferies S (2005) Genetic and economic analysis of a targeted marker-assisted wheat breeding strategy. Molecular Breeding 16 (1):67-78. doi:10.1007/s11032-005-4785-7

Kumar RS, Sakthivel N (2001) Exopolysaccharides of Xanthomonas pathovar strains that infect rice and wheat crops. Applied Microbiology and Biotechnology 55 (6):782-786

Kuti C, Lang L, Gulyas G, Karsai I, Meszaros K, Vida G, Bedo Z (2012) Bioinformatics Tool for Handling Molecular Data in Wheat Breeding. Cereal Res Commun 40 (4):573-582. doi:10.1556/Crc.40.2012.0009

Lev-Yadun S, Abbo S, Doebley J (2002) Wheat, rye, and barley on the cob? Nature biotechnology 20 (4):337-338. doi:DOI 10.1038/nbt0402-337b

Lupton F (2005) Advances in work on breeding wheat with improved grain quality in the twentieth century. Journal of Agricultural Science 143:113-116. doi:10.1017/S0021859604004617

Peng JHH, Sun DF, Nevo E (2011) Domestication evolution, genetics and genomics in wheat. Molecular Breeding 28 (3):281-301. doi:10.1007/s11032-011-9608-4

Reynolds M, Bonnett D, Chapman SC, Furbank RT, Manes Y, Mather DE, Parry MAJ (2011) Raising yield potential of wheat. I. Overview of a consortium

approach and breeding strategies. Journal of Experimental Botany 62 (2):439-452. doi:10.1093/jxb/erq311

Reynolds M, Foulkes J, Furbank R, Griffiths S, King J, Murchie E, Parry M, Slafer G (2012) Achieving yield gains in wheat. Plant Cell and Environment 35 (10):1799-1823. doi:10.1111/j.1365-3040.2012.02588.x

Tang D, Wu W, Li W, Lu H, Worland AJ (2000) Mapping of QTLs conferring resistance to bacterial leaf streak in rice. Theoretical and Applied Genetics 101 (1-2):286-291. doi:10.1007/s001220051481

Tillman BL, Kursell WS, Harrison SA, Russin JS (1999) Yield loss caused by bacterial streak in winter wheat. Plant Disease 83 (7):609-614. doi:Doi 10.1094/Pdis.1999.83.7.609

Varshney RK, Langridge P, Graner A (2007a) Application of genomics to molecular breeding of wheat and barley. Adv Genet 58:121-+. doi:10.1016/S0065-2660(06)58005-8

Varshney RK, Langridge P, Graner A (2007b) Application of genomics to molecular breeding of wheat and barley. In: Hall JC, Dunlap JC, Friedmann T, VanHeyningen V (eds) Advances in Genetics, vol 58. Advances in Genetics. pp 121-+. doi:10.1016/s0065-2660(06)58005-8

Wrigley CW (1994) Developing Better Strategies to Improve Grain Quality for Wheat. Australian Journal of Agricultural Research 45 (1):1-17. doi:Doi 10.1071/Ar9940001

Wu H, Pratley J, Lemerle D, Haig T (2001) Allelopathy in wheat (Triticum aestivum). Annals of Applied Biology 139 (1):1-9. doi:DOI 10.1111/j.1744-7348.2001.tb00124.x

CHAPTER – 2

LITERATURE REVIEW

1. Wheat

1.1. Wheat background

Wheat is the most principle food for humans, supplying more than 20% of total consumed calories. 17% of the world crop cultivated land constitutes wheat. Although optimal conditions are required for wheat to perform at its highest potential, it is a broadly adapted crop in terms of latitude, temperature, soil moisture and precipitation (Peng et al. 2011). Most wheat is used in the country in which it is produced and only a few countries produce more than they need (Salazar et al. 1996). The United States is the world's leading exporter, with about two-thirds of the crop being exported annually. Other important exporters are Canada, Australia, the European Community, Russia, India, and Argentina (USDA 2017).

Wheat belongs to the Poaceae family and Pooideae subfamily of grasses, with a center of origin in the Levant region of the Near East. Wheat is an allopolyploid species. The two predominantly cultivated forms are hexaploid bread wheat (*Triticum aestivum*, 2n=6x=42, genomes AuAuBBDD) and tetraploid pasta wheat (*T. durum*, 2n=4x=28, genomes AuAuBB) (Terzi et al. 2007; Peng et al. 2011). Wild hybridization of diploid wheat (*T. urartu*, genome AuAu) and goatgrass (*Aegilops speltoides*, genome SS), a close ancestor of the

BB genome, generated wild emmer wheat (*T. dicoccoides*, AuAuBB) (Simkova et al. 2011). Through selection, a cultivated emmer (*T. dicoccum*, genomes AABB) was created and its hybridization with *A. tauschii* (genome DD) produced *T. spelta* (genomes AuAuBBDD). Subsequent natural mutation of free-threshing ears in both emmer and spelt resulted in the rise of *T. durum* and *T. aestivum*, respectively (Peng et al. 2011).

1.2. Production uses and economic importance of wheat

Two species of wheat make up about 90% of the world crop: common bread wheat (*Triticum aestivum L.*) and durum wheat (*T. durum Desf.*) (Feng et al. 2004). Wheat is also classified on the basis of the time of the year in which it is grown, i.e., winter or spring (although wheat as a whole is a cool season crop), seed color, i.e., red or white, and on the protein content of the seed, i.e., 11 – 12% for hard wheat, 6 – 11% for soft wheat (Tudor et al. 2017). Winter wheat requires a certain period of cold temperature (vernalization) before it will produce grain whereas spring wheat does not. Almost all wheat is processed for human consumption. Hard wheat is used primarily for making bread whereas, soft wheat is used to make cakes, cookies etc. Durum wheat is used to make pasta products because of the unique coarse nature of its ground kernel (Cui et al. 2009).

1.3. Breeding and the importance of wheat improvement

Bread wheat is an almost entirely self – pollinating, allohexaploid plant (Paux et al. 2012; Jia et al. 2013). Thus, nearly all wheat cultivars are grown as homozygous lines although experimental and commercial methods of producing

hybrids exist. As of 1984, less than 0.1% of the total acreage of wheat in the United States was sown to hybrid wheat (Marshall et al. 2001).

The methods employed in wheat breeding programs are those common to self – pollinated crops (Smale et al. 2008). The major objective of wheat breeding is increased yield. Depending on the specific area, this objective may be met by improved disease resistance, better adaptation to local environmental conditions, and/or increases in genetic yield/gain potential, among others. Although there is some discussion as to whether genetic yield potentials have been reached, it is generally agreed that there is sufficient variability in wheat germplasm to expect more gains in yield due to genetic improvement (Stamp et al. 2014). Adding to the inherent variability of cultivated wheat is the use of wide – hybridization to transfer genes from several different species to wheat. In most cases, such hybridizations are used to transfer simply inherited traits such as pest or pathogen resistance (Cox et al. 1994).

2. Disease resistance in wheat

Developing disease-resistant cultivars is a very important component of any breeding program. Resistance is considered to be the most effective and economical method to protect crops from diseases caused by pathogens including bacteria and fungi (Faris et al. 1999). Inheritance of resistance to diseases can be either quantitative or qualitative. In wheat, for example, reaction to the rust, smut and powdery mildew pathogens, which cause some of the most destructive diseases of wheat, is inherited qualitatively (Keller et al. 2001). The pathogen population is made

up of genetically distinct races with each one capable of causing disease on some, but not all, members of the host population.

Many of the aforementioned diseases are caused by fungi which can be treated or have treatment measures using fungicides. With an increase in protection parameters to fungal diseases in wheat, bacterial diseases have become more prevalent with no control measure. Using bactericides to control these bacterial diseases can become an economically difficult task. The only economical way to control bacterial diseases is by identifying disease resistant cultivars and genetic markers that can be used in marker-assisted selection to breed resistant genotypes (ElAttari et al. 1996). Since the last decade or so in the United States Northern Great Plains, Bacterial Leaf Streak (BLS) has become a prominent threat to wheat production resulting in high yield losses.

3. Bacterial leaf streak of wheat

3.1. History, symptomology and pathogen taxonomy

The first report of bacterial leaf streak of cereals (barley, wheat, rye and spelt) was published in 1916. It was found that a monotrichous rod, yellow in culture caused a blight of these crops (Wonni et al. 2011). The typical BLS symptoms include water-soaked leaf streaks, blackening of the chaff and dark lesions on the peduncles, all of which normally do not occur until after the boot stage. Later research has shown that black chaff can be caused by other factors and that the most characteristic symptoms of the disease are the leaf streaks (Silva et al. 2010). The pathogen organism was originally named *Bacterium*

translucens and was later renamed to *Bacterium translucens var. undulosum*. In cross-inoculation experiments, strains of *Xanthomonas* isolated from wheat, rye and triticale were pathogenic to these three hosts and also barley but to a lesser extent. However, strains isolated from barley were pathogenic primarily on barley (El Attari et al. 1998). The genus name Bacterium was changed to *Xanthomonas* and later, over 100 different *Xanthomonas* species were condensed into one species: *Xanthomonas campestris* (Bianco et al. 2016). At the same time, a system of pathovars was adopted changing the specific epithet to the rank of pathovar. Studies report that wheat is attached by three pathovars of *Xanthomonas campestris, pvs. cerealis, undulosa,* and *translucens* (Fayette et al. 2013). However, on the basis of current evidence, it appears that the organism that attacks wheat has a wide host range and is called *Xanthomonas campestris pv. translucens*.

3.2. The lifecycle of *Xanthomonas campestris pv. translucens*

Xanthomonas campestris pv. translucens (*Xct*) survives the winter or summer in several ways. Infected seed is believed to be one of the major means by which *Xct* is disseminated (Zhao and Orser 1990). Although infected seed plays a role in the long-range dissemination of *Xct*, some of the first investigators rarely if ever saw disease in the resulting seedlings. In addition, no differences in the bacterial streak were observed between plots from infected seed and those from non – infected seed. Interestingly, it has been found that 25 to 40% of seedlings resulting from seed inoculated with both *Xct*. showed symptoms of the bacterial streak (Mellano and Cooksey 1988). Actual transmission from infested

seed lots in Montana was found to be less than 2% and was dependent on the level of seed lot infection, which ranged from 0 to 95%. Laboratory seed washing assays in Idaho found that about 1000 colony forming units (cfu) of *Xct* were needed in order for black chaff (or leaf streak) to develop in the field. Adding to the evidence for seed transmission is circumstantial evidence that the disease was apparently controlled for some years with organic mercury seed treatments. In addition to seed infection, the pathogen may pass the winter or summer in host debris, soil, or on weeds and other crops. However, in Arkansas, the pathogen was not found in any of these places *Xct* enters the wheat plant through stomata or wounds (Milus and Chalkley 1994). Recent evidence also indicates that the bacterium grows epiphytically on the plant surface. When on the leaf surface, the pathogen caused frost damage and greater leaf streak severity on wheat in the growth chamber. Thus, frost damage could be another mode of entry for *Xct* into the plant. Once *Xct* has infected the wheat plant, it can spread to other plants by driving winds and splash rains and possibly aphids (Kawahara and Obata 1998). The pathogen is capable of spreading about $28m^2$ within 39 days from a single plant (Stromberg et al. 1999).

3.3. Yield losses by bacterial leaf streak in wheat

Estimates of crop loss caused by bacterial leaf streak in sprinkler irrigated fields in Idaho are as high as 30% to 40% (Afolabi et al. 2014). Several investigators found that leaf streak decreases test weight. In Minnesota, $500^{-kernel}$ weight and seed plumpness were inversely correlated with disease severity on flag leaves in wheat and barley (Duveiller and Maraite 1993). On the other hand,

data from yield loss studies in Louisiana have not indicated a significant reduction in grain test weight. In addition, if the heads are attacked they may become sterile (Tillman and Harrison 1996). Yield loss estimates from single tiller studies in Mexico indicate that 11 – 29% of the potential grain weight/spike may be lost given that 50% of the flag leaf area is diseased. In two out of three years, in the same study, the number of grains/spike decreased as leaf streak severity on flag leaves increased (Tillman et al. 1996).

3.4. Control of *Xct* in wheat

Evidently, the bacterial leaf streak was controlled for some years with mercuric chloride seed treatments. Curtailment of these treatments due to the toxicity of mercury to humans may be responsible for the recent epidemics of bacterial leaf streak on wheat in the United States (Tillman and Harrison 1996).

Early investigators suggested plowing under crop residues and destroying perennial weeds that may harbor Xct as well as using clean seed as possible control measures. In Arkansas, the pathogen apparently does not over – summer in crop debris or on weeds so this may not be an effective control measure in the Southern USA (UNL CropWatch).

Other possible seed treatment chemicals have been tested for activity against *Xct* in Idaho. Of eight compounds tested, only acidified cupric acetate controlled *Xct* on seed. However, it also adversely affected seed germination and plant stand (Tillman et al. 1996). Seed treatments can be an effective means of control in the absence of other sources of inoculum. The use of pathogen-free seed is also a possible control method. Currently, no chemical, for either seed or

foliar application, is recommended for managing bacterial leaf streak in wheat. Given the current lack of practical chemical control, it is likely that cultivar resistance will play an important role in the control of bacterial streak.

3.5. Screening for resistance to bacterial leaf streak in wheat

Many methods have been used to artificially inoculate wheat plants with *Xct* in the field and greenhouse. These methods include spraying plants with bacterial suspensions, vacuum infiltration, rubbing a suspension on the leaf with fingers, injection of a suspension using a needless syringe, piercing the leaf with a needle and flooding with a bacterial suspension, and mowing off the tops of the plants and spraying with a suspension (Alizadeh et al. 1994). In Louisiana, a greenhouse study indicated that misting plants with a bacterial suspension at Feekes growth stage 7 gave the highest level of leaf streak. Unfortunately, greenhouse reactions to *Xct* inoculation failed to correlate well with field reaction in barley and this appears to be true with wheat (Adhikari et al. 2012). A more recent technique uses disease reactions from a needless syringe inoculation technique to rank cultivars (Raja et al. 2010). In Mexico, field inoculation is accomplished in the summer season by spraying plants after the tillering stage (Feekes growth stage 3) with an inoculum mixture containing about 109 cfu^{-ml}. Measurement of disease is usually done on the flag leaves or adult plant stage (3^{rd} – 5^{th} leaf) and two guides have been published guides to aid researchers in estimating the leaf streak severity wheat (Adhikari et al. 2011).

4. Association mapping in plant breeding

4.1. Association mapping

Association mapping is based on linkage disequilibrium (LD), where correlations between alleles in a population occur as a result of non – random segregation at different loci, and though physical linkage may increase LD, LD is not necessarily due to physical linkage (Le Couviour et al. 2011). Association mapping (AM) is a complementary strategy to QTL mapping to identify associations between genotype and phenotype and takes advantage of this "historical" LD to identify marker – trait relationships (Varshney et al. 2007).

The basic objective of AM is to detect correlations between genotypes and phenotypes in a sample of unrelated individuals. This technique is been practiced in humans and animals due to the impracticality and non – feasibility of creating large segregating individuals or populations (Arif et al. 2012). Association mapping is more advantageous to traditional linkage mapping as has an increased speed of sampling allelic variation, uses increased mapping resolution and uses lesser computational resources (Purcell et al. 2003).

4.2. Sampling of allelic variation

Linkage mapping is restricted to sampling only the alleles differing between the two parents. In contrast, AM populations are generally comprised of a diverse collection of accessions and breeding lines, providing a greater number of alleles for sampling (Raghavan et al. 2017). For example, in an AM population of common wheat, the number of alleles averaged 4.8 per microsatellite locus.

An attractive feature of AM is that marker – trait associations can be studied in well-phenotyped germplasm pools and breeding populations of locally adapted varieties. Diverse populations of germplasm such as found in AM offer a greater number of alleles for sampling as a result of more recombination events present and greater genetic diversity as compared to populations of narrow germplasm (Yu et al. 2011). Comparatively, NILs offer greater resolution than either F2 or RIL mapping populations, however, all remain limited by the number of alleles that may be sampled. Association mapping is further advantageous for its application in populations of unrelated individuals, in contrast to related populations studied in QTL mapping (Brbaklic et al. 2015). Studying populations of unrelated individuals facilitates increased sampling of meiotic events, and provides the opportunity to identify novel alleles that may be contributing to a trait.

Further, large populations increase power by providing the opportunity to identify alleles at a higher frequency. Small sample sizes often cause reduced power, with higher levels of LD decay anticipated in smaller populations of low sequence diversity (Marone et al. 2012). Increasing sample size further facilitates increased power that may normally be reduced by interactions between alleles, such as those caused by epistasis, by allowing for interaction terms to be included in models.

4.3. Resolution of association mapping

Association mapping theoretically allows mapping with higher resolution than achieved using biparental crosses. The degree of resolution depends on the

extent of LD and higher resolution is expected when LD declines rapidly with increasing genetic distance (Zanke et al. 2014). Understanding the extent of LD in the genome is required prior to conducting AM studies as the extent of genotyping required increases with rapid LD decay. Marker availability may be a limiting factor, particularly if LD is low. The best genotyping method must be chosen on the basis of the specific requirements of the envisioned genotyping project, and the resources available (Wingen et al. 2017). Single nucleotide polymorphisms (SNPs) are preferred for genotyping as a result of their abundance, providing high marker densities for mapping. Technologies currently exist with the ability to genotype thousands of sites simultaneously (for example, Perlegen Sciences Inc. genotyping arrays, Affymetrix Inc. GeneChip arrays, and Illumina Inc. BeadArray technology coupled with the GoldenGate genotyping assay), however, they are not necessarily cost-effective for genotyping large panels with a modest number of SNPs (Borner et al. 2011). The majority of studies have found that simple sequence repeats (SSRs) or SNPs are the markers of choice when performing association studies, as a result of their ability to detect genetic variability. The high level of polymorphism that SSRs provide increases the power to detect LD and facilitates higher resolution mapping (Mochida et al. 2008).

4.4. Approaches in association mapping

Recently, several AM studies have been published on a variety of crops like wheat, potato, maize and rice (Varshney et al. 2007) provided support for the

potential of AM in barley with a number of the associations identified in their study in regions of QTL previously identified through linkage analysis.

Whole genome and candidate gene analysis are the two approaches employed while conducting an association mapping study. Whole genome scans are accomplished by saturating the genome with adequate marker coverage (Sabiel et al. 2017), in order to identify associations between markers and phenotypes of interest. This approach is best suited for situations in which the availability of markers is a limiting factor or when the linkage extends for large distances, thus allowing for identification of potential candidate regions associated with the trait of interest. For example, (Cheung et al. 1992) estimated LD to extend approximately 10 cM in a collection of barley cultivars, comparatively greater than other inbreeding crop species. The high level of LD in their study was not conducive to fine resolution mapping but was useful for identifying regions which may be the subject of further fine mapping experiments. If LD decays too rapidly, the number of markers required to conduct genome-wide AM analysis increases significantly, resulting in AM focused on a candidate gene as an alternative approach for attaining high resolution. Candidate genes that have been shown or are suspected to have a functional role in the expression of a phenotype of interest can be used in AM studies where allelic variants are associated with phenotypic variation (Zanke et al. 2014). In cases where LD among single nucleotide polymorphisms (SNPs) within the gene decays rapidly, AM could be used to identify the causal molecular polymorphism(s) responsible for trait differences. 92 maize inbred

lines were analyzed using a candidate gene approach in which SNPs in dwarf8 were identified and evaluated at the time of flowering (Gupta et al. 2010). In maize, molecular differences at Y1 were associated with phenotypic variation in grain carotenoid concentration and this gene has since been identified as the causal factor for elevated carotenoids in maize. However, an association of SNPs with a trait still requires verification, as the SNP could be in disequilibrium with the causal factor, particularly if LD is high in the genomic region surrounding the gene. Thus candidate gene approaches are generally utilized to eliminate putative candidates for detailed functional studies. For example, a candidate gene approach was useful at eliminating three of eight candidates in a 70 kb region conferring resistance to *Xanthomonas oryzae* (Singh et al. 2000).

5. Genomic selection in plant breeding

5.1. Genomic prediction

Unlike QTL mapping and associated MAS techniques, genomic prediction methods attempt to predict phenotypes utilizing all available SNP marker data collected from a population, using one of many possible statistical models to predict the marker – trait associations in a data-driven way (Plaha and Sethi 1993). The accuracy of genomic prediction relies on an appropriate choice of a statistical model to capture the relationship between the genetic architecture of a trait and the underlying marker calls in a panel of high-density marker data. It is likely that the best statistical model for genomic prediction is dependent on the genetic architecture of the predicted trait (Randhawa et al. 2013). From a

mathematical perspective, models incorporating interactions between marker features have the capacity to achieve higher accuracy by capturing non-additive effects.

Alternative prediction methods continue to be an active area of research in plant and animal breeding. Once an accurate and predictive model of a QTL is discovered and an SNP marker assay has been conducted on an individual, it is trivial to convert the underlying predictions into a selection index (Lagudah et al. 2001). If the predictive model is selected such that it captures only additive effects, the resulting predictions can be considered to be an estimate of the breeding value of the assayed individual.

5.2. Choosing a genomic prediction model

Genomic prediction presents a distinct mathematical challenge compared to MAS. When conducting MAS, a large number of individual's n are evaluated at a comparatively smaller number of loci p. In a general sense, this corresponds to solving an overdetermined system of linear equations (Michel et al. 2017). The large family of regression techniques that minimize a least-squares loss function is well behaved on overdetermined systems. Genomic prediction is characterized by the opposite scenario where $n < p$. Typically a smaller number of individuals are genotyped at a larger number of marker loci. These problems can be solved using least squares regression, but also require that a regularization penalty is included in the calculations in addition to the least-squares loss function that is used to select between possible solutions to the underdetermined system (Heffner et al. 2011).

There are many forms of regularization. Perhaps the best known is L2 regularization, which penalizes large regression coefficients in the least squares regression problems. This results in a trained model that tends to place a small coefficient on all available input features.

Ridge regression is an example of an L2 regularized ordinary least squares regression. L1 regularization is another common form in which the sum of regression coefficients are penalized. As a result, L1 regularization tends to produce solutions that set non-informative feature's coefficients to zero. Least absolute shrinkage and selection operator (LASSO) regression is an L1 regularized ordinary least squares regression. Different regularization techniques such as L1 and L2 regularization have a relationship to the genetic architecture of the trait they were used to predict. If a trait is associated with many small effect markers, models incorporating L2 regularization are likely to perform better than unregularized models.

Classical MAS traits with a small number of large effect markers may be best predicted by algorithms incorporating L1 regularization. Traits falling somewhere in between may do well with models incorporating a combination of L1 and L2 regularization such as elastic net regression. A wide variety of regularization techniques exist. Some are broadly used and simple to reason about like L1 and L2 regularization. Others are applicable only to certain classes of mathematical models such as assumed prior distributions in Bayesian regression methods.

When choosing a tool for genomic prediction, it is critical to evaluate the available regularization techniques with multiple prediction methods in a data-driven way. These comparisons will ideally identify a single best model with zero or more regularization techniques which can be used to make accurate predictions for the traits of interest (Poland et al. 2012).

5.3. Genomic selection in a breeding program

Genomic selection is practiced by all major plant breeding programs today. Typically, this is accomplished by increasing the number of progeny evaluated early in a breeding program and practicing intense selection based on genomic prediction values (Uauy 2017). It is now feasible to phenotypically evaluate a randomly selected subset of a cohort of progeny while genotyping the entire cohort.

It is then trivial to build a genomic prediction model from the subset of the progeny with both phenotypic and genotypic data and use the resulting model to make selections for the entire cohort. One advantage to using genomic prediction methods over MAS is that the patterns in genotypic data that are used for selections naturally regenerate new haplotypes after each recombination event (Guzman et al. 2016). It has been hypothesized that selecting directly on this information rather than on phenotypic measurements alone may help maintain diversity in a breeding program. Other work using either a theoretical high-investment maize breeding program or a low-investment winter wheat breeding program has demonstrated that genetic gain per year could be improved by utilizing genomic selection rather than MAS (Miedaner and Korzun 2012).

Beyond maintaining genetic diversity and increasing genetic gain in a breeding program, genomic selection may also allow breeders to characterize the performance of allele combinations in environments that are critical to a target market but are rarely observed. (Heffner et al. 2011) suggest that by capturing genotype by environment interaction by modeling genotype performance in severe weather years it may be possible to characterize lines in non-severe years while still enabling selections for traits such as severe weather hardiness or severe drought tolerance.

The adoption of genomic selection and the use of GEBVs in commercial plant breeding has been rapidly increasing as molecular marker technology such as dense marker arrays has become less expensive. (Heffner et al. 2011) offers the possibility that breeding programs may eventually transition to using genomic selection as a primary selection method in a breeding program with phenotypic evaluation, at least early in a breeding program, used primarily for training statistical models of genotypic performance or updating models to improve predictions on new genotypes or recombination. These same models could then be used to identify parent candidates without performing expensive field trials. Yield trials would only be strictly needed at the end of a breeding program prior to verifying general agronomic performance prior to cultivar release.

The future state described by (Gupta et al. 2005) where breeding program selections are driven primarily by data from predictive models rather than direct measurements is not unlike the transformation that is currently underway in other

industries. Both of these transformations are driven by the growth of data science as a field, though the moniker itself has not been adopted as widely as the techniques it encompasses. Small percentage improvements in accuracy could generate much larger improvements in genetic gain over the lifetime of a breeding program (Poland et al. 2012).

6. Literature cited

Adhikari TB, Gurung S, Hansen JM, Jackson EW, Bonman JM (2012) Association Mapping of Quantitative Trait Loci in Spring Wheat Landraces Conferring Resistance to Bacterial Leaf Streak and Spot Blotch. Plant Genome 5 (1):1-16. doi:10.3835/plantgenome2011.12.0032

Adhikari TB, Jackson EW, Gurung S, Hansen JM, Bonman JM (2011) Association Mapping of Quantitative Resistance to Phaeosphaeria nodorum in Spring Wheat Landraces from the USDA National Small Grains Collection. Phytopathology 101 (11):1301-1310. doi:10.1094/Phyto-03-11-0076

Afolabi O, Milan B, Amoussa R, Koebnik R, Poulin L, Szurek B, Habarugira G, Bigirimana J, Silue D (2014) First Report of Xanthomonas oryzae pv. oryzicola Causing Bacterial Leaf Streak of Rice in Burundi. Plant Disease 98 (10):1426-1426. doi:10.1094/Pdis-05-14-0504-Pdn

Alizadeh A, Benetti V, Sarrafi A, Barrault G, Albertini L (1994) GENETIC-ANALYSIS FOR PARTIAL RESISTANCE TO AN IRANIAN STRAIN OF BACTERIAL LEAF STREAK (XANTHOMONAS-CAMPESTRIS PV HORDEI) IN BARLEY. Plant Breeding 113 (4):323-326. doi:10.1111/j.1439-0523.1994.tb00743.x

Arif MAR, Neumann K, Nagel M, Kobiljski B, Lohwasser U, Borner A (2012) An association mapping analysis of dormancy and pre-harvest sprouting in wheat. Euphytica 188 (3):409-417. doi:10.1007/s10681-012-0705-1

Bianco MI, Toum L, Yaryura PM, Mielnichuk N, Gudesblat GE, Roeschlin R, Marano MR, Lelpi L, Vojnov AA (2016) Xanthan Pyruvilation Is Essential for the Virulence of Xanthomonas campestris pv. campestris. Mol Plant Microbe In 29 (9):688-699. doi:10.1094/Mpmi-06-16-0106-R

Borner A, Neumann K, Kobiljski B (2011) Wheat Genetic Resources - How to Exploit? Czech Journal of Genetics and Plant Breeding 47:S43-S48

Brbaklic L, Trkulja D, Kondic-Spika A, Mikic S, Tomicic M, Kobiljski B (2015) Determination of Population Structure of Wheat Core Collection for Association Mapping. Cereal Res Commun 43 (1):22-28. doi:10.1556/Crc.2014.0027

Cheung WY, Moore G, Money TA, Gale MD (1992) Hpaii Library Indicates Methylation-Free Islands in Wheat and Barley. Theoretical and Applied Genetics 84 (5-6):739-746

Cox TS, Raupp WJ, Gill BS (1994) Leaf Rust-Resistance Genes Lr41, Lr42, and Lr43 Transferred from Triticum-Tauschii to Common Wheat. Crop Science 34 (2):339-343

Cui ZL, Zhang FS, Dou ZX, Miao Y, Sun QP, Chen XP, Li JL, Ye YL, Yang ZP, Zhang Q, Liu CS, Huang SM (2009) Regional Evaluation of Critical Nitrogen Concentrations in Winter Wheat Production of the North China Plain. Agron J 101 (1):159-166. doi:10.2134/agronj2008.0102

Duveiller E, Maraite H (1993) Study on Yield Loss Due to Xanthomonas-Campestris Pv Undulosa in Wheat under High Rainfall Temperate Conditions. Zeitschrift Fur Pflanzenkrankheiten Und Pflanzenschutz-Journal of Plant Diseases and Protection 100 (5):453-459

El Attari H, Rebai A, Hayes PM, Barrault G, Dechamp-Guillaume G, Sarrafi A (1998) Potential of doubled-haploid lines and localization of quantitative trait loci (QTL) for partial resistance to bacterial leaf streak (Xanthomonas campestris pv. hordei) in barley. Theoretical and Applied Genetics 96 (1):95-100

ElAttari H, Sarrafi A, Garrigues S, DechampGuillaume S, Barrault G (1996) Diallel analysis of partial resistance to an Iranian strain of bacterial leaf streak (Xanthomonas campestris pv cerealis) in wheat. Plant Pathology 45 (6):1134-1138. doi:10.1046/j.1365-3059.1996.d01-197.x

Faris JD, Li WL, Liu DJ, Chen PD, Gill BS (1999) Candidate gene analysis of quantitative disease resistance in wheat. Theoretical and Applied Genetics 98 (2):219-225

Fayette J, Raid R, Roberts P, Jones J (2013) Detection and characterization of Xanthomonas campestris pv. vitians strains. Phytopathology 103 (5):3-4

Feng DS, Xia GM, Zhao SY, Chen FG (2004) Two quality-associated HMW glutenin subunits in a somatic hybrid line between Triticum aestivum and Agropyron elongatum. Theoretical and Applied Genetics 110 (1):136-144. doi:10.1007/s00122-004-1810-x

Gupta PK, Kulwal PL, Rustgi S (2005) Wheat cytogenetics in the genomics era and its relevance to breeding. Cytogenetic and Genome Research 109 (1-3):315-327. doi:10.1159/000082415

Gupta PK, Langridge P, Mir RR (2010) Marker-assisted wheat breeding: present status and future possibilities. Molecular Breeding 26 (2):145-161. doi:10.1007/s11032-009-9359-7

Guzman C, Pena RJ, Singh R, Autrique E, Dreisigacker S, Crossa J, Rutkoski J, Poland J, Battenfield S (2016) Wheat quality improvement at CIMMYT and the use of genomic selection on it. Applied and Translational Genomics 11:3-8. doi:10.1016/j.atg.2016.10.004

Heffner EL, Jannink JL, Iwata H, Souza E, Sorrells ME (2011) Genomic Selection Accuracy for Grain Quality Traits in Biparental Wheat Populations. Crop Science 51 (6):2597-2606. doi:10.2135/cropsci2011.05.0253

Jia JZ, Zhao SC, Kong XY, Li YR, Zhao GY, He WM, Appels R, Pfeifer M, Tao Y, Zhang XY, Jing RL, Zhang C, Ma YZ, Gao LF, Gao C, Spannagl M, Mayer KFX, Li D, Pan SK, Zheng FY, Hu Q, Xia XC, Li JW, Liang QS, Chen J, Wicker T, Gou CY, Kuang HH, He GY, Luo YD, Keller B, Xia QJ, Lu P, Wang JY, Zou HF, Zhang RZ, Xu JY, Gao JL, Middleton C, Quan ZW, Liu GM, Wang J, Yang HM, Liu X, He ZH, Mao L, Wang J, Consor IWGS (2013) Aegilops tauschii draft genome sequence reveals a gene repertoire for wheat adaptation. Nature 496 (7443):91-95. doi:10.1038/nature12028

Kawahara H, Obata H (1998) Production of xanthan gum and ice-nucleating material from whey by Xanthomonas campestris pv. translucens (vol 49, pg 353, 1998). Applied Microbiology and Biotechnology 49 (6):800-800

Keller B, Stein N, Feuillet C (2001) Comparative genetics and disease resistance in wheat. Euphytica 119 (1-2):129-131. doi:Doi 10.1023/A:1017525901703

Lagudah ES, Dubcovsky J, Powell W (2001) Wheat genomics. Plant Physiology and Biochemistry 39 (3-4):335-344. doi:10.1016/s0981-9428(00)01233-x

Le Couviour F, Faure S, Poupard B, Flodrops Y, Dubreuil P, Praud S (2011) Analysis of genetic structure in a panel of elite wheat varieties and relevance for association mapping. Theoretical and Applied Genetics 123 (5):715-727. doi:10.1007/s00122-011-1621-9

Marone D, Laido G, Gadaleta A, Colasuonno P, Ficco DBM, Giancaspro A, Giove S, Panio G, Russo MA, De Vita P, Cattivelli L, Papa R, Blanco A, Mastrangelo AM (2012) A high-density consensus map of A and B wheat genomes. Theoretical and Applied Genetics 125 (8):1619-1638. doi:10.1007/s00122-012-1939-y

Marshall DR, Langridge P, Appels R (2001) Wheat breeding in the new century - Preface. Australian Journal of Agricultural Research 52 (11-12):I-Iv. doi:DOI 10.1071/ARv52n12_PR

Mellano VJ, Cooksey DA (1988) Development of Host Range Mutants of Xanthomonas campestris pv. translucens. Appl Environ Microbiol 54 (4):884-889

Michel S, Ametz C, Gungor H, Akgol B, Epure D, Grausgruber H, Loschenberger F, Buerstmayr H (2017) Genomic assisted selection for enhancing line breeding: merging genomic and phenotypic selection in winter wheat breeding programs with preliminary yield trials. Theoretical and Applied Genetics 130 (2):363-376. doi:10.1007/s00122-016-2818-8

Miedaner T, Korzun V (2012) Marker-Assisted Selection for Disease Resistance in Wheat and Barley Breeding. Phytopathology 102 (6):560-566. doi:10.1094/Phyto-05-11-0157

Milus EA, Chalkley DB (1994) Virulence of Xanthomonas-Campestris Pv Translucens on Selected Wheat Cultivars. Plant Disease 78 (6):612-615

Mochida K, Saisho D, Yoshida T, Sakurai T, Shinozaki K (2008) TriMEDB: A database to integrate transcribed markers and facilitate genetic studies of the tribe Triticeae. Bmc Plant Biology 8. doi:10.1186/1471-2229-8-72

Paux E, Sourdille P, Mackay I, Feuillet C (2012) Sequence-based marker development in wheat: Advances and applications to breeding. Biotechnology Advances 30 (5):1071-1088. doi:10.1016/j.biotechadv.2011.09.015

Peng JHH, Sun DF, Nevo E (2011) Domestication evolution, genetics and genomics in wheat. Molecular Breeding 28 (3):281-301. doi:10.1007/s11032-011-9608-4

Plaha P, Sethi GS (1993) Adaptive Advantage to 6r Chromosome of Rye in the Genomic Background of Bread Wheat. Cereal Res Commun 21 (2-3):239-245

Poland J, Endelman J, Dawson J, Rutkoski J, Wu SY, Manes Y, Dreisigacker S, Crossa J, Sanchez-Villeda H, Sorrells M, Jannink JL (2012) Genomic Selection in Wheat Breeding using Genotyping-by-Sequencing. Plant Genome 5 (3):103-113. doi:10.3835/plantgenome2012.06.0006

Purcell S, Cherny SS, Sham PC (2003) Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. Bioinformatics 19 (1):149-150. doi:DOI 10.1093/bioinformatics/19.1.149

Raghavan C, Mauleon R, Lacorte V, Jubay M, Zaw H, Bonifacio J, Singh RK, Huang BE, Leung H (2017) Approaches in Characterizing Genetic Structure and

Mapping in a Rice Multiparental Population. G3-Genes Genomes Genetics 7 (6):1721-1730. doi:10.1534/g3.117.042101

Raja NI, Rashid H, Khan MH, Chaudhry Z, Shah M, Bano A (2010) Screening of Local Wheat Varieties against Bacterial Leaf Streak Caused by Different Strains of Xanthomonas Translucens Pv. Undulosa (Xtu). Pakistan Journal of Botany 42 (3):1601-1612

Randhawa HS, Asif M, Pozniak C, Clarke JM, Graf RJ, Fox SL, Humphreys DG, Knox RE, DePauw RM, Singh AK, Cuthbert RD, Hucl P, Spaner D (2013) Application of molecular markers to wheat breeding in Canada. Plant Breeding 132 (5):458-471. doi:10.1111/pbr.12057

Sabiel SAI, Huang SS, Hu X, Ren XF, Fu CJ, Peng JH, Sun DF (2017) SNP-based association analysis for seedling traits in durum wheat (Triticum turgidum L. durum (Desf.)). Breeding Sci 67 (2):83-94. doi:10.1270/jsbbs.16074

Salazar GM, Moreno RO, Salazar GR, Carrillo ML (1996) Wheat production as affected by Seeding rate x Fertilization interaction. Cereal Res Commun 24 (2):231-237

Silva IT, Rodrigues FA, Oliveira JR, Pereira SC, Andrade CCL, Silveira PR, Conciecao MM (2010) Wheat Resistance to Bacterial Leaf Streak Mediated by Silicon. Journal of Phytopathology 158 (4):253-262. doi:10.1111/j.1439-0434.2009.01610.x

Simkova H, Safar J, Kubalakova M, Suchankova P, Cihalikova J, Robert-Quatre H, Azhaguvel P, Weng YQ, Peng JH, Lapitan NLV, Ma YQ, You FM, Luo MC, Bartos J, Dolezel J (2011) BAC Libraries from Wheat Chromosome 7D: Efficient Tool for Positional Cloning of Aphid Resistance Genes. J Biomed Biotechnol. doi:Artn 302543

10.1155/2011/302543

Singh RP, Nelson JC, Sorrells ME (2000) Mapping Yr28 and other genes for resistance to stripe rust in wheat. Crop Science 40 (4):1148-1155

Smale M, Singh J, Di Falco S, Zambrano P (2008) Wheat breeding, productivity and slow variety change: evidence from the Punjab of India after the Green Revolution. Aust J Agr Resour Ec 52 (4):419-432. doi:10.1111/j.1467-8489.2008.00435.x

Stamp P, Fossati D, Mascher F, Hund A (2014) The future of wheat breeding. Agrarforsch Schweiz+ 5 (7-8):286-291

Stromberg KD, Kinkel LL, Leonard KJ (1999) Relationship between phyllosphere population sizes of Xanthomonas translucens pv translucens and bacterial leaf

streak severity on wheat seedlings. Phytopathology 89 (2):131-135. doi:Doi 10.1094/Phyto.1999.89.2.131

Terzi V, Morcia C, Stanca AM, Kucera L, Fares C, Codianni P, Di Fonzo N, Faccioli P (2007) Assessment of genetic diversity in emmer (Triticum dicoccon Schrank) x durum wheat (Triticum durum Desf.) derived lines and their parents using mapped and unmapped molecular markers. Genetic Resources and Crop Evolution 54 (7):1613-1621. doi:10.1007/s10722-006-9173-6

Tillman BL, Harrison SA (1996) Heritability of resistance to bacterial streak in winter wheat. Crop Science 36 (2):412-418

Tillman BL, Harrison SA, Russin JS, Clark CA (1996) Relationship between bacterial streak and black chaff symptoms in winter wheat. Crop Science 36 (1):74-78

Tudor VC, Popa D, Gimbasanu GF (2017) THE ANALYSIS OF THE CULTIVATED AREAS, THE PRODUCTION AND THE SELLING PRICE FOR MAIZE CROPS DURING THE PRE- AND POST-ACCESSION PERIODS OF ROMANIA TO THE EUROPEAN UNION AND TRENDS OF EVOLUTION OF THESE INDICATORS. Scientific Papers-Series Management Economic Engineering in Agriculture and Rural Development 17 (2):387-394

Uauy C (2017) Wheat genomics comes of age. Current opinion in plant biology 36:142-148. doi:10.1016/j.pbi.2017.01.007

10.1016/j.pbi.2017.01.007. Epub 2017 Mar 24.

Varshney RK, Langridge P, Graner A (2007) Application of genomics to molecular breeding of wheat and barley. In: Hall JC, Dunlap JC, Friedmann T, VanHeyningen V (eds) Advances in Genetics, vol 58. Advances in Genetics. pp 121-+. doi:10.1016/s0065-2660(06)58005-8

Wingen LU, West C, Leverington-Waite M, Collier S, Orford S, Goram R, Yang CY, King J, Allen AM, Burridge A, Edwards KJ, Griffiths S (2017) Wheat Landrace Genome Diversity. Genetics 205 (4):1657-1676. doi:10.1534/genetics.116.194688

Wonni I, Ouedraogo L, Verdier V (2011) First Report of Bacterial Leaf Streak Caused by Xanthomonas oryzae pv. oryzicola on Rice in Burkina Faso. Plant Disease 95 (1):72-73. doi:10.1094/Pdis-08-10-0566

Yu LX, Lorenz A, Rutkoski J, Singh RP, Bhavani S, Huerta-Espino J, Sorrells ME (2011) Association mapping and gene-gene interaction for stem rust resistance in CIMMYT spring wheat germplasm. Theoretical and Applied Genetics 123 (8):1257-1268. doi:10.1007/s00122-011-1664-y

Zanke CD, Ling J, Plieske J, Kollers S, Ebmeyer E, Korzun V, Argillier O, Stiewe G, Hinze M, Neumann K, Ganal MW, Roder MS (2014) Whole Genome Association Mapping of Plant Height in Winter Wheat (Triticum aestivum L.). Plos One 9 (11). doi:ARTN e113287

10.1371/journal.pone.0113287

Zhao JL, Orser CS (1990) Conserved repetition in the ice nucleation gene inaX from Xanthomonas campestris pv. translucens. Mol Gen Genet 223 (1):163-166

CHAPTER – 3

MOLECULAR CHARACTERIZATION OF BACTERIAL LEAF STREAK (BLS) RESISTANCE IN HARD WINTER WHEAT

Abstract

Bacterial leaf streak (BLS) caused by *Xanthomonas campestris pv. translucens* is one of the major bacterial disease threatening to wheat production in the United States Northern Great Plains region. It is a sporadic but widespread disease of wheat that can cause significant loss depending on the location and year. Unlike fungal diseases, bacterial diseases cannot be effectively managed using chemicals and thus developing disease resistant cultivars would be the most economical control for BLS. Identification and characterization of genomic regions in wheat that confer resistance to BLS can be an effective way to mobilize resistance genes in wheat breeding. In this study, we evaluated a hard winter wheat association mapping panel (HWWAMP) of 300 hard winter wheat cultivars or advanced breeding lines representing the entire US hard winter wheat region for their reaction to BLS. Only four percent (11) of the lines showed a resistant reaction and eight percent (24) were moderately resistant to BLS whereas 88 percent (265) genotypes were moderately susceptible to susceptible. Genome-wide association analysis with 15,990 SNPs was conducted using an exponentially compressed mixed linear model and five genomic regions ($p < 0.001$) that regulate resistance to BLS were identified on chromosomes 1AL, 1BS, 3AL, 4AL, and 7AS. The QTLs *Q.bls.sdsu.1AL, Q.bls.sdsu.1BS, Q.bls.sdsu.3AL, Q.bls.sdsu.4AL* and *Q.bls.sdsu.7AS* explaining a total of

42 % of the variation. Comparative analysis with rice showed possible syntenic regions that harbor genes for bacterial leaf streak resistant. The 11 BLS resistant genotypes and SNP markers linked to the QTLs identified in our study could facilitate breeding for BLS resistance in wheat.

1. Introduction

Wheat (*Triticum aestivum L.*) is one of the major cereal crops worldwide. Hard winter wheat contributes 54% (USDA) of the total wheat production in the USA but is challenged by several biotic and abiotic factors which limit its yield potential. Bacterial leaf streak (BLS) caused by *Xanthomonas campestris pv. translucens* (Xct.), is emerging as a potential threat to wheat production in the Midwest of the United States in recent years because most of the commercial wheat varieties grown in the region appeared to be susceptible to this pathogen. BLS can lead yield loss up to 40% (Tillman et al. 1999) and can also affect protein content, degrading the grain quality (Shane at al. 1987). The pathogen is both residue and seed borne and may disperse long distance via wheat germplasm exchange (Tillman et al. 1999). Chemical control to manage this disease is neither economical nor environmentally friendly (Milus and Mirlohi 1993). Therefore, identifying genes or quantitative traits loci conferring BLS resistance and developing resistant cultivars is the best approach to manage BLS in wheat (ElAttari et al. 1996).

Association mapping (AM) is an effective strategy to detect quantitative trait loci (QTL) particularly in genetically diverse germplasm or wild relatives. In this approach, no prior information on the marker – trait associations are necessary, and

multiple loci can be readily identified (Bordes et al. 2011). Several robust statistical tools and modeling methods such as mixed linear models, Bayesian clustering, principal component analysis (PCA), and Q+K (terms that abbreviate gross structure based on given number of principal components [Q] and finer structure based on kinship [K]) mixed models have been developed and used to enhance our understanding of complex traits in animal and plant genetic systems.

Several AM studies have been conducted on wheat to characterize resistance to stem rust (Zhang et al. 2014; Yu et al. 2011; Muleta et al. 2017; Bariana et al. 2001; Letta et al. 2014), solid stem for sawfly (Varella et al. 2015) leaf rust (Singh et al. 2000; Turner et al. 2017; William et al. 2006) and several other diseases (Bentley et al. 2014; Dababat et al. 2016; Arif et al. 2012). A large assortment of wheat germplasm, including cultivars, breeding lines, and landraces have been evaluated for reaction to BLS in the field and/or under greenhouse conditions. Even though there was a high variation of reaction among the genotypes no high resistant or immune genotype was observed (Milus and Mirlohi 1995; Tillman et al. 1996; Kandel et al. 2012; Adhikari et al. 2012a). Five genes were reported, namely Bls1, Bls2, Bls3, Bls4 and Bls5 to condition resistance in three wheat cultivars (Duveiller et al. 1993). A couple studies on the genetics of BLS resistance in spring wheat have reported QTLs on chromosome 1A, 4A, 4B and 6A explaining a variation ranging from 1.4% to 2.6% (Adhikari et al. 2012b). In another study Kandel et al. 2015 reported QTLs on chromosomes 1B, 2A, 3B and 6A in a multifamily population of spring wheat. The significant SNPs in this study explained a variation ranging from 0.5% to 23%. There

has been no report on identification or characterization of BLS resistance in winter wheat.

The main objective of this study was to identify new sources and characterize BLS resistance in hard winter wheat and develop SNP markers for facilitating marker-assisted selection.

2. Material and Methods

2.1. Plant material

In the present study, we used a hard winter wheat association mapping panel (HWWAMP) of 300 winter wheat accessions developed under the USDA TCAP project (Guttieri et al. 2015). The geographic diversity of these 300 HWW accessions are provided in Appendix Figure 1. The experiments were conducted in both in the greenhouse for controlled conditions and also in the field at the South Dakota State University Agriculture Experimental Station at Aurora (SD) between the fall of 2015 and 2017.

2.2. Genotyping

The HWWAMP has been genotyped using the Infinium 90k iSELECT array (Illumina Inc. San Diego, CA) under the USDA-TCAP (Guttieri et al. 2017) and we obtained the genotype data from T3 Toolbox (https://triticeaetoolbox.org/wheat/). To avoid spurious marker-trait associations, SNP markers with MAF < 0.05 and missing data >10% were excluded from further analyses. The genetic and physical positions of SNP markers from the wheat 90 K array were obtained from the consensus map with 46,977 SNPs

developed using a combination of 8 mapping populations (Wang et al. 2014) and the International Wheat Genome Sequencing Consortium website (https://www.wheatgenome.org/).

2.3. Phenotypic evaluation and statistical analysis

2.3.1. Planting and experimental design

The entire experiment was conducted in greenhouse experiment constituting four replications (2015 – 2016) and field experiment constituting two replications (2016 – 2017). Both in the field and the greenhouse, spring wheat cv. Briggs (susceptible check: SC1) and germplasm accession SD1001 (pedigree: PFAU/MILAN//TROST susceptible check: SC2); SD52 (pedigree: CNO79//PF70354/MUS/3/PASTOR/4/BAV92*2/5/HAR311 resistant check: RC1) were included in the experiment as susceptible and moderately resistant checks, respectively.

Three seeds of each accession were planted in cones (Stuewe and Sons, Inc., Corvallis, OR). To the soil in each container, 2.5 g of multicote slow release commercial fertilizer with a 14–14–16 Nitrogen Phosphorous and Potassium composition (Sungro Horticulture Distribution Inc. Agawam, MA) was applied at the time of planting. Each cone consisted of three plants per replication. Each cone was considered as an experimental unit, and the third leaf of each plant was regarded as a sampling unit. In the greenhouse the entire experiment was performed under controlled temperature: 30/18°C diurnal cycle (day/night) with a 16h photoperiod and

relative humidity (>85%) (Silva et al. 2010) in a completely randomized design. In the field, the 300 wheat accessions were planted in a three feet rows in two replications along with resistant and susceptible checks.

2.3.2. Inoculum, infiltration and disease assessment

The highly virulent strain *isolate Xct -017* of Xct, was used in this study. The isolate was provided by Dr. Karl Glover of South Dakota State University. A fresh culture of the isolate *Xct -017* was initiated from a frozen at -80°C by streaking on a KG agar media (agar=20g; magnesium sulphate=1.5g; proteose peptone=20g; potassium phosphate=1.5g) (HiMediaLabs Inc. Mumbai, India). The bacterial cells were obtained from a 2 – day old culture by adding 20 – 25ml of distilled water into each plate and scrapping the surface using a flamed microscope slide. The inoculum density was adjusted to 3×10^8 colonies forming unit ml^{-1} using a turbidimeter (BIOLOG). In the greenhouse, infiltration was performed by injecting 10 to 15 μL of inoculum into a fully expanded third leaf using a needleless disposable syringe (Faris et al. 1996). The infiltrated areas were marked by a non – toxic sharpie, permanent marker and the plants were placed in trays with water. These plants were kept in a moisture chamber for 24 hours post-infiltration to enhance the infection process and then moved to a growth chamber (Silva et al. 2010). In the field when the plants reached the third leaf stage and then were inoculated using a blast sprayer. 30g of carborundum was added to 1gal of inoculum to create non – lethal wounds on the leafs.

2.3.3. Disease rating

BLS being a very complex disease required stringent and highly accurate phenotyping. We developed a disease rating scale of 1 – 5 under controlled environments. The rating of the plants was done after 14 days of infiltration. The initial area of infiltration was noted and then the increased area was also noted. These two distances were subtracted and based on this difference plants were classified into five different categories (Figure 3.1). This difference was used as the phenotype value for each line. 0 – 0.49cm increase was classified as Resistant (R) category 1, 0.5 – 0.9cm as Moderately Resistant (MR) category 2, 1 – 1.49cm Moderately Susceptible MS) category 3, 1.5 – 1.99cm was classified as Susceptible (S) category 4 and >=2cm was classified as Highly Susceptible (HS) category 5. The data was analyzed by standard ANOVA and META – R was used to calculate BLUEs (Vargas et al. 2013). In field evaluation, we rated the plants on the scale of continuous scale of 0 – 90% based on percent leaf area affected. 0 – 20% was considered as R, 20 – 40% as MR, 40 – 50% as MS, 50 – 70% as S and >70% as HS.

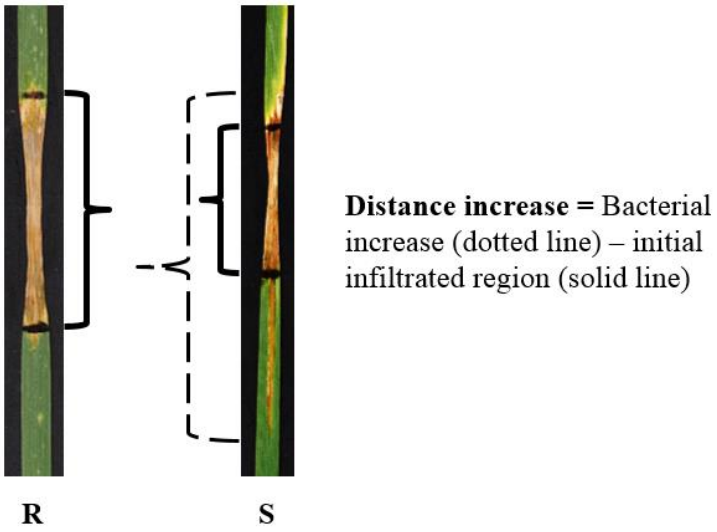

Figure 3.1. Diagrammatic representation of the categorization used to evaluate the lines in the Hard Winter Wheat Association Mapping Panel. The dotted lines represent the bacterial increase from the infiltrated region while the solid lines represent the initial region of infiltration.

2.3.4. Association mapping analysis

GAPIT software was used to analyze the marker properties, LD, principal component (PC) matrix, hierarchical clustering, and Q+K mixed model (Brbaklic et al. 2013) and TASSEL 5 was used for validating the results (Bradbury et al. 2007; Zhang et al. 2010). The ECMLM algorithm in GAPIT was used to analyze the marker-trait association which was then validated using other GAPIT algorithms like SUPER and CMLM. These results were also again validated using the GLM (General Linear Model) and MLM (Mixed Linear Model) (Zhang et al. 2010) algorithms in TASSEL. The ECMLM model is described by Henderson's notation (Li et al. 2014; Singh 2005).

$$y = X\beta + Zu + e \quad \ldots\ldots\ldots\ldots\ldots \quad \text{(Equation. 1)}$$

Where y represents a vector of the phenotype, β represents the unknown fixed effects like population structure and marker effects, u represents the unknown polygenic effects like kinship, X and Z are the incidence matrices for β and u respectively and e represents the error.

Data from the appropriate number of PCs and the allele sharing similarity matrix accounting for Q and K were fit into the linear model to associate numeric SNP genotypes with ordinal phenotypes. The "*Q*" parameter was accounted by the PC and the PC scores were used in the model as random components. Since dominance or additive effects were not assumed, association tests were performed by treating genotypes as categorical variables in ANOVA (dominance model) and as quantitative variables in regression (additive model) analyses. The negative $\log_{10}$ (p – value) conversion was used on all calculated p – values. QQ – plots assuming a uniform distribution of p – values under the null hypothesis of no QTLs were used to evaluate the models. Briefly, the observed p – values were plotted against the expected theoretical values (i.e. cumulative density function) for a uniform distribution. This is a standard methodology to evaluate the model's ability to control for spurious association (Hirschhorn and Daly 2005). These statistical analyses were also performed in R statistical software (https://www.r-project.org/).

2.3.5. Comparative analysis with rice

The wheat genome assembly used for the comparative analysis was IWGSC wheat genome assembly v1.0 (Mayer et al. 2014). The wheat genome sequence was repeat masked with RepeatMasker (http://www.repeatmasker.org/). BLAST was performed using a downloaded BLAST API onto a Linux based high-performance cluster (Altschul et al. 1990). blastn and blastx were the modules used within the BLAST API. The wheat and rice synteny was pictographically represented using a Perl based software CIRCOS (Krzywinski et al. 2009).

3. Results

3.1. Phenotypic analysis

A continuous range of response to BLS infiltration was observed in HWWAMP accessions. As expected, SC1, SC2, and MRC1 exhibited susceptible and moderately resistant reactions respectively. In the greenhouse experiment, the mean disease score was 3.2 whereas in the field the mean disease score was 47.4% (Appendix Table 1). Off 300 genotypes, 11 genotypes (3.6 %) exhibited a consistent resistant reaction to BLS in greenhouse and field experiments (score 1), whereas another 24 (8%) and 46 (15.3%) accessions demonstrated a moderate resistance response in greenhouse and field respectively (score 2). 152 (50.6%) and 95 (31.6%) lines showed a moderately susceptible reaction (score 3). Nearly 87 (29%) and 139 (46.3%) lines showed susceptible (score 4) response in greenhouse and field respectively, whereas 26 (8.6%) and 9 (3%) lines showed a

highly susceptible reaction (score 5) in the greenhouse and field experiments respectively (Figure 3.2). Analysis of variance (ANOVA) for BLS scores revealed significant differences among genotypes in GH and field experiments (P value = $1.6e^{-8}$). The correlation between the GH and field experiment was $r^2 = 0.62$.

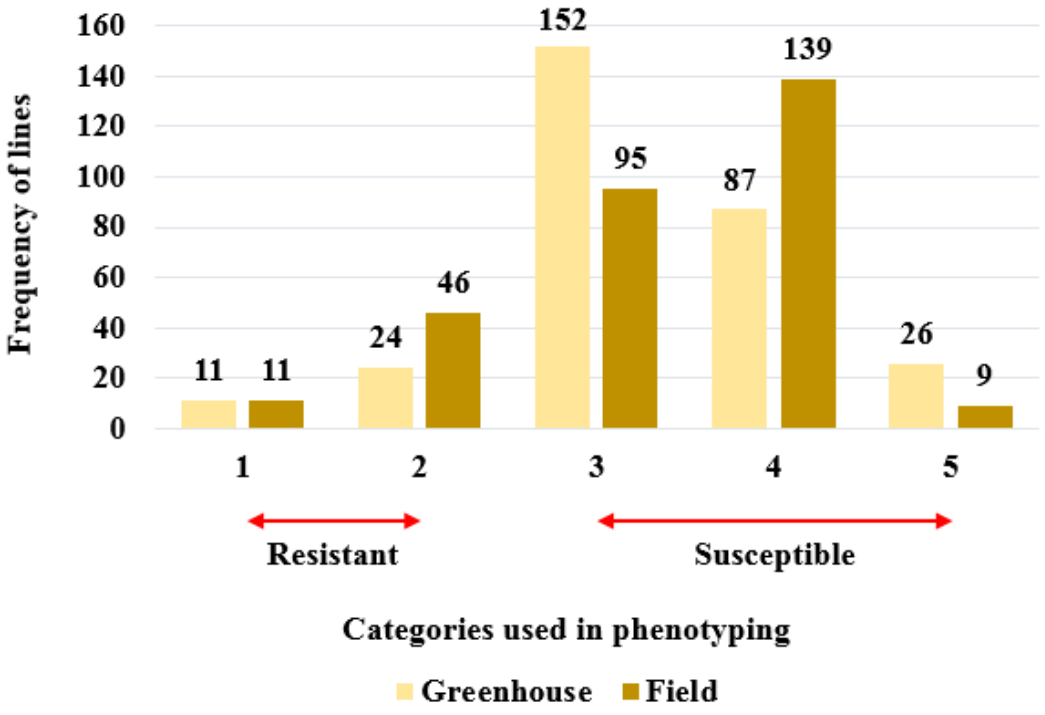

Figure 3.2. Distribution of the 300 genotypes of the HWWAMP into the 5 categories. The bars represent the mean response of the genotypes to BLS in both the greenhouse and the field (1 – R, 2 – MR, 3 – MS, 4 – S, 5 – HS).

3.2. Marker statistics and linkage disequilibrium

Linkage disequilibrium statistics (r^2) were calculated using the TASSEL program (Figure 3.3). Among 21,500 polymorphic SNP markers, 5,510 markers were eliminated as they had a MAFs of <0.05, greater than 20% missing

genotypic data and they didn't have any map positions. This resulted in a total of 15,990 markers which was used in the association analysis. The majority of SNP markers were distributed across wheat A and B genomes (40% and 50%, respectively) while the D genome had the fewest (10 %). The highest number of SNP markers was distributed on chromosome 5B followed by 1B, 6B, and 2B (Figure 3.4). The average number of SNP markers per chromosome was 750.

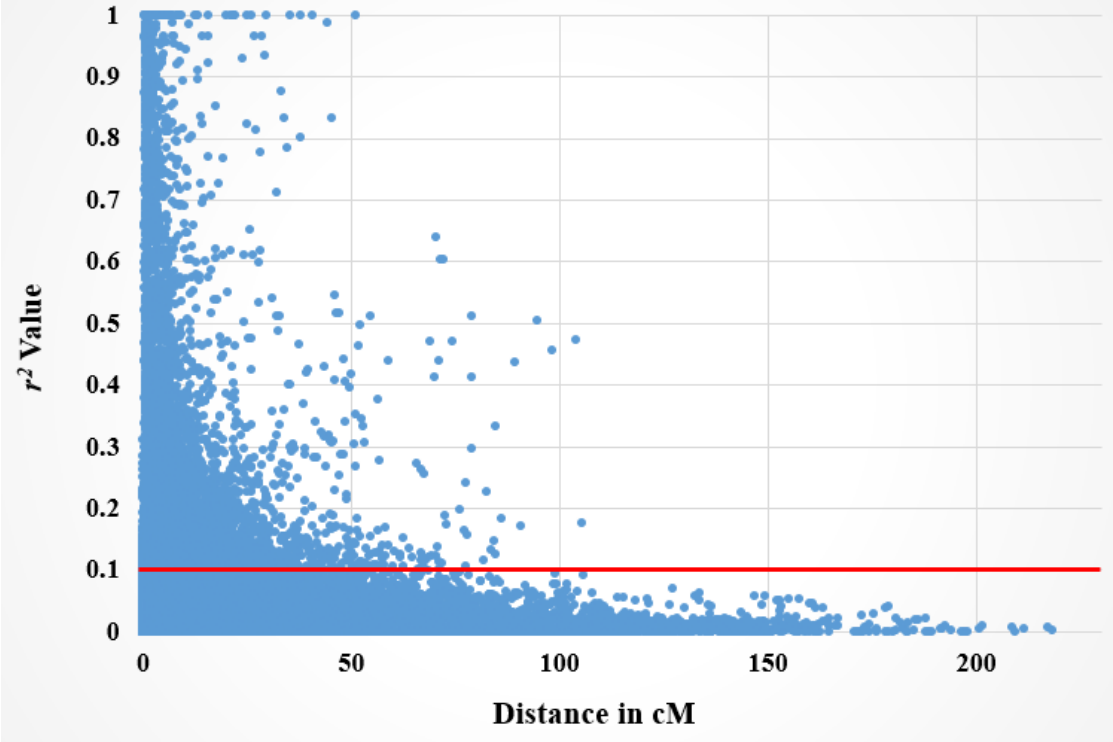

Figure 3.3. Genome – wide linkage disequilibrium (LD) decay plot with 15,990 markers on the HWWAMP. The red line indicates the LD value.

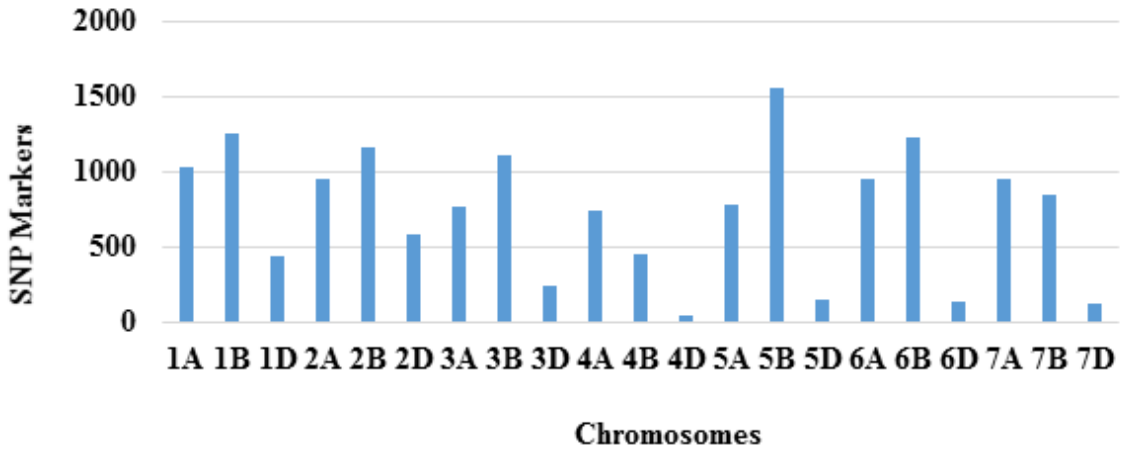

Figure 3.4. Distribution of 15,590 SNP markers across the 21 wheat chromosomes. Chromosomal locations and positions of SNP markers obtained from the wheat 90k consensus genetic map.

3.3. Association analysis of the QTLs associated with resistance to BLS

GWAS was carried out for each experiment (one greenhouse experiment: Figure 3.5.A and Field Experiment Figure 3.5.B) separately. To ascertain consensus a BLUE value for each genotype was obtained which was then used to perform an individual analysis (Figure 3.5.C). Several algorithms and models were tested to ascertain the consensus in the marker-trait associations (MTAs). The consensus QTLs were identified based on its significance in all experiments and by using multiple algorithms and methods. In the greenhouse experiment, a total of four significant regions were obtained each on chromosome 1A, 1B, 4A and 7A (Figure 3.5A). Whereas in the field data, we identified only three significant regions 1A, 4A, and 7A (Figure 3.5B). When BLUE values were used for GWAS we identified five genomic regions namely 1A, 1B, 3A, 4A, and 7A for the (Figure 3.5.C). A total of 38 significant MTAs across these five

chromosomes were identified to be linked to BLS resistance ($p < 0.001$) in the HWWAMP. These five genomic regions were further validated using GLM and MLM models in TASSEL and SUPER algorithms of GAPIT to ascertain consensus and significance. All five QTL, *Q.bls.sdsu.1AL, Q.bls.sdsu.1BS, Q.bls.sdsu.3AL, Q.bls.sdsu.4AL,* and *Q.bls.sdsu.7AS* were consistent across all the algorithms in TASSEL and GAPIT (Table 3.1). These genomic regions showed significant association with 3, 2, 1, 10 and 6 significant SNPs respectively. The most significant SNPs on the chromosome 1A, 1B, 3A, 4A and 7A are BS00084995_51, Ku_c17846_363, IWA7541, IAAV1943 and tplb0032m13_1358 respectively. The QTLs *Q.bls.sdsu.1AL, Q.bls.sdsu.1BS, Q.bls.sdsu.3AL, Q.bls.sdsu.4AL* and *Q.bls.sdsu.7AS* explained 8.3%, 8.5%, 7.9%, 8.3% and 9.3% of the variation respectively (Table 3.1). These QTLs spanning regions was estimated to be 2.05cM (1AL), 3.5cM (1BS), 0.13cM (3AL), 2.41cM (4AL) and 1.53cM (71AL), corresponding to 11Mb, 4.2Mb, 60Mb, 20Mb and 6Mb respectively.

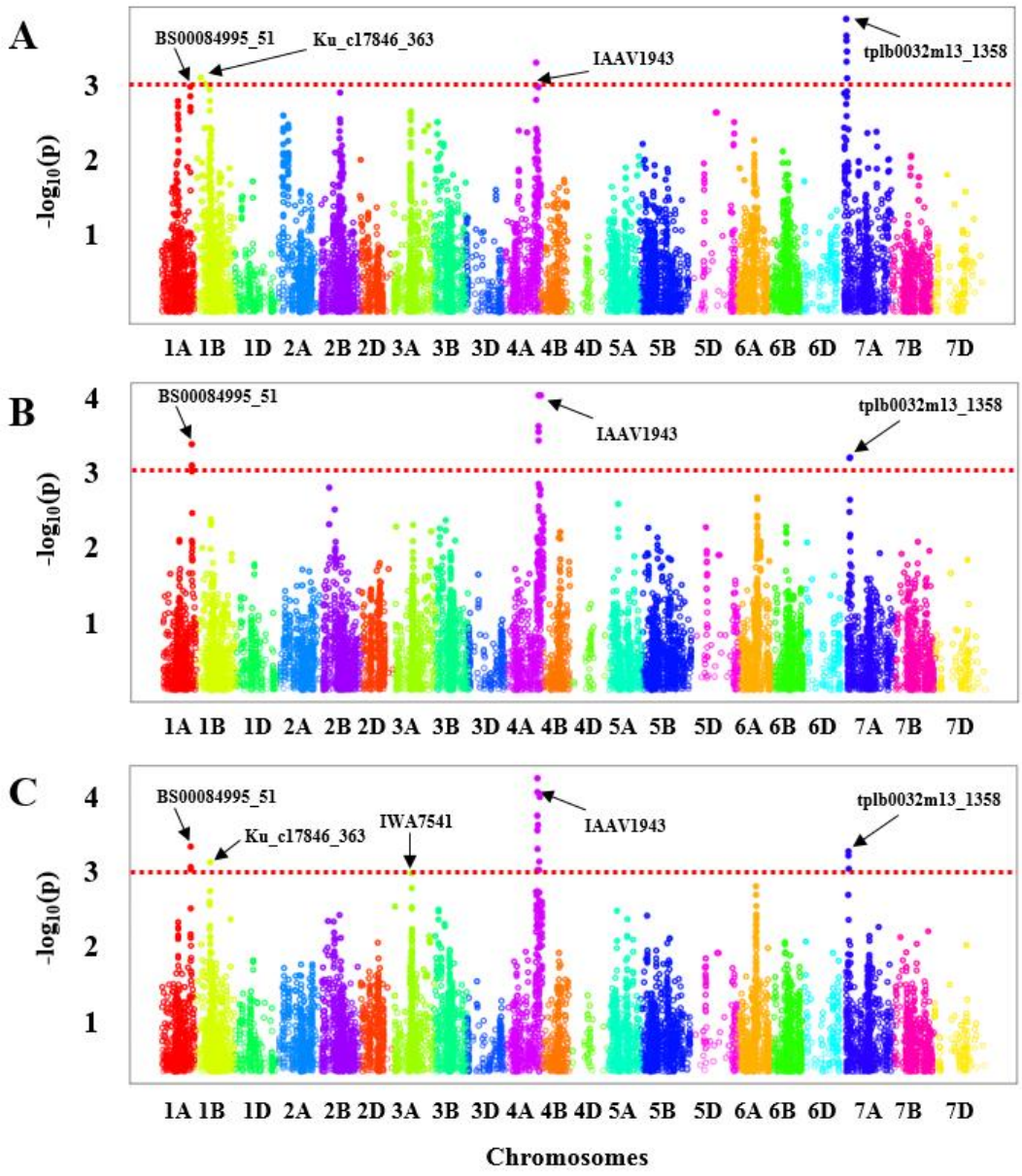

Figure 3.5. Manhattan plot of BLS reaction from GWAS using the ECMLM model. Manhattan plot with the $-\log_{10}$P – value of all SNP used in GWAS with 300 genotypes of the HWWAMP using ECMLM model. Greenhouse experiment (A), Field experiment 1 (C) and BLUE values (D). The red color line in the figure shows the threshold of $-\log_{10}$ (P – value) of three and all the significantly associated SNP markers are above the red line.

Table 3.1. Summary of significant SNPs linked to QTLs for BLS resistance detected from the BLUEs values of greenhouse and field evaluations**.**

QTL	Marker	cM†	Mb‼	Allele‡	p - value	R^2 %	Additive effect	T – test¥
Q.bls.sdsu.1AL	BS00084995_51	139.47	580.42	T	$1.16e^{-3}$	8.3	-0.20	$9.82e^{-6}$
Q.bls.sdsu.1BS	Ku_c17846_363	24.54	9.54	C	$8.91e^{-4}$	8.5	-0.26	$1.10e^{-3}$
Q.bls.sdsu.3AL	IWA7541	89.48	533.07	G	$1.31e^{-3}$	7.9	-0.22	$5.74e^{-3}$
Q.bls.sdsu.4AL	IAAV1943	144.37	726.44	T	$1.03e^{-4}$	8.8	-0.21	$5.66e^{-4}$
Q.bls.sdsu.7AS	tplb0032m13_1358	43.47	10.25	T	$2.53e^{-4}$	9.3	-0.22	$2.21e^{-3}$

† The cM location is from Wang et al 2014. ‼ The SNP position (Mb) was identified by BLAST on IWGSC RefSeq (https://www.wheatgenome.org/). ‡ Resistant allele. ¥ P – value obtained from the 5 – fold cross-validation.

3.4. 5 – fold cross-validation

A 5 – fold cross-validation was performed to ascertain the significance of the obtained SNP markers in the genomic regions. The entire HWWAMP was randomly split into five different parts without repetition and four parts were used for MTA and fifth part was used for validation of the significant SNPs correlated to BLS resistance. All five markers BS00084995_51, Ku_c17846_363, IWA7541, IAAV1943 and tplb0032m13_1358 of chromosome 1A, 1B, 3A, 4A and 7A respectively were significantly associated with respective QTLs. These SNPs were significant at a confidence interval of $\alpha = 0.01$ with the p – values of $9.8e^{-6}$, $1.1e^{-3}$, $5.7e^{-3}$, $1.4e^{-3}$ and $2.2e^{-3}$ respectively. From these results, it is evident that all the SNPs were significantly different for each allelic group they represent. These SNP markers could be useful in marker-assisted selection for BLS resistance. The sequence of these SNPs are provided in Table 3.3 and can be used to design KBioscience Competitive Allele-Specific Polymerase chain reaction. (KASPar) based markers (Table 3.2).

Table 3.2. Nucleotide sequence flanking the SNPs associated with BLS resistance in the HWWAMP.

SNP Marker Name	Nucleotide Sequence
BS00084995_51	TGACAACAACTTTGGAGTTCAAGGCGTTAATAAGA GATACGACAACTGCA**[G/T]**CAAGGAAGGTTATAAC AATGCGTTTGGATTCGATAGCATAACTAAGACAT
Ku_c17846_363	CTGGTCACTTCTTAATAGCGTCACCCGGTTGTACCT GGCATTCTCCAATA**[C/T]**TGATGAGGGAACCTTTTC ATGTCTGTGTGTGTTACGTGCTCTAAGCACCC
IWA7541	TCAAATTCATCACTGGCTAAAGCCATGCTAATTGC TTCATCCATTGAGCTTTTGTAATCATCGGTTAGCAC ACTGTCAAGTCCCAGATCATTTGTTTCAG**[A/G]**ATT AAGATGATTATCACCTACAGATTGCAGAATAGAAA GCACATCCTCCAAATAGAAAGTTTTGACCGGATAG GTATGGCCTGGGACTCGAATAACCGGG
IAAV1943	ATTGACAGAGAGGAAGCGAAGCTAAAGAATGGTG GAGCGTGAGACAGATGATGGTGATGAACAATTGAT TAGACCACTGCAAATTTACCCATTCTTATTA**[C/T]**T TATAAGGGAGTCTACATCCTGAAATTATGGTAGGA CAGAGCTCATCTAGGTATTT
tplb0032m13_1358	CTGCTTATGAACCCTCTTAACATCCCCAGCTCCGGG CGCCATTTCTACCT**[C/T]**GCCGTTGACCGCCTCCAG TTCAAGATGAGGACACTACTGGAGCTCCTAGG

3.5. Comparative analysis of the QTL regions in rice and wheat

BLS is a very threatening disease in rice and several QTLs for bacterial leaf streak resistant have been reported (Tang et al. 2000). We performed a comparative syntenic analysis between the candidates QTL regions identified in

our study by extracting the sequence of the three candidate regions from the IWGSC wheat genome assembly v1.0 (Mayer et al. 2014). The candidate regions were repeat masked with RepeatMasker (http://www.repeatmasker.org/) using mipsREdat 9.3p Poaceae TEs repeat database (Nussbaumer et al. 2013). The low copy region was used to identify coding sequences in the candidate region by a blastn against wheat CDS databases. To identify a syntenic relationship of these wheat QTLs in the rice genome the wheat CDS sequences from the candidate regions were compared against the rice CDS sequences (Matsumoto et al. 2005) over layered on chromosomes. Of the five QTLs identified in our study QTLs *Q.bls.sdsu.1AL, Q.bls.sdsu.3AL* and *Q.bls.sdsu.4AL* (Figure 3.6) were likely syntenic to known BLS resistance QTLs *qBlsr5b, qBlsr1 and qBlsr3d* on chromosomes 5R, 1R and 3R in rice respectively (He et al. 2012; Tang et al. 2000).

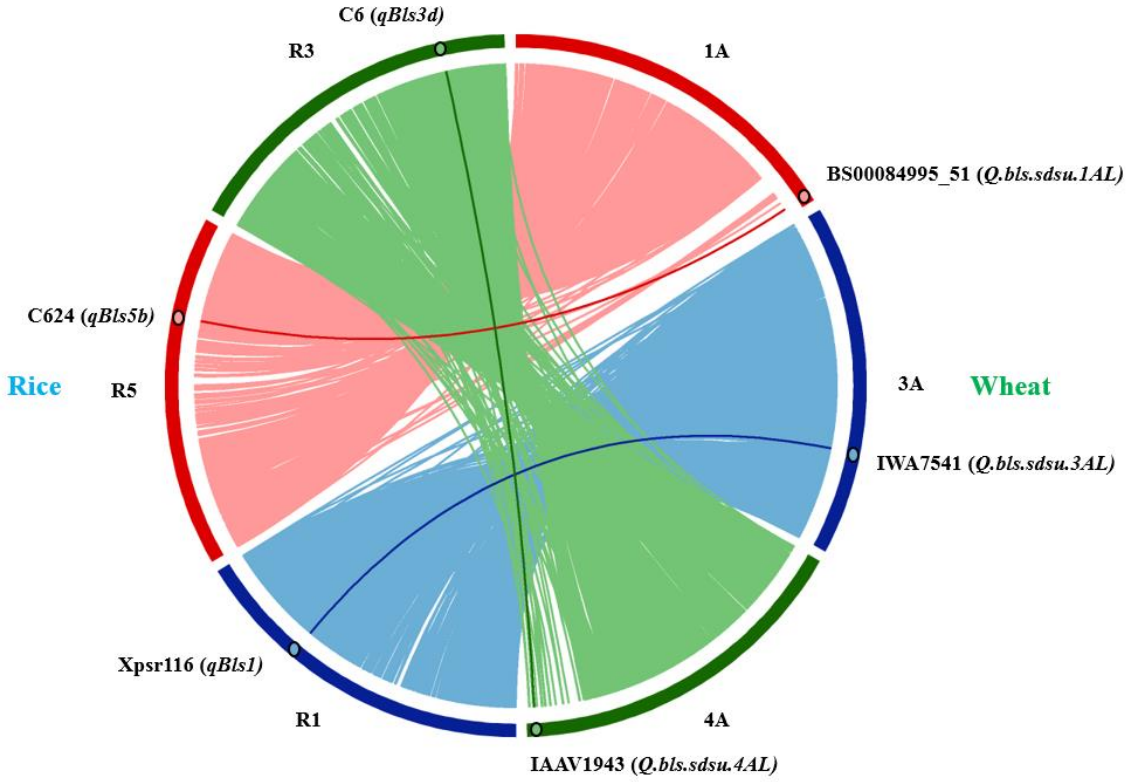

Figure 3.6. Synteny analysis of wheat chromosomes 3A, 4A, and 6A and corresponding rice chromosomes R1, R2 and R3 along with the location of their respective BLS resistance QTLs.

4. Discussion

BLS impacts many crop species including wheat, Triticale, rice and Brassica (Wechter et al. 2014; Jayawardana et al. 2016; Tang et al. 2000; Alizadeh et al. 1995). Development of disease resistant cultivars seems to be most effective management strategy in the absence of effective chemical control (Milus and Mirlohi 1995). However, the progress in the development of BLS resistant cultivars is hampered by complex inheritance of BLS resistance, poor understanding of the resistance mechanism and unavailability of good molecular markers. Some major genes and

several QTLs for BLS resistance have been reported in rice (Tang et al. 2000; Chen et al. 2006; He et al. 2012). Recently, Wen et al. (2017) reported a major QTL on 5R of Triticale. In wheat in one of the earlier studies five genes (Bls1, Bls2, Bls3, Bls4, and Bls5) for BLS resistance were suggested by Duveiller et al. (1993). Disease evaluations have shown that resistance to BLS in wheat is partial (Milus and Mirlohi 1995; Tillman et al. 1996; Kandel et al. 2012; Adhikari et al. 2012a). Most of these studies were performed in spring wheat with no extensive characterization of BLS resistance has been reported in winter wheat. To our knowledge, this is the first study that characterizes QTLs for BLS resistance in winter wheat. In the present study, we evaluated HWWAMP of 300 winter wheat accessions for their response to BLS in field and greenhouse conditions. The frequency distributions observed for both greenhouse and field BLS severity showed a normal distribution though it was slightly skewed towards susceptibility as expected because most of the germplasm in wheat is known to be moderately to highly susceptible to BLS, further suggesting that BLS is a quantitatively inherited trait in wheat (Duveiller et al. 1993; Tillman and Harrison 1996; Adhikari et al. 2012) similar to rice (Poulin et al. 2014; Tran et al. 2015; Tang et al. 2000; Li et al. 2008).

The mean disease incidence in the greenhouse was slightly lower than the field which could be attributed to the differences in the infiltration and inoculation methods or environmental conditions but there was a significantly positive correlation (0.62) observed between disease scores from these two locations. We identified only eleven accessions (3.6%) showing resistance to BLS (Table 3.3). Most of the resistant lines to BLS have the pedigree or a selection involving the genotype ‘Scout’. The

cultivar Scout released in 1963 by the University of Nebraska at Lincoln and later a selection was made for earliness and better quality. This selection was released as Scout 66 (1987). The cultivar Scout 66 is one of the 11 resistant genotypes of the HWWAMP. This indicates that Scout may be a major source of BLS resistance in the evaluated wheat genotypes. Scout also is reported to be resistant to Hessian fly, tan spot, and soil-borne diseases. It is moderately resistant to leaf rust and susceptible to stem rust and stripe rust (2016 UNL Fall Seed guide). A large number of genotypes showed a moderately susceptible reaction (152) to susceptible reaction (113) demonstrating the importance of resistant germplasm. Majority of these genotypes show on their pedigree cv. '2180'. The cultivar 2180 (PI 532912) was released by Kansas State University in 1989 and itself shows a moderately susceptible reaction to BLS in our study. It is possible that 2180 could be a source of susceptible to bacterial leaf streak in our panel.

Table 3.3. Details of the 11 identified BLS resistant lines from the HWWAMP.

Genotype	Pedigree	PI
EAGLE (KS: 1970)	Selection from Scout	Cltr15068
GOODSTREAK (NE: 2002)	Len//Butte/ND526/6/Agent/3/ ND441//Waldron/Bluebird/4/Butte/5/ Len/7/KS88H164 /8/NE89646	PI632434
LARNED (KS: 1976)	Ottawa /5* Scout	CItr17650
NE04490	NE95589/3/Abilene/Norkan//Rawhide/4/ Abilene/ Arapahoe	-
OK05723W	SWM866442/Betty	-
OK1068112	Farmec/Jagalene	-
ROBIDOUX (NE: 2010)	Odesskaya P/Cody// Pavon 76/3* Scout 66/3/ Wahoo	PI659690
SCOUT66 (NE: 1976)	Nebred//Hope/Turkey/3/ Cheyenne/Ponca	CI13996
TX07A001420	U1254-1-5-2-1/ TX81V6582// Desconocido	-
VISTA (NE: 1992)	Warrior//Atlas66/Comanche/3/Comanche/Ottawa/5/Ponca/2* Cheyenne/3/Illinois No. 1//2* Chinese Spring /T. timopheevii/4/ Cheyenne/Tenmarq// Mediterranean/Hope/3/ Sando60/6/Centurk/ Brule	PI562653
WENDY (SD: 2004)	Gent/Siouxland// Abilene	PI638521

Our GWAS for BLS resistance showed significant SNPs associated with five QTLs conferring resistance to BLS in winter wheat. The five QTLs on chromosome 1AL (*Q.bls.sdsu.1AL*), IBS (*Q.bls.sdsu.1BS*), 3AL (*Q.bls.sdsu.3AL*), 4AL (*Q.bls.sdsu.4AL*) and 7AS (*Q.bls.sdsu.7AS*) explaining 8.3%, 8.5%, 7.9%, 8.8% and 9.3% of the variation respectively which accounts for 42.3% of the total variation. Two QTLs *Q.bls.sdsu.1AL* and *Q.bls.sdsu.4AL* identified in our study are in a similar region to the QTLs reported by Adhikari et al. 2012b in GWAS analysis of BLS in spring wheat, *w*hereas another QTL *Q.bls.sdsu.1BS* in a similar region as reported by Kandel et al. (2015). The additional genomic regions conferring resistance to BLS on chromosomes 4B and 6B (Adhikari et al. 2012b) and chromosomes 2A and 6B (Kandel et al. 2015) were not significant in our HWWAMP. Three QTLs (*Q.bls.sdsu.1AL, Q.bls.sdsu.1BS, Q.bls.sdsu.3AL*) identified in the present study were located in regions similar to reported earlier but with high markers coverage, we identified of high-quality SNPs associated with these QTLs. Further, we identified two QTLs *Q.bls.sdsu.3AL,* and *Q.bls.sdsu.7AS* in novel regions not been reported in previous studies. Three *Q.bls.sdsu.1AL, Q.bls.sdsu.3AL* and *Q.bls.sdsu.4AL* of the five QTLs were mapped in syntenic regions when compared to rice, therefore, comparative genomic approached could help in fine mapping of these QTLs and understand the mechanism of BLS resistance in cereals. The high-quality SNPs identified in our study (Table 3.1 and 3.2) could be used to develop KASPar based markers for marker-assisted selection for BLS resistance.

5. Conclusions

In conclusion, our works not only identified sources of resistance to BLS in hard winter wheat germplasm but also characterizes genomic regions associated with resistance to Bacterial Leaf Streak. Further, the SNP markers associated with these genomic regions would be useful in marker-assisted selection in developing BLS resistant wheat varieties. The information from this work will help enhance our understanding of the molecular basis of BLS resistance in wheat.

6. Acknowledgments

The authors would like to thank the South Dakota Agriculture Experimental Station (Brookings, SD, USA) for providing the resources to conduct the field experiments. This project was partially funded by the USDA hatch project SD00H538-15 and South Dakota Wheat Commission 3X7261.

7. Literature cited

Adhikari TB, Gurung S, Hansen JM, Bonman JM (2012a) Pathogenic and Genetic Diversity of Xanthomonas translucens pv. undulosa in North Dakota. Phytopathology 102 (4):390-402. doi:10.1094/phyto-07-11-0201

Adhikari TB, Gurung S, Hansen JM, Jackson EW, Bonman JM (2012b) Association Mapping of Quantitative Trait Loci in Spring Wheat Landraces Conferring Resistance to Bacterial Leaf Streak and Spot Blotch. Plant Genome 5 (1):1-16. doi:10.3835/plantgenome2011.12.0032

Alizadeh A, Barrault G, Sarrafi A, Rahimian H, Albertini L (1995) IDENTIFICATION OF BACTERIAL LEAF STREAK OF CEREALS BY THEIR PHENOTYPIC CHARACTERISTICS AND HOST-RANGE IN IRAN. European Journal of Plant Pathology 101 (3):225-229. doi:10.1007/bf01874778

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215 (3):403-410. doi:10.1016/S0022-2836(05)80360-2

Arif MAR, Neumann K, Nagel M, Kobiljski B, Lohwasser U, Borner A (2012) An association mapping analysis of dormancy and pre-harvest sprouting in wheat. Euphytica 188 (3):409-417. doi:10.1007/s10681-012-0705-1

Bariana HS, Hayden MJ, Ahmed NU, Bell JA, Sharp PJ, McIntosh RA (2001) Mapping of durable adult plant and seedling resistances to stripe rust and stem rust diseases in wheat. Australian Journal of Agricultural Research 52 (11-12):1247-1255. doi:Doi 10.1071/Ar01040

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

Bordes J, Ravel C, Le Gouis J, Lapierre A, Charmet G, Balfourier F (2011) Use of a global wheat core collection for association analysis of flour and dough quality traits. Journal of Cereal Science 54 (1):137-147. doi:10.1016/j.jcs.2011.03.004

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

Brbaklic L, Trkulja D, Kondic-Spika A, Treskic S, Kobiljski B (2013) Detection of QTLs for Important Agronomical Traits in Hexaploid Wheat Using Association Analysis. Czech Journal of Genetics and Plant Breeding 49 (1):1-8

Dababat AA, Ferney GBH, Erginbas-Orakci G, Dreisigackee S, Imree M, Toktay H, Elekciogle HI, Mekete T, Nicol JM, Ansari O, Ogbonnaya F (2016) Association analysis of resistance to cereal cyst nematodes (Heterodera avenae) and root lesion nematodes (Pratylenchus neglectus and P-thornei) in CIMMYT advanced spring wheat lines for semi-arid conditions. Breeding Sci 66 (5):692-702. doi:10.1270/jsbbs.15158

Duveiller E, Vanginkel M, Thijssen M (1993) Genetic-Analysis of Resistance to Bacterial Leaf Streak Caused by Xanthomonas-Campestris Pv Undulosa in Bread Wheat. Euphytica 66 (1-2):35-43. doi:Doi 10.1007/Bf00023506

ElAttari H, Sarrafi A, Garrigues S, DechampGuillaume S, Barrault G (1996) Diallel analysis of partial resistance to an Iranian strain of bacterial leaf streak (Xanthomonas campestris pv cerealis) in wheat. Plant Pathology 45 (6):1134-1138. doi:10.1046/j.1365-3059.1996.d01-197.x

Faris JD, Anderson JA, Francl LJ, Jordahl JG (1996) Chromosomal location of a gene conditioning insensitivity in wheat to a necrosis-inducing culture filtrate from Pyrenophora tritici-repentis. Phytopathology 86 (5):459-463. doi:DOI 10.1094/Phyto-86-459

Guttieri MJ, Baenziger PS, Frels K, Carver B, Arnall B, Waters BM (2015) Variation for Grain Mineral Concentration in a Diversity Panel of Current and Historical Great Plains Hard Winter Wheat Germplasm. Crop Science 55 (3):1035-1052. doi:10.2135/cropsci2014.07.0506

Guttieri MJ, Frels K, Regassa T, Waters BM, Baenziger S (2017) Variation for nitrogen use efficiency traits in current and historical great plains hard winter wheat. Euphytica 213 (4). doi:10.1007/s10681-017-1869-5

He WA, Huang DH, Li RB, Qiu YF, Song JD, Yang HN, Zheng JX, Huang YY, Li XQ, Liu C, Zhang YX, Ma ZF, Yang Y (2012) Identification of a Resistance Gene bls1 to Bacterial Leaf Streak in Wild Rice Oryza rufipogon Griff. Journal of Integrative Agriculture 11 (6):962-969

Hirschhorn JN, Daly MJ (2005) Genome-wide association studies for common diseases and complex traits. Nat Rev Genet 6 (2):95-108. doi:10.1038/nrg1521

Jayawardana M, Li X, Fiedler J, Liu Z (2016) Genetic mapping of a gene conditioning resistance to bacterial leaf streak in triticale. Phytopathology 106 (12):189-189

Kandel YR, Glover KD, Osborne LE, Gonzalez-Hernandez JL (2015) Mapping quantitative resistance loci for bacterial leaf streak disease in hard red spring wheat using an identity by descent mapping approach. Euphytica 201 (1):53-65. doi:10.1007/s10681-014-1174-5

Kandel YR, Glover KD, Tande CA, Osborne LE (2012) Evaluation of Spring Wheat Germplasm for Resistance to Bacterial Leaf Streak Caused by Xanthomonas campestris pv. translucens. Plant Disease 96 (12):1743-1748. doi:10.1094/Pdis-03-12-0303-Re

Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA (2009) Circos: An information aesthetic for comparative genomics. Genome Research 19 (9):1639-1645. doi:10.1101/gr.092759.109

Letta T, Olivera P, Maccaferri M, Jin Y, Ammar K, Badebo A, Salvi S, Noli E, Crossa J, Tuberosa R (2014) Association Mapping Reveals Novel Stem Rust Resistance Loci in Durum Wheat at the Seedling Stage. Plant Genome 7 (1). doi:10.3835/plantgenome2013.08.0026

Li DX, Mao Q, Guo XL, Chen L (2008) Proteomic analysis of differentially expressed proteins from rice leaves in response to bacterial leaf streak. J Biotechnol 136:S626-S626. doi:10.1016/j.jbiotec.2008.07.1451

Li M, Liu XL, Bradbury P, Yu JM, Zhang YM, Todhunter RJ, Buckler ES, Zhang ZW (2014) Enrichment of statistical power for genome-wide association studies. Bmc Biol 12. doi:ARTN 73

10.1186/s12915-014-0073-5

Matsumoto T, Wu JZ, Kanamori H, Katayose Y, Fujisawa M, Namiki N, Mizuno H, Yamamoto K, Antonio BA, Baba T, Sakata K, Nagamura Y, Aoki H, Arikawa K, Arita K, Bito T, Chiden Y, Fujitsuka N, Fukunaka R, Hamada M, Harada C,

Hayashi A, Hijishita S, Honda M, Hosokawa S, Ichikawa Y, Idonuma A, Iijima M, Ikeda M, Ikeno M, Ito K, Ito S, Ito T, Ito Y, Ito Y, Iwabuchi A, Kamiya K, Karasawa W, Kurita K, Katagiri S, Kikuta A, Kobayashi H, Kobayashi N, Machita K, Maehara T, Masukawa M, Mizubayashi T, Mukai Y, Nagasaki H, Nagata Y, Naito S, Nakashima M, Nakama Y, Nakamichi Y, Nakamura M, Meguro A, Negishi M, Ohta I, Ohta T, Okamoto M, Ono N, Saji S, Sakaguchi M, Sakai K, Shibata M, Shimokawa T, Song JY, Takazaki Y, Terasawa K, Tsugane M, Tsuji K, Ueda S, Waki K, Yamagata H, Yamamoto M, Yamamoto S, Yamane H, Yoshiki S, Yoshihara R, Yukawa K, Zhong HS, Yano M, Sasaki T, Yuan QP, Shu OT, Liu J, Jones KM, Gansberger K, Moffat K, Hill J, Bera J, Fadrosh D, Jin SH, Johri S, Kim M, Overton L, Reardon M, Tsitrin T, Vuong H, Weaver B, Ciecko A, Tallon L, Jackson J, Pai G, Van Aken S, Utterback T, Reidmuller S, Feldblyum T, Hsiao J, Zismann V, Iobst S, de Vazeille AR, Buell CR, Ying K, Li Y, Lu TT, Huang YC, Zhao Q, Feng Q, Zhang L, Zhu JJ, Weng QJ, Mu J, Lu YQ, Fan DL, Liu YL, Guan JP, Zhang YJ, Yu SL, Liu XH, Zhang Y, Hong GF, Han B, Choisne N, Demange N, Orjeda G, Samain S, Cattolico L, Pelletier E, Couloux A, Segurens B, Wincker P, D'Hont A, Scarpelli C, Weissenbach J, Salanoubat M, Quetier F, Yu Y, Kim HR, Rambo T, Currie J, Collura K, Luo MZ, Yang TJ, Ammiraju JSS, Engler F, Soderlund C, Wing RA, Palmer LE, de la Bastide M, Spiegel L, Nascimento L, Zutavern T, O'Shaughnessy A, Dike S, Dedhia N, Preston R, Balija V, McCombie WR, Chow TY, Chen HH, Chung MC, Chen CS, Shaw JF, Wu HP, Hsiao KJ, Chao YT, Chu MK, Cheng CH, Hour AL, Lee PF, Lin SJ, Lin YC, Liou JY, Liu SM, Hsing YI, Raghuvanshi S, Mohanty A, Bharti AK, Gaur A, Gupta V, Kumar D, Ravi V, Vij S, Kapur A, Khurana P, Khurana P, Khurana JP, Tyagi AK, Gaikwad K, Singh A, Dalal V, Srivastava S, Dixit A, Pal AK, Ghazi IA, Yadav M, Pandit A, Bhargava A, Sureshbabu K, Batra K, Sharma TR, Mohapatra T, Singh NK, Messing J, Nelson AB, Fuks G, Kavchok S, Keizer G, Llaca ELV, Song RT, Tanyolac B, Young S, Il KH, Hahn JH, Sangsakoo G, Vanavichit A, de Mattos LAT, Zimmer PD, Malone G, Dellagostin O, de Oliveira AC, Bevan M, Bancroft I, Minx P, Cordum H, Wilson R, Cheng ZK, Jin WW, Jiang JM, Leong SA, Iwama H, Gojobori T, Itoh T, Niimura Y, Fujii Y, Habara T, Sakai H, Sato Y, Wilson G, Kumar K, McCouch S, Juretic N, Hoen D, Wright S, Bruskiewich R, Bureau T, Miyao A, Hirochika H, Nishikawa T, Kadowaki K, Sugiura M, Project IRGS (2005) The map-based sequence of the rice genome. Nature 436 (7052):793-800. doi:10.1038/nature03895

Mayer KFX, Rogers J, Dolezel J, Pozniak C, Eversole K, Feuillet C, Gill B, Friebe B, Lukaszewski AJ, Sourdille P, Endo TR, Dolezel J, Kubalakova M, Cihalikova J, Dubska Z, Vrana J, Sperkova R, Simkova H, Rogers J, Febrer M, Clissold L, McLay K, Singh K, Chhuneja P, Singh NK, Khurana J, Akhunov E, Choulet F, Sourdille P, Feuillet C, Alberti A, Barbe V, Wincker P, Kanamori H, Kobayashi F, Itoh T, Matsumoto T, Sakai H, Tanaka T, Wu JZ, Ogihara Y, Handa H, Pozniak C, Maclachlan PR, Sharpe A, Klassen D, Edwards D, Batley J, Olsen OA, Sandve SR, Lien S, Steuernagel B, Wulff B, Caccamo M, Ayling S, Ramirez-Gonzalez RH, Clavijo BJ, Steuernagel B, Wright J, Pfeifer M, Spannagl M, Mayer KFX, Martis MM, Akhunov E, Choulet F, Mayer KFX,

Mascher M, Chapman J, Poland JA, Scholz U, Barry K, Waugh R, Rokhsar DS, Muehlbauer GJ, Stein N, Gundlach H, Zytnicki M, Jamilloux V, Quesneville H, Wicker T, Mayer KFX, Faccioli P, Colaiacovo M, Pfeifer M, Stanca AM, Budak H, Cattivelli L, Glover N, Martis MM, Choulet F, Feuillet C, Mayer KFX, Pfeifer M, Pingault L, Mayer KFX, Paux E, Spannagl M, Sharma S, Mayer KFX, Pozniak C, Appels R, Bellgard M, Chapman B, Pfeifer M, Pfeifer M, Sandve SR, Nussbaumer T, Bader KC, Choulet F, Feuillet C, Mayer KFX, Akhunov E, Paux E, Rimbert H, Wang SC, Poland JA, Knox R, Kilian A, Pozniak C, Alaux M, Alfama F, Couderc L, Jamilloux V, Guilhot N, Viseux C, Loaec M, Quesneville H, Rogers J, Dolezel J, Eversole K, Feuillet C, Keller B, Mayer KFX, Olsen OA, Praud S, Iwgsc (2014) A chromosome-based draft sequence of the hexaploid bread wheat (Triticum aestivum) genome. Science 345 (6194). doi:ARTN 1251788

10.1126/science.1251788

Milus EA, Mirlohi AF (1993) A Test Tube Assay for Estimating Populations of Xanthomonas-Campestris Pv Translucens on Individual Wheat Leaves. Phytopathology 83 (2):134-139. doi:DOI 10.1094/Phyto-83-134

Milus EA, Mirlohi AF (1995) Survival of Xanthomonas-Campestris Pv Translucens between Successive Wheat Crops in Arkansas. Plant Disease 79 (3):263-265

Muleta KT, Rouse MN, Rynearson S, Chen XM, Buta BG, Pumphrey MO (2017) Characterization of molecular diversity and genome-wide mapping of loci associated with resistance to stripe rust and stem rust in Ethiopian bread wheat accessions. Bmc Plant Biology 17. doi:10.1186/s12870-017-1082-7

Nussbaumer T, Martis MM, Roessner SK, Pfeifer M, Bader KC, Sharma S, Gundlach H, Spannagl M (2013) MIPS PlantsDB: a database framework for comparative plant genome research. Nucleic Acids Research 41 (D1):D1144-D1151. doi:10.1093/nar/gks1153

Poulin L, Raveloson H, Sester M, Raboin LM, Silue D, Koebnik R, Szurek B (2014) Confirmation of Bacterial Leaf Streak Caused by Xanthomonas oryzae pv. oryzicola on Rice in Madagascar. Plant Disease 98 (10):1423-1423. doi:10.1094/Pdis-02-14-0132-Pdn

Silva IT, Rodrigues FA, Oliveira JR, Pereira SC, Andrade CCL, Silveira PR, Conciecao MM (2010) Wheat Resistance to Bacterial Leaf Streak Mediated by Silicon. Journal of Phytopathology 158 (4):253-262. doi:10.1111/j.1439-0434.2009.01610.x

Singh B (2005) Applications of linear models in animal breeding - A review. Indian J Anim Sci 75 (8):999-1007

Singh RP, Nelson JC, Sorrells ME (2000) Mapping Yr28 and other genes for resistance to stripe rust in wheat. Crop Science 40 (4):1148-1155

Tang D, Wu W, Li W, Lu H, Worland AJ (2000) Mapping of QTLs conferring resistance to bacterial leaf streak in rice. Theoretical and Applied Genetics 101 (1-2):286-291. doi:10.1007/s001220051481

Tillman BL, Harrison SA, Clark CA, Milus EA, Russin JS (1996) Evaluation of bread wheat germplasm for resistance to bacterial streak. Crop Science 36 (4):1063-1068

Tillman BL, Kursell WS, Harrison SA, Russin JS (1999) Yield loss caused by bacterial streak in winter wheat. Plant Disease 83 (7):609-614. doi:Doi 10.1094/Pdis.1999.83.7.609

Tran TT, Nga NV, Ngan PT, Hong NT, Szurek B, Koebnik R, Ham LH, Cuong HV, Cunnac S (2015) Confirmation of Bacterial Leaf Streak of Rice Caused by Xanthomonas oryzae pv. oryzicola in Vietnam. Plant Disease 99 (12):1853-1853. doi:10.1094/Pdis-03-15-0289-Pdn

Turner MK, Kolmer JA, Pumphrey MO, Bulli P, Chao S, Anderson JA (2017) Association mapping of leaf rust resistance loci in a spring wheat core collection. Theoretical and Applied Genetics 130 (2):345-361. doi:10.1007/s00122-016-2815-y

Varella AC, Weaver DK, Sherman JD, Blake NK, Heo HY, Kalous JR, Chao S, Hofland ML, Martin JM, Kephart KD, Talbert LE (2015) Association Analysis of Stem Solidness and Wheat Stem Sawfly Resistance in a Panel of North American Spring Wheat Germplasm. Crop Science 55 (5):2046-2055. doi:10.2135/cropsci2014.12.0852

Vargas M, Combs E, Alvarado G, Atlin G, Mathews K, Crossa J (2013) META: A Suite of SAS Programs to Analyze Multienvironment Breeding Trials. Agron J 105 (1):11-19. doi:10.2134/agronj2012.0016

Wang SC, Wong DB, Forrest K, Allen A, Chao SM, Huang BE, Maccaferri M, Salvi S, Milner SG, Cattivelli L, Mastrangelo AM, Whan A, Stephen S, Barker G, Wieseke R, Plieske J, Lillemo M, Mather D, Appels R, Dolferus R, Brown-Guedira G, Korol A, Akhunova AR, Feuillet C, Salse J, Morgante M, Pozniak C, Luo MC, Dvorak J, Morell M, Dubcovsky J, Ganal M, Tuberosa R, Lawley C, Mikoulitch I, Cavanagh C, Edwards KJ, Hayden M, Akhunov E, Sequencing IWG (2014) Characterization of polyploid wheat genomic diversity using a high-density 90 000 single nucleotide polymorphism array. Plant Biotechnology Journal 12 (6):787-796. doi:10.1111/pbi.12183

Wechter WP, Keinath AP, Smith JP, Farnham MW, Bull CT, Schofield DA (2014) First Report of Bacterial Leaf Blight on Mustard Greens (Brassica juncea) Caused by Pseudomonas cannabina pv. alisalensis in Mississippi. Plant Disease 98 (8):1151-1151. doi:10.1094/Pdis-09-13-0966-Pdn

William HM, Singh RP, Huerta-Espino J, Palacios G, Suenaga K (2006) Characterization of genetic loci conferring adult plant resistance to leaf rust and stripe rust in spring wheat. Genome 49 (8):977-990. doi:10.1139/G06-052

Yu LX, Lorenz A, Rutkoski J, Singh RP, Bhavani S, Huerta-Espino J, Sorrells ME (2011) Association mapping and gene-gene interaction for stem rust resistance in CIMMYT spring wheat germplasm. Theoretical and Applied Genetics 123 (8):1257-1268. doi:10.1007/s00122-011-1664-y

Zhang DD, Bowden RL, Yu JM, Carver BF, Bai GH (2014) Association Analysis of Stem Rust Resistance in US Winter Wheat. PloS one 9 (7). doi:ARTN e103747

10.1371/journal.pone.0103747

Zhang ZW, Ersoz E, Lai CQ, Todhunter RJ, Tiwari HK, Gore MA, Bradbury PJ, Yu JM, Arnett DK, Ordovas JM, Buckler ES (2010) Mixed linear model approach adapted for genome-wide association studies. Nat Genet 42 (4):355-U118. doi:10.1038/ng.546

CHAPTER – 4

GENOMIC SELECTION FOR GRAIN YIELD IMPROVEMENT IN THE SOUTH DAKOTA WINTER WHEAT BREEDING PROGRAM

Abstract

Recent studies suggest that genomic selection (GS) holds the potential to increase genetic gain for quantitative trait breeding in crop species. This technique uses all available genome-wide markers as predictors in a statistical model intended to predict the breeding value of complex traits such as yield in wheat without referring the underlying QTLs. The objective of our study was to evaluate the relative efficiency of genomic selection versus phenotypic selection for grain yield in the South Dakota winter wheat breeding program. A total of 434 unique advanced breeding lines or cultivars over the span of 4 years (2014 – 2017) were selected and genotyped – by – sequencing (GBS). The lines were grown under the Preliminary Yield Trial (PYT) and Advanced Yield Trial (AYT) nurseries and their grain yield from a 34 – year × locations combination were used in developing the genomic selection models. We developed several training and the validation sets to test the efficiency of the GS models. The single and multi-year analysis were done using several GS models (rrBLUP, PLSR, ELNET and Random Forest). The average predictions accuracies within a single year across locations were 0.62. However, with the multi-year-location analysis, the average genomic prediction accuracies were 0.26 for two-year combination, 0.32 for three-year combination and 0.36 for the four-year

combination. The rrBLUP algorithm demonstrated to show overall a better average prediction accuracy of 0.39 when compared to other models. Our results suggested several years of data is required to develop better genome-wide selection models.

1. Introduction

Wheat is the most widely planted crop in the world thus making it one the most important food source for the majority of the world population (Briggle LW and Curtis BC 1987). It is predicted that with the current world population would reach 9 billion by 2050 (Gerland et al. 2014). With the current rate of wheat growth across the globe it would be difficult to tackle this increase in demand for food supply (Gilbert and Morgan 2010). Further, challenges (abiotic and biotic stresses) faced by wheat makes it essential to develop wheat genotypes that can adapt to the frequently changing environment (Pandey et al. 2017).

The main objectives of breeding programs across the world are to develop improved high yielding varieties, with good physical characteristics of grains, and resistance to biotic and abiotic stresses (Arruda et al. 2016; Moose and Mumm 2008; Gupta et al. 2010; Rajaram 2001). Wheat breeding strategies generally aim to release pure line cultivars and it normally requires 10 – 12 years by a traditional winter wheat breeding program to release a variety for commercial purpose (Kuchel et al. 2005; Reynolds et al. 2011; Kirigwi et al. 2004; Wrigley 1994). Thus it is essential to incorporate recent state of the art computational techniques with the standard breeding protocols to shorten this lengthy breeding cycle and hence lead to enhanced genetic gain which would help meet the increased demand of food (Edgerton 2009).

With the sudden availability of a plethora of genetic and genomic data, several bioinformatics approaches have shed light to tackle this situation. Amongst these approaches is Genomic Selection (Poland and Rife 2012; Xu and Crouch 2008). Genomic selection was first suggested by Meuwissen et al. (2001) for animal breeding. While the traditional phenotypic selection is considered cumbersome and time-consuming, the GS uses genome-wide molecular markers to predicting phenotypes (GEBVs) to substitute the phenotype-dependent field evaluation. GS would substantially reduce the effort and investment on field assessment in quantitative traits and could help enhance the scale of breeding (). Further, breeding programs test genotypes in several environments to select the best line based on a particular for specific environments or best lines overall environments (Michel et al. 2017). Genomic Selection takes into account the genotype into environment interaction thus helping in selecting lines across various locations (Jarquin et al. 2017).

Genomic Selection has become a common practice in animal breeding (VanRaden et al. 2009; Pryce and Daetwyler 2012; Villumsen et al. 2009). In last few years GS has been also studied in many crops like Soybeans (Shu et al. 2013; Jarquin et al. 2014), wheat (Heffner et al. 2011a; Storlie and Charmet 2013; Huang et al. 2016), rice (Spindel et al. 2015), corn (Technow et al. 2013), sugar beet (Wurschum et al. 2013) for different traits including quality, plant architecture, disease resistance and grain yield etc. (Jannink et al. 2010; Heffner et al. 2011b; Heffner et al. 2011a; Varshney et al. 2016). These studies have shown the power of genomic selection in various plant species. In wheat de los Campos et al. (2013) first suggested the

inclusion of SNP marker into a genomic selection model increases the accuracy of prediction. Ever since then genomic selection has been used to predict several characteristics of wheat such as grain yield (Poland et al. 2012a), plant height, heading date, lodging, pre-harvest sprouting (PHS), flour yield and flour protein (Heffner et al. 2011b). Genomic selection has also been implemented in disease resistance-related studies like stem rust (Rutkoski et al. 2011), Fusarium head blight (FHB) (Rutkoski et al. 2012) and stripe rust (Juliana et al. 2017). Several models have been developed and tested for GS pipelines. Most commonly used models are ridge regression best linear unbiased predictions (Meuwissen et al. 2001), random forest (Breiman 2001) and Bayesian statistics (Dekkers et al. 2009).

The major focus of these studies was to increase the accuracy of the prediction models. Very few studies have focused on the impact genomic selection would have on a traditional breeding program. To bridge this gap, our study focuses on the evaluation of various genomic prediction models, the size of training and validation sets and establish a genomic selection pipeline to predict yield across the environment and over years in the South Dakota Winter Wheat Breeding program.

2. Material and Methods

2.1. South Dakota Winter Wheat breeding program

Winter wheat breeding at South Dakota State University (SDSU) focuses on breeding red and white winter wheat for the state of South Dakota. Majorly a modified bulk selection method is followed in developing new improved winter wheat cultivars. Each year 500 to 800 new cross combinations are developed

from which the F_1 progenies are obtained and multiplied in the Arizona winter nursery. In the F_2 generation, 600 populations are advanced with selections at one location. Selected 400 to 500 F_3 populations are evaluated at three locations and single head selections are performed. Each head is planted as $F_{3:4}$ rows and the selected lines are then advanced to the $F_{3:5}$ Early Observation Trial (EOT). In EOT up to 2,000 lines are evaluated in a single location (planted in 4 short row plots). The selected entries (500 to 800 lines) are then advanced to an unreplicated Early Yield Trials nursery (EYT) at two locations. The best performing 120 lines are selected from the EYT which are taken forth to the Preliminary Yield Trial nursery (PYT). This PYT trial is conducted at seven locations with two replications across South Dakota. From here a total of 35 lines are promoted to the Advanced Yield Trials nursery (AYT) which are tested at seven locations in South Dakota in three replications. Finally, from these 35 advanced lines, 10 SDSU elite experimental lines are advanced to the Crop Performance Trial nursery (CPT) where the elite lines are evaluated at 14 locations for three years before a variety is released.

2.2. Phenotypic data analysis

Implementation of multiple environments is a key factor in a genomic selection pipeline. In the South Dakota winter wheat breeding program the PYT and AYT nurseries are the only two nurseries that are planted in multiple locations including multiple replications in each environment. Thus this makes these two nurseries good candidates for the genomic selection pipeline. Grain yield data from four to seven locations across four years (2014 – 2017) from the

AYTs and PYT nurseries were obtained. The number of entries and locations in each nursery varied depending on the year. Not all the locations are present through the four years. Amongst all the seven locations Hayes is present in one year, while the other six locations are present in a minimum of 3 years. On an average AYTs and PYTs contained between 30 – 36 and 90 – 120 lines respectively. Each trial is planted in a randomized complete block design with 2 – 3 replications of PYT and AYT respectively. The grain yield data were analyzed separately using the statistical tool PROC GLM (SAS/STAT(R) 9.2) and the BLUPs, BLUEs were calculated using META – R (Vargas et al. 2013). BLUP value for the grain yield for each of the 434 genotypes was calculated to account for variable genotypes grown per year and the location. A linear mixed model was used in META – R to calculate the BLUPs with genotypes as random effects and locations as fixed effects. These BLUPs were used as the phenotype values for each line in the genomic selection pipeline.

2.3. Genotypic data analysis

A total of 434 lines from AYT and PYT nurseries were genotyping – by – sequencing. The DNA extraction was performed on bulked leaf tissue from each line using the BioSprint 96 DNA Plant Kit (Qiagen. Hilden, Germany) with the BioSprint 96 Workstation (Qiagen. Hilden, Germany). Genotyping – by – sequencing (GBS) performed by preparing GBS libraries form individually digested DNA of each genotype (*Pst*-I and *Msp*-I) followed by adapter ligation, and amplification. The GBS library was sequenced on Ion Proton system and the single nucleotide polymorphisms (SNPs) was called using a TASSEL 5 reference

based pipeline (Poland et al. 2012a). The missing data were imputed with the beagle software (Browning and Browning 2007).

2.4. Models implemented to establish the genomic selection pipeline

The 'lme4' package in R was used to fit these mixed linear models. The R package 'GSwGBS' was used to implement the GS pipeline (https://github.com/gaynorr/GSwGBS). The ‘GS.model’ function of this package was mainly used which is a wrapper for obtaining GS predictions using statistical models implemented in other R packages (https://github.com/gaynorr/GSwGBS). The function was designed to minimize the amount of user-generated coded needed to run these models, create a consistent method for calling each model, and to allow for fast computation. Genetic data is intended to come from numerically coded markers produced by hap2marker, but any markers similarly coded can be used (Poland et al. 2012b).

Up to five different GS models are used for predictions. Predictions from all chosen methods are returned in a data frame with the average of all selected methods if more than one method was chosen. Where possible, the 'foreach' package is used for parallel computing to reduce runtime (Revolution Analytics and Watson 2014). The 'rrBLUP' package (Endelman 2011) is used to generate predictions based on estimated marker effects using a ridge regression best linear unbiased prediction approach (Equation. 1) or estimated line effects using a Gaussian kernel (Endelman 2011). Random forest regression is implemented using the 'randomForest' package (Equation 2) (Liaw and Wiener, 2002). A partial least squares regression model is implemented using the 'pls' package

(Mevik and Wehrens 2007) (Equation 3). The number of components retained in this model is determined using 10-fold cross-validation (CV) on the training population to minimize the bias-corrected CV estimate. The fifth model is an elastic net model produced using the 'glmnet' package (Friedman et al. 2010). The elastic net mixing parameter and lambda are both set using a grid selection, 10-fold CV approach. A sequence of mixing parameters ranging from 0 (equivalent to the ridge regression penalty) to 1 (lasso penalty) is examined with a sequence of lambdas generated by the glmnet function to identify a pair of values which produce the lowest CV error (Jiang and Wang 2017). The mathematical representation of the models used are as follows:

rrBLUP: $Y = \mu + Xg + e$……(Equation.1)

where Y is a N × 1 vector of phenotypic means, μ is the overall mean of the training set, X N × Nm marker matrix, g is the Nm × 1 marker effects and e is the N × 1 vector of residual effects.

Random Forest: $\hat{y} = \frac{1}{m}\sum_{i=1}^{n} W_j(x_i, \acute{x})y_i$……..(Equation. 2)

where $W(x_i, \acute{x})$ is the non – negative weight of the i^{th} training point relative to the new point $x^{'}$ in the same tree and $\hat{y}$ is the predictor.

PLSR: $y_i = \sum_{k=1}^{p} x_{ik}\beta_k + e_i$……..(Equation.3)

where y_i is the phenotype of individual i, x_i is the 1 × p vector of SNP genotypes of individual i at locus k and p loci, β_k is the effect of SNP k and e_i is the residual term.

3. Results

3.1. Phenotyping

The overall average grain yield for all lines across all locations and year is 51.6 bushels/acre. The mean grain yield in both AYT and PYT nurseries were comparable within each year for all four years of trials (Figure 4.1). The average performance was higher in 2016 trials as compared to all other years. Based on the number of shared lines across year we estimated the correlations among the shared lines. The correlations were low ranging from 0.18 to 0.3 (Table 4.1). The combined heritability, grand mean, LSD and CV was calculated for all locations across all years (Table 4.2). The genotypes were significantly different as expected (Table 4.2). When comparing across locations Selby, SD showed an overall highest grain yield of 72.71 overall 34 years – locations whereas Hayes, SD had the lowest grain yield 21.90. However, the broad sense heritability (H^2 = 0.79) was highest at Hayes followed by Dakota Lakes (0.76) and was least in Winner, SD (0.22). The heritability of a location could have an impact on the genomic prediction accuracy estimated at this location. Stability analysis was performed to test the stability of all the genotypes across all replication, locations and years. We also tested the stability of all locations across four years. The genotypes were significantly different from each other with a p-value of $2.26e^{-16}$. The genotype by environment interaction was also captured and it was significant (p-value = $1.12e^{-16}$). The blocking factor applied in the linear mixed model (replication) was significant (p-value = $2.26e^{-16}$). Genotypes showed more stable

performance in Aurora and Wall in four years whereas Dakota lakes and Onida were highly variable locations.

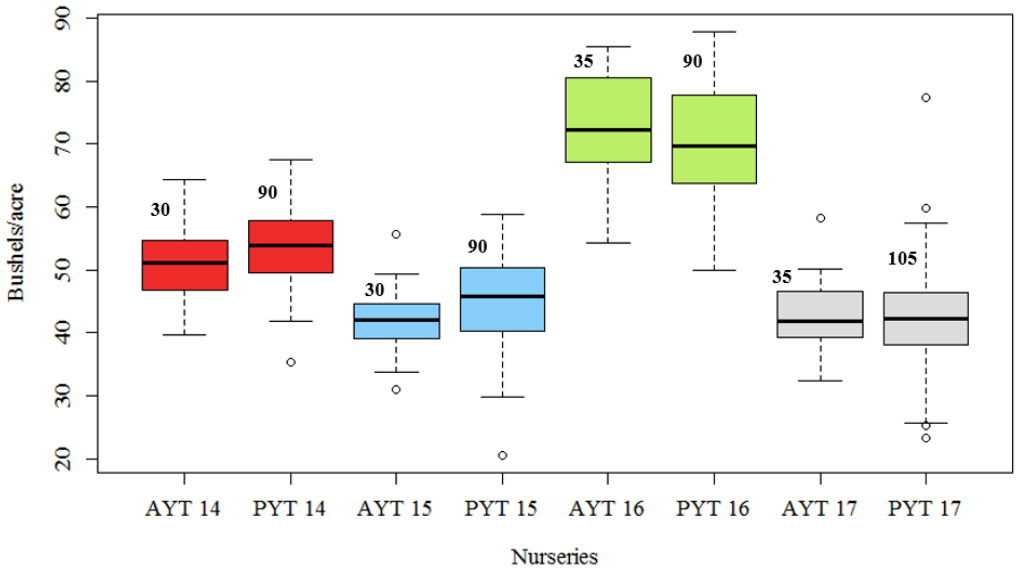

Figure 4.1. Overall grain yield in AYT and PYT nurseries in all seven locations and four years (2014-2017). The box plots show line means.

Table 4.1. Correlations among the shared lines of each succeeding year for grain yield.

Year	Shared lines between the years			Correlation
	AYT	PYT	Total	
2014 - 2015	9	27	36	0.21
2015 - 2016	17	38	55	0.30
2016 - 2017	12	20	32	0.18

Table 4.2. Heritability and other statistical data for grain yield from all locations (7) through all years (2014 – 2017).

Statistic	Aurora	D Lakes	Hayes	Onida	Selby	Wall	Winner
Heritability	0.45	0.76	0.79	0.75	0.72	0.49	0.22
Genotype Variance	43.26	174.7	15.55	160.36	248.23	43.43	17.91
Residual Variance	156.57	164.72	11.85	157.73	282.48	133.25	182.25
Grand Mean	55.3	57.71	21.90	53.82	72.71	59.18	54.20
LSD	13.49	18.96	5.54	18.33	24.31	13.51	10.41
CV	22.62	22.23	15.72	23.33	23.11	19.50	24.90
n Replicates	3	3	3	3	3	3	3
Genotype significance	$3.05e^{-5}$	0	0	0	0	$4.23e^{-8}$	$2.54e^{-3}$

3.2. Genotyping

A total of 348,682 SNPs were called via a reference based pipeline using IWGSC wheat genome assembly v1.0 (Mayer et al. 2014). The monoallelic SNPs were removed to obtain 224,169 SNPs. Further data filtering was done and SNPs with a MAF of <0.5% and a missing percentage of <90% were considered. A total of 8,164 high-quality SNPs were obtained which were then imputed using the beagle software (Browning and Browning 2007). The SNPs were then converted to numeric codes (i.e. 1 for lines homozygous for the most common allele, 0 for heterozygotes, and -1 for lines homozygous for the less frequent allele).

3.3. Training and validation set analysis

We selected subsets of entries from AYT and PYT nurseries to develop a training set (TS) and residual entries form the subset formed the validation set (VS). We evaluated several training sets of sizes ranging from 20% to 80% of the entries in that year and the analysis was performed for all four years (2014-2017). A total of 700 iterations were conducted and the optimized training set size yielding a high r^2 (predicted phenotypes and empirical phenotypes) were selected (Figure 4.2). A training set of 80% entries had the highest r^2 value of 0.78 and whereas the TS with 30% entries had the lowest of 0.44%. We observed no significant improvement in r^2 when the TS constituted 60% entries compared to TS with 80% entries. Therefore, we selected the final training set of 60% entries in further analysis. We also evaluated the prediction accuracy of the TS with 60% entries from single one year and multiyear combinations (Table 4.3).

Each prediction was performed for several iterations and we concluded that a TS of 60% of the entries is most suitable for GS in our breeding program.

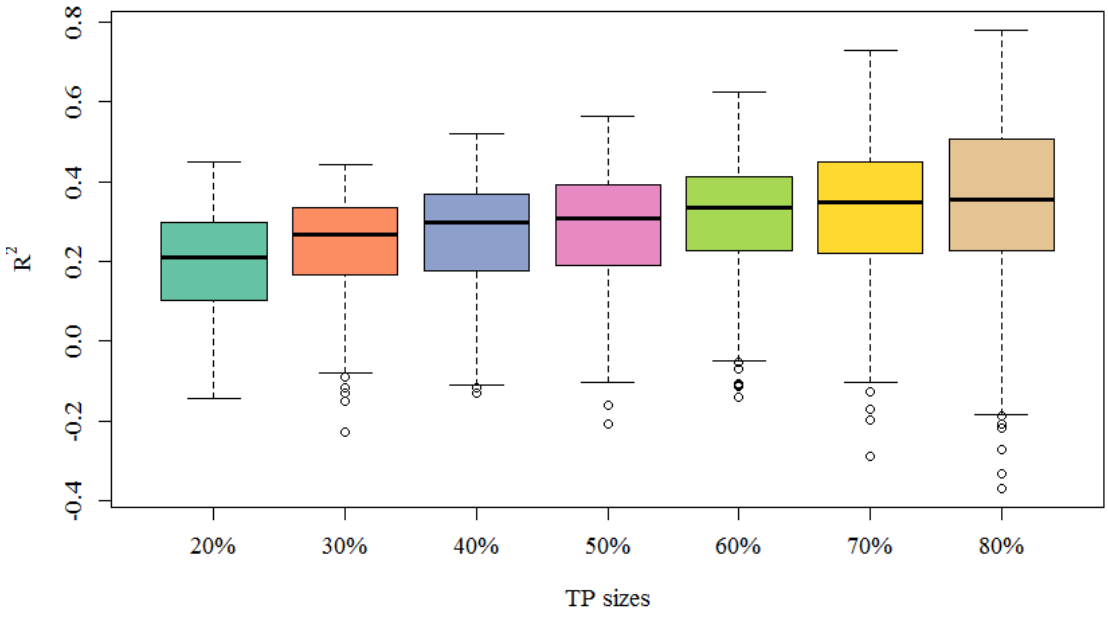

Figure 4.2. Boxplot showing the prediction accuracies (r^2) obtained for various training population sizes in the year 2014.

Table 4.3. Training set and validation set combinations used in the present study to estimate genomic prediction accuracies.

TS	Lines	VS	Lines	PCA Plot
2014 (AYT) [TS1]	30	2014 (PYT)	90	Fig. 4.3.A
2014 (AYT, PYT) + 2015 (AYT) [TS2]	114	2015 (PYT)	90	Fig. 4.3.B
2014 (AYT, PYT) + 2015 (AYT, PYT) + 2016 (AYT) [TS3]	175	2016 (PYT)	90	Fig. 4.3.C
2014 (AYT, PYT) + 2015 (AYT, PYT) + 2016 (AYT, PYT) + 2017 (AYT) [TS4]	228	2017 (PYT)	105	Fig. 4.3.D

3.4. Prediction accuracy within a single year

Four genomic selection algorithms were tested on a single year in which each location was used to predict all the other locations. Overall a good correlation was observed between the predicted and expected yield. An average prediction accuracy (r^2) of 0.62 was obtained (Table 4.4). Amongst the four tested algorithms (rrBLUP, PLSR, ELNET and Random Forest) on an average PLSR gives the highest correlation of 0.66 between all locations and ELNET was the poorest predictor with a value of 0.57 (Table 4.4). Similarly, prediction accuracies of location to location comparison (location combination) were analyzed and Dakota Lakes-Wall yielded the highest correlation of 0.71 and Winner-Aurora yielded the lowest of 0.46.

Table 4.4. Prediction accuracies obtained for 2014 PYT (VS) using the 2014 AYT as the TS. This analysis is done across all locations using rrBLUP, PLSR, ELNET and Random Forest prediction algorithms.

rrBLUP				
Location	**Aurora**	**D Lakes**	**Onida**	**Wall**
D Lakes	0.59			
Onida	0.65	0.63		
Wall	0.55	0.73	0.63	
Winner	0.46	0.8	0.45	0.72
PLSR				
Location	**Aurora**	**D Lakes**	**Onida**	**Wall**
D Lakes	0.8			
Onida	0.77	0.79		
Wall	0.74	0.82	0.5	
Winner	0.51	0.48	0.53	0.67
ELNET				
Location	**Aurora**	**D Lakes**	**Onida**	**Wall**
D Lakes	0.53			
Onida	0.58	0.66		
Wall	0.53	0.51	0.57	
Winner	0.56	0.48	0.61	0.74
Random Forest				
Location	**Aurora**	**D Lakes**	**Onida**	**Wall**
D Lakes	0.76			
Onida	0.53	0.74		
Wall	0.58	0.77	0.71	
Winner	0.46	0.6	0.55	0.55

3.5. Prediction accuracy across multiple years

Upon addition of years into the training and prediction set, the prediction accuracies were considerably lower as compared to a single year (Table 4.5). In the two – year analysis we used 2014 AYT and PYT and 2015AYT nurseries (TS2) to predict the 2015 PYT nursery (Table 4.5). The average prediction accuracy obtained was 0.26 with all four algorithms. On the contrary, to single year analysis where PSLR showed best performed (r^2=0.66), in the two-year analysis, PLSR had the lowest average correlation of 0.25 whereas rrBLUP had the highest with a correlation of 0.29. The correlation reduced drastically with across year predictions. Similarly, prediction accuracies of location to location comparison (location combination) were analyzed and Onida-Aurora yielded the highest correlation of 0.35 and Winner-Aurora yielded the lowest of 0.20.

Upon addition of another year into the GS model (three-year combination) the average prediction accuracy increased when compared to the two-year analysis. We used 2014, 2015 AYT and PYT nurseries combined with 2016 AYT nursery (TS3: three-year combination) to predict the 2016 PYT nursery (Table 4.6). The average prediction accuracy obtained was 0.32. The GS algorithm rrBLUP was consistently the best predictor and PLSR was the lowest performing predictor in this analysis as well. However, on the average prediction accuracy obtained with rrBLUP (0.32) and PLSR (0.31) are not much different. Similarly, prediction accuracies of location to location comparison (location combination) were analyzed and Dakota Lakes-Aurora yielded the highest correlation of 0.43 and Winner-Aurora yielded the lowest of 0.27.

We further analyzed a four-year combination GS model (TS4). Here we used 2014, 2015 and 2016s AYT and PYT nurseries along with the 2017 AYT nursery (TS4: three-year combination) to predict the 2017 PYT nursery. The average prediction accuracy obtained was 0.36. We didn't find a significant increase in the overall average prediction accuracy when we compared the results of the four-year combination analysis to the three-year combination analysis. However, rrBLUP still performed the best yielding an average prediction accuracy of 0.36. Similarly, prediction accuracies of location to location comparison (location combination) were analyzed and Winner-Aurora yielded the highest correlation of 0.45 and Winner-Wall yielded the lowest of 0.30.

Table 4.5. Prediction accuracies obtained for 2015 PYT nursery using the data from the 2014 (AYT, PYT) and 2015 (AYT) nurseries as the TS. This information contains all the locations.

rrBLUP				
Location	**Aurora**	**D Lakes**	**Onida**	**Wall**
D Lakes	0.32			
Onida	0.31	0.26		
Wall	0.24	0.32	0.34	
Winner	0.25	0.33	0.27	0.26
PLSR				
Location	**Aurora**	**D Lakes**	**Onida**	**Wall**
D Lakes	0.23			
Onida	0.22	0.21		
Wall	0.33	0.33	0.21	
Winner	0.26	0.2	0.27	0.33
ELNET				
Location	**Aurora**	**D Lakes**	**Onida**	**Wall**
D Lakes	0.3			
Onida	0.2	0.25		
Wall	0.28	0.2	0.31	
Winner	0.25	0.25	0.28	26
Random Forest				
Location	**Aurora**	**D Lakes**	**Onida**	**Wall**
D Lakes	0.23			
Onida	0.35	0.28		
Wall	0.23	0.28	0.21	
Winner	0.2	0.31	0.26	0.31

Table 4.6. Prediction accuracies obtained for 2016 PYT nursery using the data from the 2014 and 2015 (AYT, PYT) and 2016 (AYT) nurseries as the TS. This information contains all the locations.

rrBLUP				
Location	**Aurora**	**D Lakes**	**Onida**	**Wall**
D Lakes	0.3			
Onida	0.4	0.32		
Wall	0.4	0.27	0.28	
Winner	0.35	0.31	0.34	0.32
PLSR				
Location	**Aurora**	**D Lakes**	**Onida**	**Wall**
D Lakes	0.38			
Onida	0.39	0.26		
Wall	0.29	0.26	0.36	
Winner	0.27	0.26	0.37	0.31
ELNET				
Location	**Aurora**	**D Lakes**	**Onida**	**Wall**
D Lakes	0.43			
Onida	0.33	0.29		
Wall	0.25	0.4	0.25	
Winner	0.42	0.31	0.28	0.42
Random Forest				
Location	**Aurora**	**D Lakes**	**Onida**	**Wall**
D Lakes	0.39			
Onida	0.38	0.29		
Wall	0.34	0.36	0.26	
Winner	0.27	0.31	0.25	0.32

Table 4.7. Prediction accuracies obtained for 2017 PYT nursery using the data from 2014, 2015 and 2016 (AYT, PYT) and 2017 (AYT) nurseries as the TS. This information contains all the locations.

rrBLUP				
Location	**Aurora**	**D Lakes**	**Onida**	**Wall**
D Lakes	0.3			
Onida	0.36	0.3		
Wall	0.39	0.31	0.31	
Winner	0.45	0.4	0.44	0.36
PLSR				
Location	**Aurora**	**D Lakes**	**Onida**	**Wall**
D Lakes	0.45			
Onida	0.42	0.33		
Wall	0.34	0.33	0.33	
Winner	0.45	0.3	0.36	0.33
ELNET				
Location	**Aurora**	**D Lakes**	**Onida**	**Wall**
D Lakes	0.43			
Onida	0.44	0.44		
Wall	0.32	0.34	0.32	
Winner	0.41	0.35	0.37	0.3
Random Forest				
Location	**Aurora**	**D Lakes**	**Onida**	**Wall**
D Lakes	0.32			
Onida	0.31	0.34		
Wall	0.39	0.38	0.4	
Winner	0.35	0.38	0.42	0.38

4. Discussion

Genomic selection is a promising technique for improving qualitative traits in plant species (Heffner et al. 2009) and is better equipped when compared to traditional marker-assisted selection (Bernardo and Yu 2007).

The number of markers needed for a proper genomic selection pipeline is crucial and the number of markers depends solely on the linkage disequilibrium decay in the species and the germplasm or population under consideration (Zhong et al. 2009). Studies have shown that with a rapid LD decay up to 1 million markers are required for an effective genomic selection model (Van Inghelandt et al. 2011). On the contrary to this, studies have also shown that 100 – 7000 markers could be sufficient to achieve a good genomic prediction accuracy (Lorenzana and Bernardo 2009; Cavanagh et al. 2013). Using GBS markers in a GS study make the pipeline more efficient when compared to a study using DArT markers (Akbari et al. 2006; Poland et al. 2012b). In our study, we used a total of 8,164 high-quality GBS based SNP markers which are distributed evenly throughout the entire genome.

Training population and validation population size plays a critical role in affecting the genomic selection prediction accuracy. Generally, a TS with a large number of individuals, highly related to the validation population would give the most accurate prediction accuracy (Isidro et al. 2015; Heffner et al. 2011a; Bentley et al. 2014). Since one of the aims of genomic selection is reducing the need for phenotyping, we explored the effect of training set size to obtain useful prediction accuracies. Decreasing the training population size had a significant decrease in our prediction accuracy (Figure 4.3). We obtained a significantly high prediction

accuracy when we used a training population size of 60% entries. This was in consensus with the results obtained by a multifamily prediction for genomic selection in wheat (Isidro et al. 2015; Zhong et al. 2009; Bentley et al. 2014; Heffner et al. 2011a).

We performed a single year and multi-year analysis in our study (Table 4.3). In the single year analysis, we obtained a high average prediction accuracy of 0.62. Similar prediction accuracy in the range of 0.48 – 0.72 have been reported in elite European maize breeding population (Zhao et al. 2012). In the multi-year analysis, we performed three separate studies (two-year, three-year, and four-year combination analysis) by adding a year into the genomic selection model each time (Table 4.3). In the two-year analysis (TS2) we achieved an average prediction accuracy of 0.26. Upon addition of another year into the model (three-year combination: TS3) the average prediction accuracy significantly increased to 0.32. However, we found an only slight increase in the four-year combination (TS4) in which we obtained an average prediction accuracy of 0.36 when we compared the results of the three-year combination (TS3). These results obtained in our study is consistent to the results obtained in genomic selection studies performed to predict yield parameters in an biparental synthetic derived wheat lines (Dunckel et al. 2017), elite tropical rice breeding lines (Spindel et al. 2015) and in an F_5 derived soft winter wheat population (Heffner et al. 2011b). Genomic selection studies performed on a biparental wheat breeding population for quality traits also showed lower prediction accuracies for multi-year analysis when compared to single-year analysis (Heffner et al. 2011a). However, prediction accuracies in our study are lower when compared to genomic

selection studies for grain yield performed on elite hybrid rye populations (Wang et al. 2014). High throughput phenotyping based genomic selection studies have also been done to predict grain yield in inbred wheat lines (Haghighattalab et al. 2017; Sun et al. 2017; Rutkoski et al. 2016). The genomic prediction accuracy obtained in these studies ranged from 0.21 – 0.72 which is comparable to the results obtained by single year and multi-year analysis in our study.

Asoro et al. 2011 demonstrated that the levels of relatedness of individuals in a population can have a drastic impact on genomic selection models affecting the prediction accuracies to great extents (Asoro et al. 2011). To understand this with respect to our populations we analyzed the allelic diversity between all our training sets and validation sets (Figure 4.3). Four principal components were generated to study this. In the single year analysis, the TS and VS are scattered throughout the PCA plot and the TS set very well represented the VS thus sharing high allelic diversity (Figure 4.3.A). In addition to the effect of single year environment, the genetic relatedness and better representation of the VS may have resulted in a higher average prediction accuracy (0.62) for the single year analysis. We performed the same study on the multi-year analysis as well. When we added another year to the model (two-year combination) the VS clustered into a small confined region while the TS was scattered throughout the plot (Figure 4.3.B). The TS did not represent the VS to a great extent and hence we obtained a very low prediction accuracy (0.26). However, as we added more years into the model (three-year combination and four-year combination) the VS is more scattered through the plot and the relatedness between the TS and VS seems to be increased (Figure 4.3.C, Figure 4.3.D). Thus

better representation of training set very for the VS may have resulted in significant increase in the prediction accuracy for the three-year combination (0.32) (Figure 4.3.C) and the four-year combination (0.36) (Figure 4.3.D).

An optimal GS method should provide the highest prediction accuracy possible, solely basing itself on marker LD rather than on kinship (Habier et al. 2007). In our study, we tested four genomic selection models (rrBLUP, PLSR, ELNET and Random Forest). Among these four models rrBLUP consistently performed better resulting in an average prediction accuracy of 0.39. In our GS model, we obtain a better prediction accuracy for grain yield (0.36) over years when compared to Heffner et al (Heffner et al. 2011b) 0.22 or Gaynor (2015) 0.34 in a wheat breeding. Crossa et al. in 2010 performed a genomic selection study in 599 historical wheat lines and 284 maize inbred from the International Maize and Wheat Improvement Center (CIMMYT). Testing multiple GS models and environments, the rGS for wheat grain yield obtained in their study ranged from 0.36 – 0.61. The results obtained in our study were comparable to the results obtained in their study. Further, the genomic prediction accuracy estimated for grain yield is generally lower than for end-use quality traits like test weight, 1000-kernel weight, hardness, grain and flour protein, flour yield, sodium dodecyl sulfate sedimentation, Mixograph and Alveograph performance, and loaf volume (Sarah Battenfield et al 2016). The prediction accuracy for rrBLUP in their study ranged between 0.41 to 0.68 however they did not study the prediction accuracies for grain yield.

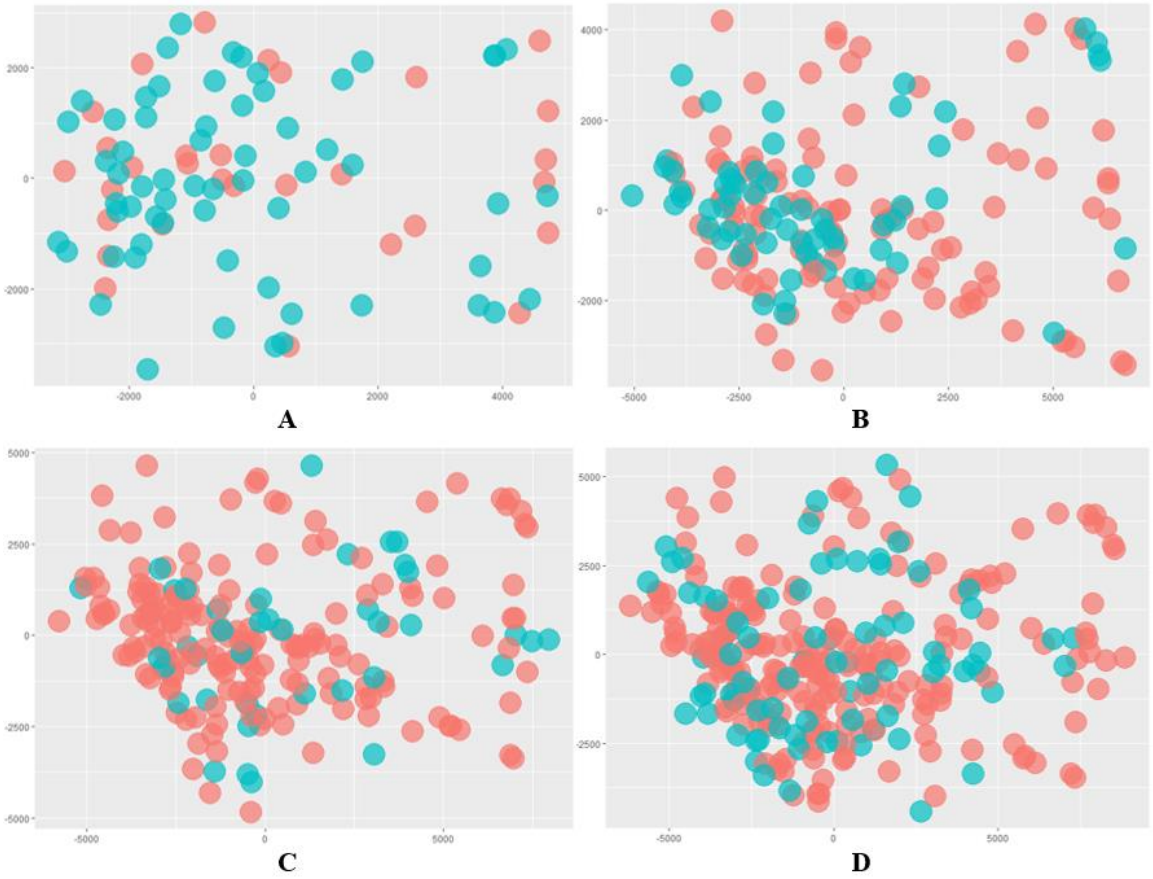

Figure 4.3. PCA plots showing the allelic diversity and genetic relatedness between the various Training set (TS) and validation set (VS) described in table 4.3. Red denoted the TS and blue denotes the VS.

5. Conclusion

In conclusion, we successfully established a genomic selection pipeline for the South Dakota winter wheat breeding program. Breeders all around the world have solely relied on highly replicative tests to evaluate the potential of a genotype. It is evident that these testing practices would be necessary to identify advance breeding lines before commercial release. However, genomic selection may result in major advantages in this process by reducing time and saving costs. Our study shows that

GS has the ability in predicting breeding values of individuals alongside saving time and expensive phenotype methodologies. Further research and software development is needed to enable widespread adoption of GS in plant breeding programs.

6. Acknowledgments

The authors would like to thank the South Dakota Agriculture Experimental Station (Brookings, SD, USA) for providing the resources to conduct the field experiments. This project was partially funded by the USDA hatch project SD00H538-15 and South Dakota Wheat Commission 3X7261.

7. Literature cited

Akbari M, Wenzl P, Caig V, Carling J, Xia L, Yang SY, Uszynski G, Mohler V, Lehmensiek A, Kuchel H, Hayden MJ, Howes N, Sharp P, Vaughan P, Rathmell B, Huttner E, Kilian A (2006) Diversity arrays technology (DArT) for high-throughput profiling of the hexaploid wheat genome. Theoretical and Applied Genetics 113 (8):1409-1420. doi:10.1007/s00122-006-0365-4

Arruda MP, Lipka AE, Brown PJ, Krill AM, Thurber C, Brown-Guedira G, Dong Y, Foresman BJ, Kolb FL (2016) Comparing genomic selection and marker-assisted selection for Fusarium head blight resistance in wheat (Triticum aestivum L.). Molecular Breeding 36 (7). doi:ARTN 84

10.1007/s11032-016-0508-5

Asoro FG, Newell MA, Beavis WD, Scott MP, Jannink JL (2011) Accuracy and Training Population Design for Genomic Selection on Quantitative Traits in Elite North American Oats. Plant Genome 4 (2):132-144. doi:10.3835/plantgenome2011.02.0007

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

Bernardo R, Yu JM (2007) Prospects for genomewide selection for quantitative traits in maize. Crop Science 47 (3):1082-1090. doi:10.2135/cropsci2006.11.0690

Breiman L (2001) Random forests. Mach Learn 45 (1):5-32. doi:Doi 10.1023/A:1010933404324

Browning SR, Browning BL (2007) Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. American Journal of Human Genetics 81 (5):1084-1097. doi:10.1086/521987

Cavanagh CR, Chao SM, Wang SC, Huang BE, Stephen S, Kiani S, Forrest K, Saintenac C, Brown-Guedira GL, Akhunova A, See D, Bai GH, Pumphrey M, Tomar L, Wong DB, Kong S, Reynolds M, da Silva ML, Bockelman H, Talbert L, Anderson JA, Dreisigacker S, Baenziger S, Carter A, Korzun V, Morrell PL, Dubcovsky J, Morell MK, Sorrells ME, Hayden MJ, Akhunov E (2013) Genome-wide comparative diversity uncovers multiple targets of selection for improvement in hexaploid wheat landraces and cultivars. Proceedings of the National Academy of Sciences of the United States of America 110 (20):8057-8062. doi:10.1073/pnas.1217133110

Dunckel S, Crossa J, Wu SY, Bonnett D, Poland J (2017) Genomic Selection for Increased Yield in Synthetic-Derived Wheat. Crop Science 57 (2):713-725. doi:10.2135/cropsci2016.04.0209

Edgerton MD (2009) Increasing Crop Productivity to Meet Global Needs for Feed, Food, and Fuel. Plant Physiology 149 (1):7-13. doi:10.1104/pp.108.130195

Endelman JB (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

Friedman J, Hastie T, Tibshirani R (2010) Regularization Paths for Generalized Linear Models via Coordinate Descent. J Stat Softw 33 (1):1-22

Gerland P, Raftery AE, Sevcikova H, Li N, Gu DA, Spoorenberg T, Alkema L, Fosdick BK, Chunn J, Lalic N, Bay G, Buettner T, Heilig GK, Wilmoth J (2014) World population stabilization unlikely this century. Science 346 (6206):234-237. doi:10.1126/science.1257469

Gupta PK, Langridge P, Mir RR (2010) Marker-assisted wheat breeding: present status and future possibilities. Molecular Breeding 26 (2):145-161. doi:10.1007/s11032-009-9359-7

Habier D, Fernando RL, Dekkers JCM (2007) The impact of genetic relationship information on genome-assisted breeding values. Genetics 177 (4):2389-2397. doi:10.1534/genetics.107.081190

Haghighattalab A, Crain J, Mondal S, Rutkoski J, Singh RP, Poland J (2017) Application of Geographically Weighted Regression to Improve Grain Yield Prediction from Unmanned Aerial System Imagery. Crop Science 57 (5):2478-2489. doi:10.2135/cropsci2016.12.1016

Heffner EL, Jannink JL, Iwata H, Souza E, Sorrells ME (2011a) Genomic Selection Accuracy for Grain Quality Traits in Biparental Wheat Populations. Crop Science 51 (6):2597-2606. doi:10.2135/cropsci2011.05.0253

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

Heffner EL, Sorrells ME, Jannink JL (2009) Genomic Selection for Crop Improvement. Crop Science 49 (1):1-12. doi:10.2135/cropsci2008.08.0512

Huang M, Cabrera A, Hoffstetter A, Griffey C, Van Sanford D, Costa J, McKendry A, Chao S, Sneller C (2016) Genomic selection for wheat traits and trait stability. Theoretical and Applied Genetics 129 (9):1697-1710. doi:10.1007/s00122-016-2733-z

Isidro J, Jannink JL, Akdemir D, Poland J, Heslot N, Sorrells ME (2015) Training set optimization under population structure in genomic selection. Theoretical and Applied Genetics 128 (1):145-158. doi:10.1007/s00122-014-2418-4

Jannink JL, Lorenz AJ, Iwata H (2010) Genomic selection in plant breeding: from theory to practice. Brief Funct Genomics 9 (2):166-177. doi:10.1093/bfgp/elq001

Jarquin D, Kocak K, Posadas L, Hyma K, Jedlicka J, Graef G, Lorenz A (2014) Genotyping by sequencing for genomic prediction in a soybean breeding population. BMC genomics 15. doi:Artn 740

10.1186/1471-2164-15-740

Jarquin D, Lemes da Silva C, Gaynor RC, Poland J, Fritz A, Howard R, Battenfield S, Crossa J (2017) Increasing Genomic-Enabled Prediction Accuracy by Modeling Genotype x Environment Interactions in Kansas Wheat. Plant Genome 10 (2). doi:10.3835/plantgenome2016.12.0130

10.3835/plantgenome2016.12.0130.

Jiang GX, Wang WJ (2017) Error estimation based on variance analysis of k-fold cross-validation. Pattern Recognition 69:94-106. doi:10.1016/j.patcog.2017.03.025

Juliana P, Singh RP, Singh PK, Crossa J, Huerta-Espino J, Lan C, Bhavani S, Rutkoski JE, Poland JA, Bergstrom GC, Sorrells ME (2017) Genomic and pedigree-based prediction for leaf, stem, and stripe rust resistance in wheat. TAG Theoretical and applied genetics Theoretische und angewandte Genetik 130 (7):1415-1430. doi:10.1007/s00122-017-2897-1

10.1007/s00122-017-2897-1. Epub 2017 Apr 9.

Kirigwi FM, van Ginkel M, Trethowan R, Sears RG, Rajaram S, Paulsen GM (2004) Evaluation of selection strategies for wheat adaptation across water regimes. Euphytica 135 (3):361-371. doi:Doi 10.1023/B:Euph.0000013375.66104.04

Kuchel H, Ye GY, Fox R, Jefferies S (2005) Genetic and economic analysis of a targeted marker-assisted wheat breeding strategy. Molecular Breeding 16 (1):67-78. doi:10.1007/s11032-005-4785-7

Lorenzana RE, Bernardo R (2009) Accuracy of genotypic value predictions for marker-based selection in biparental plant populations. Theoretical and Applied Genetics 120 (1):151-161. doi:10.1007/s00122-009-1166-3

Mayer KFX, Rogers J, Dolezel J, Pozniak C, Eversole K, Feuillet C, Gill B, Friebe B, Lukaszewski AJ, Sourdille P, Endo TR, Dolezel J, Kubalakova M, Cihalikova J, Dubska Z, Vrana J, Sperkova R, Simkova H, Rogers J, Febrer M, Clissold L, McLay K, Singh K, Chhuneja P, Singh NK, Khurana J, Akhunov E, Choulet F, Sourdille P, Feuillet C, Alberti A, Barbe V, Wincker P, Kanamori H, Kobayashi F, Itoh T, Matsumoto T, Sakai H, Tanaka T, Wu JZ, Ogihara Y, Handa H, Pozniak C, Maclachlan PR, Sharpe A, Klassen D, Edwards D, Batley J, Olsen OA, Sandve SR, Lien S, Steuernagel B, Wulff B, Caccamo M, Ayling S, Ramirez-Gonzalez RH, Clavijo BJ, Steuernagel B, Wright J, Pfeifer M, Spannagl M, Mayer KFX, Martis MM, Akhunov E, Choulet F, Mayer KFX, Mascher M, Chapman J, Poland JA, Scholz U, Barry K, Waugh R, Rokhsar DS, Muehlbauer GJ, Stein N, Gundlach H, Zytnicki M, Jamilloux V, Quesneville H, Wicker T, Mayer KFX, Faccioli P, Colaiacovo M, Pfeifer M, Stanca AM, Budak H, Cattivelli L, Glover N, Martis MM, Choulet F, Feuillet C, Mayer KFX, Pfeifer M, Pingault L, Mayer KFX, Paux E, Spannagl M, Sharma S, Mayer KFX, Pozniak C, Appels R, Bellgard M, Chapman B, Pfeifer M, Pfeifer M, Sandve SR, Nussbaumer T, Bader KC, Choulet F, Feuillet C, Mayer KFX, Akhunov E, Paux E, Rimbert H, Wang SC, Poland JA, Knox R, Kilian A, Pozniak C, Alaux M, Alfama F, Couderc L, Jamilloux V, Guilhot N, Viseux C, Loaec M, Quesneville H, Rogers J, Dolezel J, Eversole K, Feuillet C, Keller B, Mayer KFX, Olsen OA, Praud S, Iwgsc (2014) A chromosome-based draft sequence of the hexaploid bread wheat (Triticum aestivum) genome. Science 345 (6194). doi:ARTN 1251788

10.1126/science.1251788

Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics 157 (4):1819-1829

Mevik BH, Wehrens R (2007) The pls package: Principal component and partial least squares regression in R. J Stat Softw 18 (2):1-23

Michel S, Ametz C, Gungor H, Akgol B, Epure D, Grausgruber H, Loschenberger F, Buerstmayr H (2017) Genomic assisted selection for enhancing line breeding: merging genomic and phenotypic selection in winter wheat breeding programs with preliminary yield trials. Theoretical and Applied Genetics 130 (2):363-376. doi:10.1007/s00122-016-2818-8

Moose SP, Mumm RH (2008) Molecular plant breeding as the foundation for 21st century crop improvement. Plant Physiology 147 (3):969-977. doi:10.1104/pp.108.118232

Poland J, Endelman J, Dawson J, Rutkoski J, Wu SY, Manes Y, Dreisigacker S, Crossa J, Sanchez-Villeda H, Sorrells M, Jannink JL (2012a) Genomic Selection in Wheat Breeding using Genotyping-by-Sequencing. Plant Genome 5 (3):103-113. doi:10.3835/plantgenome2012.06.0006

Poland JA, Brown PJ, Sorrells ME, Jannink JL (2012b) Development of High-Density Genetic Maps for Barley and Wheat Using a Novel Two-Enzyme Genotyping-by-Sequencing Approach. PloS one 7 (2). doi:ARTN e32253

10.1371/journal.pone.0032253

Poland JA, Rife TW (2012) Genotyping-by-Sequencing for Plant Breeding and Genetics. Plant Genome 5 (3):92-102. doi:10.3835/plantgenome2012.05.0005

Pryce JE, Daetwyler HD (2012) Designing dairy cattle breeding schemes under genomic selection: a review of international research. Anim Prod Sci 52 (2-3):107-114. doi:10.1071/An11098

Rajaram S (2001) Prospects and promise of wheat breeding in the 21(st) century. Euphytica 119 (1-2):3-15. doi:Doi 10.1023/A:1017538304429

Reynolds M, Bonnett D, Chapman SC, Furbank RT, Manes Y, Mather DE, Parry MAJ (2011) Raising yield potential of wheat. I. Overview of a consortium approach and breeding strategies. Journal of Experimental Botany 62 (2):439-452. doi:10.1093/jxb/erq311

Rutkoski J, Benson J, Jia Y, Brown-Guedira G, Jannink JL, Sorrells M (2012) Evaluation of Genomic Prediction Methods for Fusarium Head Blight Resistance in Wheat. Plant Genome 5 (2):51-61. doi:10.3835/plantgenome2012.02.0001

Rutkoski J, Poland J, Mondal S, Autrique E, Perez LG, Crossa J, Reynolds M, Singh R (2016) Canopy Temperature and Vegetation Indices from High-Throughput Phenotyping Improve Accuracy of Pedigree and Genomic Selection for Grain Yield in Wheat. G3-Genes Genomes Genetics 6 (9):2799-2808. doi:10.1534/g3.116.032888

Rutkoski JE, Heffner EL, Sorrells ME (2011) Genomic selection for durable stem rust resistance in wheat. Euphytica 179 (1):161-173. doi:10.1007/s10681-010-0301-1

Shu YJ, Yu DS, Wang D, Bai X, Zhu YM, Guo CH (2013) Genomic selection of seed weight based on low-density SCAR markers in soybean. Genetics and Molecular Research 12 (3):2178-2188. doi:10.4238/2013.July.3.2

Spindel J, Begum H, Akdemir D, Virk P, Collard B, Redona E, Atlin G, Jannink JL, McCouch SR (2015) Genomic Selection and Association Mapping in Rice (Oryza sativa): Effect of Trait Genetic Architecture, Training Population Composition, Marker Number and Statistical Model on Accuracy of Rice Genomic Selection in Elite, Tropical Rice Breeding Lines. Plos Genet 11 (2). doi:ARTN e1004982

10.1371/journal.pgen.1004982

Storlie E, Charmet G (2013) Genomic Selection Accuracy using Historical Data Generated in a Wheat Breeding Program. Plant Genome 6 (1). doi:10.3835/plantgenome2013.01.0001

Sun J, Rutkoski JE, Poland JA, Crossa J, Jannink JL, Sorrells ME (2017) Multitrait, Random Regression, or Simple Repeatability Model in High-Throughput Phenotyping Data Improve Genomic Prediction for Wheat Grain Yield. Plant Genome 10 (2). doi:10.3835/plantgenome2016.11.0111

10.3835/plantgenome2016.11.0111.

Technow F, Burger A, Melchinger AE (2013) Genomic Prediction of Northern Corn Leaf Blight Resistance in Maize with Combined or Separated Training Sets for Heterotic Groups. G3-Genes Genomes Genetics 3 (2):197-203. doi:10.1534/g3.112.004630

Van Inghelandt D, Reif JC, Dhillon BS, Flament P, Melchinger AE (2011) Extent and genome-wide distribution of linkage disequilibrium in commercial maize germplasm. Theoretical and Applied Genetics 123 (1):11-20. doi:10.1007/s00122-011-1562-3

VanRaden PM, Van Tassell CP, Wiggans GR, Sonstegard TS, Schnabel RD, Taylor JF, Schenkel FS (2009) Invited review: Reliability of genomic predictions for North American Holstein bulls. J Dairy Sci 92 (1):16-24. doi:10.3168/jds.2008-1514

Vargas M, Combs E, Alvarado G, Atlin G, Mathews K, Crossa J (2013) META: A Suite of SAS Programs to Analyze Multienvironment Breeding Trials. Agron J 105 (1):11-19. doi:10.2134/agronj2012.0016

Varshney RK, Singh VK, Hickey JM, Xun X, Marshall DF, Wang J, Edwards D, Ribaut JM (2016) Analytical and Decision Support Tools for Genomics-Assisted Breeding. Trends in Plant Science 21 (4):354-363. doi:10.1016/j.tplants.2015.10.018

Villumsen TM, Janss L, Lund MS (2009) The importance of haplotype length and heritability using genomic selection in dairy cattle. J Anim Breed Genet 126 (1):3-13. doi:10.1111/j.1439-0388.2008.00747.x

Wang Y, Mette MF, Miedaner T, Gottwald M, Wilde P, Reif JC, Zhao YS (2014) The accuracy of prediction of genomic selection in elite hybrid rye populations surpasses the accuracy of marker-assisted selection and is equally augmented by multiple field evaluation locations and test years. BMC genomics 15. doi:Artn 556

10.1186/1471-2164-15-556

Wrigley CW (1994) Developing Better Strategies to Improve Grain Quality for Wheat. Australian Journal of Agricultural Research 45 (1):1-17. doi:Doi 10.1071/Ar9940001

Wurschum T, Reif JC, Kraft T, Janssen G, Zhao YS (2013) Genomic selection in sugar beet breeding populations. Bmc Genetics 14. doi:Artn 85

10.1186/1471-2156-14-85

Xu YB, Crouch JH (2008) Marker-assisted selection in plant breeding: From publications to practice. Crop Science 48 (2):391-407. doi:10.2135/cropsci2007.04.0191

Zhao YS, Gowda M, Liu WX, Wurschum T, Maurer HP, Longin FH, Ranc N, Reif J (2012) Accuracy of genomic selection in European maize elite breeding populations. Theoretical and Applied Genetics 124 (4):769-776. doi:10.1007/s00122-011-1745-y

Zhong SQ, Dekkers JCM, Fernando RL, Jannink JL (2009) Factors Affecting Accuracy From Genomic Selection in Populations Derived From Multiple Inbred Lines: A Barley Case Study. Genetics 182 (1):355-364. doi:10.1534/genetics.108.098277

APPENDICES

Appendix Figure 1. Demographic distribution of the 300 winter wheat lines constituting the Association Mapping Panel.

Appendix Table 1. The reaction of the HWWAMP to BLS in the greenhouse and field experiments.

Line information		Greenhouse experiments					Field	BLUEs
		Experiment 1					Experiment 1	
Genotype	ID	Rep 1	Rep 2	Rep 3	Rep 4	Mean	Mean	
1	**TRIUMPH64**	3	3	3	3	3	60	3
2	**CHISHOLM**	2	2	2	2	2	40	2
3	**CENTURY**	3	3	3	3	3	60	3
4	**CUSTER**	3	3	3	3	3	60	3
5	**2174-05**	3	3	3	3	3	60	3
6	**INTRADA**	3	3	3	3	3	60	3
7	**OK101**	3	3	3	3	3	50	3
8	**OK102**	2	2	2	2	2	30	2
9	**ENDURANCE**	3	3	3	3	3	60	3
10	**DELIVER**	3	3	3	3	3	60	3
11	**OK_BULLET**	2	2	2	2	2	40	2
12	**CENTERFIELD**	2	2	2	2	2	40	2
13	**GUYMON**	3	3	3	3	3	60	3
14	**DUSTER**	3	3	3	3	3	60	3
15	**OK_RISING**	3	3	3	3	3	60	3
16	**OK02405**	3	3	3	3	3	60	3

17	**PETE**	3	3	3	3	3	60	3
18	**BILLINGS**	3	3	3	3	3	60	3
19	**OK04505**	2	2	2	2	2	40	2
20	**OK04525**	2	2	2	2	2	30	2
21	**OK04507**	3	3	3	3	3	60	3
22	**OK05830**	3	3	3	3	3	60	3
23	**OK04111**	3	3	3	3	3	60	3
24	**OK04415**	2	2	2	2	2	40	2
25	**OK05711W**	3	3	3	3	3	60	3
26	**OK05723W**	1	1	1	1	1	5	1
27	**OK05108**	4	4	4	4	4	90	5
28	**OK05122**	3	3	3	3	3	60	3
29	**OK05526**	4	4	4	4	4	90	5
30	**OK05134**	2	2	2	2	2	30	2
31	**OK05303**	3	3	3	3	3	60	3
32	**OK05312**	3	3	3	3	3	60	3
33	**OK05511**	2	2	2	2	2	40	2
34	**OK05204**	2	2	2	2	2	40	2
35	**GARRISON**	2	2	2	2	2	30	2
36	**OK06114**	2	2	2	2	2	40	2
37	**OK06210**	2	2	2	2	2	40	2

38	**OK06319**	4	4	4	4	4	90	5
39	**OK06318**	3	3	3	3	3	60	3
40	**OK06336**	3	3	3	3	3	60	3
41	**AGATE**	3	3	3	3	3	60	3
42	**ALLIANCE**	3	3	3	3	3	60	3
43	**ANTELOPE**	5	5	5	5	5	90	5
44	**ARAPAHOE**	3	3	3	3	3	60	3
45	**BENNETT**	3	3	3	3	3	60	3
46	**BUCKSKIN**	2	2	2	2	2	40	2
47	**CENTURA**	2	2	2	2	2	30	2
48	**CENTURK78**	3	3	3	3	3	60	3
49	**CHEYENNE**	3	3	3	3	3	60	3
50	**COLT**	2	2	2	2	2	40	2
51	**COUGAR**	2	2	2	2	2	40	2
52	**CULVER**	2	2	2	2	2	40	2
53	**GAGE**	2	2	2	2	2	30	2
54	**GOODSTREAK**	1	1	1	1	1	5	1
55	**HALLAM**	2	2	2	2	2	30	2
56	**HARRY**	2	2	2	2	2	40	2
57	**HOMESTEAD**	2	2	2	2	2	30	2.5
58	**INFINITY_CL**	2	2	2	2	2	30	2.5

59	**KHARKOF**	3	3	3	3	3	60	3
60	**MILLENNIUM**	2	2	2	2	2	40	3
61	**CAMELOT**	3	3	3	3	3	60	3
62	**OVERLAND**	2	2	3	2	2.5	40	3
63	**NE99495**	3	3	3	3	3	60	3
64	**NIOBRARA**	2	2	3	2	2.5	40	3
65	**NUPLAINS**	3	3	3	3	3	60	3
66	**PRONGHORN**	3	3	3	3	3	60	3
67	**RAWHIDE**	2	2	3	2	2.5	20	3
68	**REDLAND**	3	3	3	3	3	60	3
69	**SCOUT66**	1	1	1	1	1	5	1
70	**SIOUXLAND**	3	4	3	3	3.5	60	4
71	**TURKEY_NEBSEL**	2	2	3	2	2.5	40	3
72	**VISTA**	1	1	1	1	1	5	1
73	**WAHOO**	2	2	3	2	2.5	30	3
74	**WARRIOR**	2	2	3	2	2.5	40	3
75	**WESLEY**	3	4	3	3	3.5	60	4
76	**WICHITA**	3	4	3	3	3.5	60	3.5
77	**WINDSTAR**	3	4	3	3	3.5	50	4
78	**JAGGER**	3	4	3	3	3.5	60	4
79	**LANCER**	2	2	3	2	2.5	40	3

80	**SETTLER_CL**	2	2	3	2	2.5	40	3
81	**ANTON**	3	4	4	4	4	60	4
82	**MACE**	3	4	4	4	4	60	4
83	**JERRY**	3	4	4	4	4	60	4
84	**TAM107-R7**	5	5	5	5	5	80	5
85	**ARLIN**	3	4	4	4	4	60	4
86	**ALICE**	2	2	3	2	2.5	40	3
87	**DARRELL**	2	2	3	2	2.5	30	3
88	**EXPEDITION**	2	2	3	2	2.5	40	3
89	**WENDY**	1	1	1	1	1	5	1
90	**SD00111-9**	2	3	3	2	2.5	40	3
91	**SD01237**	2	3	3	3	3	40	3
92	**SD01058**	3	4	4	4	4	60	4
93	**SD05118**	4	4	4	4	4	60	4
94	**SD05210**	2	3	3	3	3	40	3
95	**SD05W018**	2	3	3	3	3	40	3
96	**NEKOTA**	2	3	3	3	3	40	3
97	**TANDEM**	4	4	4	4	4	60	4
98	**CRIMSON**	4	4	4	4	4	60	4
99	**ROSE**	4	4	4	4	4	50	4
100	**DAWN**	2	3	3	3	3	30	3

101	**WINOKA**	2	3	3	3	3	30	3
102	**NELL**	5	5	5	5	5	90	5
103	**RITA**	4	4	4	4	4	60	4
104	**BRONZE**	2	3	3	3	3	40	3
105	**HUME**	2	3	3	3	3	40	3
106	**GENT**	4	4	4	4	4	60	4
107	**HARDING**	2	3	3	3	3	40	3
108	**HV9W03-1551WP**	4	4	4	4	4	60	4
109	**G1878**	4	4	4	4	4	60	4
110	**HV9W03-1379R**	2	3	3	3	3	40	3
111	**HV9W03-1596R**	4	4	4	4	4	60	4
112	**HV9W05-1280R**	4	4	4	4	4	60	4
113	**HV9W06-504**	4	4	4	4	4	60	4
114	**SPARTAN**	4	4	4	4	4	60	4
115	**HV906-865**	2	3	3	3	3	40	3
116	**TARKIO**	4	4	4	4	4	50	4
117	**SMOKYHILL**	4	4	4	4	4	50	4
118	**SHOCKER**	4	4	4	4	4	60	4
119	**VONA**	2	3	3	3	3	30	3
120	**CO940610**	2	3	3	3	3	40	3
121	**AVALANCHE**	4	4	4	4	4	60	4

122	**BOND_CL**	4	4	4	4	4	60	4
123	**PLATTE**	2	3	3	3	3	30	3
124	**LINDON**	4	4	4	4	4	60	4
125	**CO03W043**	2	3	3	3	3	40	3
126	**CO03W054**	4	4	4	4	4	60	4
127	**THUNDER_CL**	2	3	3	3	3	40	3
128	**CO04025**	4	4	4	4	4	60	4
129	**CO04393**	2	3	3	3	3	40	3
130	**CO04499**	4	4	4	4	4	60	4
131	**CO04W320**	4	4	4	4	4	50	4
132	**LAMAR**	2	3	3	3	3	40	3
133	**CARSON**	4	4	4	4	4	60	4
134	**HAIL**	4	4	4	4	4	60	4
135	**SANDY**	3	3	3	3	3	40	3
136	**DUKE**	4	4	4	4	4	60	4
137	**HALT**	3	3	3	3	3	30	3
138	**HATCHER**	3	3	3	3	3	30	3
139	**PRAIRIE_RED**	3	3	3	3	3	40	3
140	**YUMAR**	4	4	4	4	4	60	4
141	**ABOVE**	3	3	3	3	3	40	3
142	**CO03064**	3	3	3	3	3	30	3

143	**BILL_BROWN**	3	3	3	3	3	30	3
144	**RIPPER**	4	4	4	4	4	60	4
145	**PROWERS**	3	3	3	3	3	40	3
146	**AKRON**	4	4	4	4	4	60	4
147	**JULES**	4	4	4	4	4	60	4
148	**YUMA**	4	4	4	4	4	60	4
149	**TAMW-101**	4	4	4	4	4	60	4
150	**TAM105**	4	4	4	4	4	60	4
151	**TAM107**	3	3	3	3	3	30	3
152	**TAM109**	4	4	4	4	4	60	4
153	**TAM110**	4	4	4	4	4	60	4
154	**TAM111**	4	4	4	4	4	60	4
155	**TAM112**	3	3	3	3	3	40	3
156	**TAM200**	4	4	4	4	4	60	4
157	**TAM202**	3	3	3	3	3	30	3
158	**TAM203**	4	4	4	4	4	60	4
159	**TAM302**	3	3	3	3	3	40	3
160	**TAM303**	4	4	4	4	4	50	4
161	**TAM304**	5	5	5	5	5	90	5
162	**TAM400**	3	3	3	3	3	40	3
163	**LOCKETT**	5	5	5	5	5	90	5

164	**STURDY**	3	3	3	3	3	40	3
165	**STURDY_2K**	4	4	4	4	4	60	4
166	**MIT**	5	5	5	5	5	90	5
167	**CAPROCK**	3	3	3	3	3	30	3
168	**TX01A5936**	4	4	4	4	4	50	4
169	**TAM401**	3	3	3	3	3	40	3
170	**TX02A0252**	3	3	3	3	3	40	3
171	**TX03A0148**	3	3	3	3	3	40	3
172	**TX03A0563**	3	3	3	3	3	30	3
173	**TX04A001246**	3	3	3	3	3	30	3
174	**TX01V5134RC-3**	4	4	4	4	4	60	4
175	**TX04M410164**	4	4	4	4	4	50	4
176	**TX04M410211**	3	3	3	3	3	40	3
177	**TX04V075080**	4	4	4	4	4	60	4
178	**TX99A0153-1**	3	3	3	3	3	40	3
179	**TX01M5009-28**	3	3	3	3	3	40	3
180	**TX00V1131**	3	3	3	3	3	30	3
181	**TX99U8618**	3	3	3	3	3	30	3
182	**TX96D1073**	4	4	4	4	4	60	4
183	**2180**	4	4	4	4	4	60	4
184	**HG-9**	4	4	4	4	4	60	4

185	**TX86A5606**	4	4	4	4	4	60	4
186	**TX86A6880**	3	3	3	3	3	30	3
187	**TX86A8072**	4	4	4	4	4	60	4
188	**CREST**	4	4	4	4	4	60	4
189	**ROSEBUD**	4	4	4	4	4	60	4
190	**JUDITH**	3	3	3	3	3	30	3
191	**MT85200**	4	4	4	4	4	60	4
192	**NUSKY**	3	3	3	3	3	40	3
193	**MT9513**	4	4	4	4	4	60	4
194	**MT9904**	3	3	3	3	3	40	3
195	**MT9982**	3	3	3	3	3	40	3
196	**GENOU**	4	4	4	4	4	60	4
197	**NORRIS**	4	4	4	4	4	60	4
198	**YELLOWSTONE**	3	3	3	3	3	40	3
199	**MT0495**	4	4	4	4	4	60	4
200	**MTS0531**	4	4	4	4	4	60	4
201	**DECADE**	3	3	3	3	3	40	3
202	**MT06103**	4	4	4	4	4	60	4
203	**JUDEE**	3	3	3	3	3	40	3
204	**LAKIN**	3	3	3	3	3	30	3
205	**STANTON**	4	4	4	4	4	60	4

206	**TREGO**	3	3	3	3	3	40	3
207	**KARL_92**	4	4	4	4	4	60	4
208	**DODGE**	3	3	3	3	3	40	3
209	**NORKAN**	3	3	3	3	3	40	3
210	**CHENEY**	3	3	3	3	3	40	3
211	**NEWTON**	3	3	3	3	3	40	3
212	**LARNED**	1	1	1	1	1	5	1
213	**PARKER76**	3	3	3	3	3	30	3
214	**KIRWIN**	4	4	4	4	4	50	4
215	**SAGE**	3	3	3	3	3	40	3
216	**TRISON**	4	4	4	4	4	60	4
217	**EAGLE**	1	1	1	1	1	5	1
218	**SHAWNEE**	4	4	4	4	4	60	4
219	**PARKER**	3	3	3	3	3	40	3
220	**KAW61**	4	4	4	4	4	60	4
221	**TASCOSA**	4	4	4	4	4	60	4
222	**BISON**	4	4	4	4	4	60	4
223	**KIOWA**	4	4	4	4	4	60	4
224	**WICHITA**	3	3	3	3	3	40	3.5
225	**COMANCHE**	3	3	3	3	3	30	3
226	**BAKERS_WHITE**	4	4	4	4	4	60	4

227	**BURCHETT**	3	3	3	3	3	40	3
228	**CUTTER**	3	3	3	3	3	40	3
229	**DUMAS**	4	4	4	4	4	60	4
230	**HONDO**	4	4	4	4	4	60	4
231	**JAGALENE**	3	3	3	3	3	30	3
232	**LONGHORN**	3	3	3	3	3	40	3
233	**NEOSHO**	4	4	4	4	4	60	4
234	**OGALLALA**	4	4	4	4	4	60	4
235	**POSTROCK**	3	3	3	3	3	40	3
236	**THUNDERBOLT**	3	3	3	3	3	20	3
237	**W04-417**	4	4	4	4	4	60	4
238	**NUFRONTIER**	3	3	3	3	3	40	3
239	**NUHORIZON**	3	3	3	3	3	40	3
240	**ONAGA**	3	3	3	3	3	40	3
241	**RONL**	3	3	3	3	3	40	3
242	**2145**	3	3	3	3	3	40	3
243	**HEYNE**	3	3	3	3	3	30	3
244	**KS00F5-20-3**	4	4	4	4	4	60	4.5
245	**OVERLEY**	4	4	4	4	4	60	5
246	**FULLER**	3	3	3	3	3	40	3
247	**COSSACK**	4	5	4	4	4.5	60	5

248	**ENHANCER**	3	3	3	3	3	40	3
249	**SANTA_FE**	4	5	4	4	4.5	60	5
250	**VENANGO**	4	5	4	4	4.5	60	5
251	**WB411W**	3	3	3	3	3	30	3
252	**KEOTA**	3	3	3	3	3	40	3
253	**TX05A001822**	4	5	4	4	4.5	60	5
254	**TX06A001263**	3	3	3	3	3	40	3
255	**TX06A001132**	3	3	3	3	3	40	3
256	**TX06A001281**	4	5	4	4	4.5	60	5
257	**TX06A001386**	3	3	3	3	3	40	3
258	**TX05V7259**	4	5	5	5	5	60	5
259	**TX05V7269**	3	3	3	3	3	40	3
260	**TX05A001188**	3	3	3	3	3	40	3
261	**TX07A001279**	4	5	5	5	5	60	5
262	**TX07A001318**	3	3	3	3	3	40	3
263	**TX07A001420**	1	1	1	1	1	5	1
264	**TX06V7266**	3	3	3	3	3	30	3
265	**OK1067071**	3	3	3	3	3	30	3
266	**OK1067274**	4	5	5	5	5	60	5
267	**OK1068002**	3	3	3	3	3	40	3
268	**OK1068009**	5	5	5	5	5	60	5

269	**OK1068026**	3	3	3	3	3	40	3
270	**OK1068112**	1	1	1	1	1	5	1
271	**OK1070275**	3	3	3	3	3	30	3
272	**OK1070267**	5	5	5	5	5	60	5
273	**OK09634**	5	5	5	5	5	60	5
274	**OK10119**	5	5	5	5	5	60	5
275	**GALLAGHER**	3	3	3	3	3	30	3
276	**OK07231**	3	3	3	3	3	30	3
277	**OK07S117**	3	3	3	3	3	40	3
278	**OK08328**	5	5	5	5	5	60	5
279	**BIG_SKY**	3	3	3	3	3	40	3
280	**DANBY**	3	3	3	3	3	30	3
281	**E2041**	3	3	3	3	3	30	3
282	**DENALI**	5	5	5	5	5	50	5
283	**CO050337-2**	3	3	3	3	3	30	3
284	**BYRD**	5	5	5	5	5	60	5
285	**CO07W245**	3	3	3	3	3	30	3
286	**MCGILL**	3	3	3	3	3	40	3
287	**NE02558**	3	3	3	3	3	40	3
288	**NW03666**	5	5	5	5	5	60	5
289	**NE04490**	1	1	1	1	1	5	1

290	**NE05430**	3	3	3	3	3	40	3
291	**NE05496**	3	3	3	3	3	40	3
292	**NE05548**	5	5	5	5	5	60	5
293	**NE06545**	5	5	5	5	5	60	5
294	**NE06607**	5	5	5	5	5	60	5
295	**ROBIDOUX**	1	1	1	1	1	5	1
296	**NI06736**	3	3	3	3	3	40	3
297	**NI06737**	5	5	5	5	5	60	5
298	**NI07703**	3	3	3	3	3	40	3
299	**NI08707**	3	3	3	3	3	30	3
300	**NI08708**	3	3	3	3	3	40	3

Appendix Table 3. Prediction accuracies obtained for validation populations 2015, 2016 and 2017 PYT nurseries through the single year analysis. This analysis was across all locations using all the four algorithms.

	2015 PYT Nursery ‡				2016 PYT Nursery †				2017 PYT Nursery ¥			
rrBLUP												
Location	**Aurora**	**D Lakes**	**Onida**	**Wall**	**Aurora**	**D Lakes**	**Onida**	**Wall**	**Aurora**	**D Lakes**	**Onida**	**Wall**
D Lakes	0.73				0.7				0.64			
Onida	0.58	0.65			0.57	0.68			0.82	0.54		
Wall	0.53	0.52	0.47		0.74	0.78	0.55		0.57	0.48	0.66	
Winner	0.78	0.7	0.82	0.53	0.64	0.64	0.54	0.78	0.84	0.5	0.53	0.48
PLSR												
Location	**Aurora**	**D Lakes**	**Onida**	**Wall**	**Aurora**	**D Lakes**	**Onida**	**Wall**	**Aurora**	**D Lakes**	**Onida**	**Wall**
D Lakes	0.58				0.51				0.85			
Onida	0.49	0.56			0.54	0.78			0.63	0.63		
Wall	0.55	0.72	0.6		0.68	0.52	0.77		0.77	0.62	0.81	
Winner	0.61	0.73	0.74	0.52	0.65	0.57	0.45	0.5	0.72	0.78	0.7	0.55

ELNET

Location	Aurora	D Lakes	Onida	Wall	Aurora	D Lakes	Onida	Wall	Aurora	D Lakes	Onida	Wall
D Lakes	0.73				0.85				0.59			
Onida	0.58	0.49			0.45	0.73			0.53	0.45		
Wall	0.53	0.82	0.63		0.66	0.79	0.74		0.62	0.59	0.76	
Winner	0.78	0.54	0.45	0.62	0.71	0.63	0.52	0.68	0.7	0.49	0.52	0.57

Random Forest

Location	Aurora	D Lakes	Onida	Wall	Aurora	D Lakes	Onida	Wall	Aurora	D Lakes	Onida	Wall
D Lakes	0.53				0.84				0.47			
Onida	0.58	0.61			0.67	0.49			0.67	0.5		
Wall	0.53	0.77	0.8		0.82	0.79	0.76		0.58	0.8	0.83	
Winner	0.56	0.6	0.55	0.66	0.6	0.76	0.46	0.61	0.67	0.77	0.64	0.77

The training populations are ‡ AYT (2015), † AYT (2016) and ¥ AYT (2017).

Appendix Table 4. Prediction accuracies obtained for validation populations 2016 and 2017 PYT nurseries through the multiple year analysis (two – year combination). This analysis was across all locations using rrBLUP, PLSR, ENLET and Random Forest prediction algorithms.

(TP) 2015 (AYT, PYT) + 2016 (AYT) = (VP) 2016 (PYT)					**(TP) 2016 (AYT, PYT) + 2017 (AYT) = (VP) 2017 (PYT)**				
rrBLUP									
Location	**Aurora**	**D Lakes**	**Onida**	**Wall**	**Location**	**Aurora**	**D Lakes**	**Onida**	**Wall**
D Lakes	0.32				**D Lakes**	0.3			
Onida	0.34	0.32			**Onida**	0.21	0.27		
Wall	0.25	0.27	0.26		**Wall**	0.23	0.28	0.21	
Winner	0.32	0.25	0.26	0.28	**Winner**	0.28	0.3	0.33	0.25
PLSR									
Location	**Aurora**	**D Lakes**	**Onida**	**Wall**	**Location**	**Aurora**	**D Lakes**	**Onida**	**Wall**
D Lakes	0.2				**D Lakes**	0.26			
Onida	0.23	0.28			**Onida**	0.31	0.2		
Wall	0.31	0.3	0.28		**Wall**	0.33	0.27	0.31	
Winner	0.29	0.29	0.22	0.34	**Winner**	0.35	0.29	0.28	0.35

ENLET

Location	Aurora	D Lakes	Onida	Wall	Location	Aurora	D Lakes	Onida	Wall
D Lakes	0.32				D Lakes	0.22			
Onida	0.28	0.2			Onida	0.28	0.25		
Wall	0.2	0.28	0.3		Wall	0.32	0.34	0.3	
Winner	0.33	0.31	0.24	0.2	Winner	0.33	0.31	0.3	0.29

Random Forest

Location	Aurora	D Lakes	Onida	Wall	Location	Aurora	D Lakes	Onida	Wall
D Lakes	0.24				D Lakes	0.31			
Onida	0.31	0.2			Onida	0.27	0.22		
Wall	0.27	0.2	0.21		Wall	0.25	0.34	0.23	
Winner	0.22	0.29	0.26	0.23	Winner	0.2	0.23	0.34	0.21

Appendix Table 5. Prediction accuracies obtained for validation populations 2016 and 2017 PYT nurseries through the multiple year analysis (three – year combination). This analysis was across all locations using rrBLUP, PLSR, ENLET and Random Forest prediction algorithms.

2015 (AYT, PYT) + 2016 (AYT, PYT) + 2017 (AYT) = 2017 (PYT)									
	rrBLUP					PLSR			
Location	**Aurora**	**D Lakes**	**Onida**	**Wall**	**Location**	**Aurora**	**D Lakes**	**Onida**	**Wall**
D Lakes	0.42				**D Lakes**	0.32			
Onida	0.28	0.34			**Onida**	0.26	0.4		
Wall	0.32	0.43	0.42		**Wall**	0.25	0.39	0.29	
Winner	0.37	0.34	0.29	0.37	**Winner**	0.31	0.35	0.37	0.32
	ELNET					**Random Forest**			
Location	**Aurora**	**D Lakes**	**Onida**	**Wall**	**Location**	**Aurora**	**D Lakes**	**Onida**	**Wall**
D Lakes	0.36				**D Lakes**	0.4			
Onida	0.39	0.43			**Onida**	0.35	0.26		
Wall	0.29	0.33	0.41		**Wall**	0.29	0.26	0.33	
Winner	0.41	0.4	0.41	0.4	**Winner**	0.4	0.31	0.28	0.29