

USING MODERN GENOMIC TECHNOLOGIES TO IMPROVE MONTANA SPRING
WHEAT

by
Jared Sherman Lile

A thesis submitted in partial fulfillment
of the requirements for the degree

of
Doctor of Philosophy
in
Plant Science

MONTANA STATE UNIVERSITY
Bozeman, Montana

May 2026

©COPYRIGHT

by

Jared Sherman Lile

2026

All Rights Reserved

DEDICATION

This thesis is dedicated to my friends, my family, and everyone who supported me on my journey to get here.

ACKNOWLEDGEMENTS

I would like to thank my advisors Dr. Jennifer Lachowiec and Dr. Jason Cook for years of continual support, training, and belief that I could succeed. Thank you to my committee members Dr. Sue Mondal, Dr. Shinjini Nandi, and my former committee member Dr. Venu Mendu for support, feedback, and pushing me to produce the best possible work that I could. I want to thank my numerous fellow lab members for their friendship and support. I would like to thank the USDA National Institute of Food and Agriculture for funding this work. A huge thanks to the WheatCAP group for training, education, and for being a continual example of excellence in the field of wheat breeding. Finally, a huge thank you to my friends, family, and church for emotional support, love, and fellowship over these last five years.

TABLE OF CONTENTS

1. INTRODUCTION	1
References.....	10
2. FINE MAPPING A QTL FOR PRODUCTIVE TILLER NUMBER.....	21
Introduction.....	21
Materials and Methods.....	22
Generating Plant Materials	22
Greenhouse Phenotyping Trials	24
Greenhouse Data Analysis	25
New Greenhouse Grow Lights Pilot Study.....	26
RNA-seq Analysis.....	28
RNA-seq Variant Calling with GATK	29
Orthologous Gene Identification.....	29
Results.....	31
Discussion	41
Greenhouse Phenotyping Trials	41
Light Regulation of Tiller Number	42
RNA-seq Analysis.....	44
Orthologous Gene Analysis	44
Conclusion	47
References.....	48
3. EVALUATING REACTION NORM AND FIXED EFFECT MODELS FOR GENOMIC SELECTION OF END-USE QUALITY IN WHEAT	56
Contribution of Authors and Co-Authors	56
Manuscript Information	57
Abstract	58
Introduction.....	58
Materials and Methods.....	61
Phenotypic Data Collection	61
Genetic Data Collection.....	63
Two-Step Analysis Method.....	63
GWAS	65
Known QTL as Fixed Effects	65
Single-Step Analysis Method.....	66
Cross-Validation.....	67
Results.....	68
Two-Step Analysis Results	69
Single-Step Analysis Results	72

TABLE OF CONTENTS CONTINUED

Discussion	74
References	79
4. GENOMIC SELECTION FOR WHEAT STEM SAWFLY RESISTANCE	86
Introduction	86
Methods	88
Phenotypic Data Collection	88
Genetic Data Collection	93
Genomic Selection Models	94
Model Validation	95
Genome Wide Association Study	96
Epistatic Analyses	97
Removal of the F6 Data	97
Results	97
Discussion	103
Conclusion	108
References	108
5. CONCLUSION	115
CUMULATIVE REFERENCES CITED	118

LIST OF TABLES

Table	Page
1. Table 1.1. Results for Type II analysis of deviance for early tiller number collected at three different time points across four different experiments. Three separate generalized linear mixed models were fit; one for each time point of collection shown in column 1. Column 2 gives the model fixed effect term for the given time point, column 3 gives the associated Wald chi-square test statistic for the model term, column 4 gives the degrees of freedom for the chi-square test, and column 5 gives the unadjusted the p-values.	33
2. Table 1.2. Table of possible orthologous tillering genes from rice and <i>Arabidopsis</i> as well as potential tillering genes already identified in wheat. The gene names, chromosomes, and positions are provided from IWGSC RefSeq v2.1 as well as the mean expression, differential expression adjusted p-value, and detected variants from the RNA-seq data. Bolded rows indicate genes that are located on chromosome 6B.....	41
3. Table 3.1. Summary statistics and Cullis broad-sense heritability values for each year of the full phenotypic sawfly resistance dataset (including the F6 trials).	98
4. Table 3.2. Forward validation results for predicting sawfly resistance using seven genomic selection models. The first five columns denote results from the full dataset while the last five columns are results from the reduced dataset with the F6 trials removed. The first column in each section gives the GS model tested followed by the prediction accuracy in the second column. The third column gives the relative sawfly resistance in percent of the most resistant line that was selected for removal by the model compared to the mean resistance of all the predicted lines. The fourth column gives the mean resistance of all the lines selected for removal by the model relative to the overall mean of the predicted lines. Finally, the last column in each section denotes the number of lines selected for removal by the model that have true resistance values in the best 25% of all the predicted lines (designated “elites” relative to the group).	100
5. Table 3.3. MLM GWAS results for the full dataset and the F6 removed dataset. Significant markers (Bonferroni threshold $p < 0.05$) are given in the first column with their corresponding chromosomes, positions, minor allele frequencies (MAF), estimated effect of the alternate allele on percent lodging, standard error of the effect (SE) and raw p-value.....	103

LIST OF FIGURES

Figure	Page
1. Figure 1.1. A) Table of 14 entries evaluated in greenhouse trials and their respective haplotypes. Entries 1-12 consist of F ₆ -HIF lines that either non-recombinant (haplotypes A, B) or recombinant (haplotypes C, D). Entries 13 and 14 are the original NIL sister lines crossed to develop the F ₆ -HIF population and serve as the check lines for the experiments. B) Table of genotypes across eight different KASP markers spanning the narrow QTL region of interest. Genotypes are either “1” representing the allele of the high-ETN NIL sister line, or “2” for the allele of the low-ETN NIL sister line. Haplotypes A, B, C, and D represent non-recombinant (A,B) and recombinant (C,D) groupings.	24
2. Figure 1.2. Light output metrics between the metal halide grow lights (MVR1000 Multi-Vapor; top chart) and the replacement LED lights (Arize Element L1000 Gen2; bottom chart). Graphs show the measured photon flux density at each light wavelength from 300 to 900 nm. The total photon output (Integrated Total), fraction of total output captured, photosynthetic photon flux (PPF), yield photon flux (YPF), photon efficacy (PPE), and Red to Far Red light ratio are also reported for each grow light.	27
3. Figure 1.3. Early tiller number (ETN) counts are given for four, five, and six weeks after planting. Data is combined from four different greenhouse trials, and a single point indicates one plant-level observation within a pot, block, and trial. Different entries are denoted on the X-axis, and entries are grouped within shared haplotypes by the letters at the top of the plots. The blue horizontal line indicates the overall mean tiller number within the time after planting, across all trials. The red horizontal bars indicate the entry-level mean tiller numbers across all trials within the given time after planting.	32
4. Figure 1.4. Estimated marginal log mean early tiller numbers are given between different haplotypes on the X-axis grouped within one of four greenhouse trials with data taken either four, five, or six weeks after planting. Bars indicate 95% confidence intervals around the estimated means and letters indicate grouping based on pairwise contrasts between haplotypes within a given trial with p-values adjusted by Sidak’s method.	34

LIST OF FIGURES CONTINUED

Figure	Page
5. Figure 1.5. Results from the winter 2023 pilot study using the new LED grow lights. The left figure shows the raw ETN counts for each plant grouped by haplotype on the X-axis. The blue line indicates the overall mean ETN and individual red bars indicate the mean tiller counts within one haplotype. The right panel shows estimated log mean ETN values for the haplotype groups with bars representing 95% confidence intervals. Letters indicate grouping based on pairwise contrasts between haplotype groups with p-values adjusted by Sidak's method.....	35
6. Figure 1.6. A) Differential expression analysis results from RNA-seq between high and low tiller NIL sister lines. Each point represents one gene on chromosome 6B of spring wheat. The Y-axis gives the log base 2 fold expression change of a gene in the high tiller line compared to the low tiller line. The X-axis gives the base pair position of the gene. A red point indicates the gene was differentially expressed (adjusted p-value > 0.01) whereas blue indicates the gene was not. The green vertical bars indicate the original range of <i>QTn.mst-6B</i> where the orange, vertical lines indicate the fine-mapped narrow region of the QTL. B) Table of the four DEGs detected in the narrow QTL range, including their names, positions, associated statistics and functional annotations.....	36
7. Figure 2.1. A) Table of broad-sense heritabilities for twelve end-use quality traits calculated using the Cullis method. B) Pearson correlation matrix between the BLUEs of the same twelve end-use quality traits with darker red indicating a stronger, positive correlation and darker blue indicating a stronger, negative correlation.....	69
8. Figure 2.2. Five-fold cross-validation results (n=100) for twelve end-use quality traits. The mean prediction accuracy is given on the y-axis and error bars denote one standard deviation. Colors represent results from one of seven different genomic selections that were tested.	70

LIST OF FIGURES CONTINUED

Figure	Page
9. Figure 2.3. Five-fold cross-validation results (n=100) for the two-step method showing the mean prediction accuracy for twelve different end-use quality traits. The BRR genomic selection model was used in all cases with different colors representing different fixed effect specifications. The first, orange bars are the baseline model with no added fixed effect terms which are compared against the inclusion of markers associated to potential known QTL, markers identified by GWAS on each training fold, markers identified by GWAS on the full phenotypic BLUEs dataset, and 20 or 3 principal components of a PCA across all markers associated with a potential known QTL.....	71
10. Figure 2.4. Leave-one-year-out (LOYO) validation results for the two-step analysis approach. The x-axis denotes the year in which the experimental lines were left out of the training data and predicted. The y-axis shows the accuracy as the correlation between the predicted breeding values and known BLUEs for the experimental lines in the left-out year. Each box represents one end-use quality trait and for each trait, results are given for seven different GS models denoted with different colors.	72
11. Figure 2.5. Five-fold cross-validation results (n=100) for the single-step analysis method for twelve end-use quality traits. The mean prediction accuracy across all cycles of cross-validation, within a location-year, represented by each point. The x-axis denotes the quality trait tested and each color represents a different genomic selection model. Box plots show the overall median of the means prediction accuracy for each trait across all location-years and models.....	73
12. Figure 2.6. Leave-one-year-out (LOYO) validation results for the single-step method of analysis. The correlation between predicted and observed phenotypic values within a location-trial are shown on the y-axis with boxplots showing the median and range of correlations between different locations within the left-out year, denoted by the x-axis. Two representative genomic selection models (BRR and RKHS) are shown in different colors.	74

LIST OF FIGURES CONTINUED

Figure	Page
13. Figure 3.1. Incidence plot of each individual line phenotyped for sawfly resistance on the X-axis and the phenotyping trials in which a line was observed on the Y-axis for each year of data collected. The top bar indicates if a line was observed only one time across all trials (grey color), if the line was observed more than one time (green color), or if the line was a released variety used as a check in some trials (orange color).	92
14. Figure 3.2. Five-fold cross validation results for seven different genomic selection models evaluated by their ability to predict sawfly resistance BLUEs. The model tested is on the X-axis and the mean prediction accuracy ($n=100$) $\pm$ one standard deviation is shown on the Y-axis. Results are shown for models fit using the full phenotypic dataset and using the phenotypic dataset without the included F6 trials.	99
15. Figure 3.3. Pairwise significant marker interactions (Bonferroni corrected p -value < 0.05) that were detected in the full dataset (top panel) and in the F6 removed dataset (bottom panel). Each line represents one significant marker-marker interaction with the color of the line denoting the relative size of the p -value associated with that marker test.....	101
16. Figure 3.4. MLM GWAS results for A) models fit using the full dataset and B) models fit with the F6 trials removed. In both panels, the X-axis denotes the physical position (bp) a marker as well as the chromosome on which it sits. The Y-axis denotes the $-\log_{10}$ of the p -value associated with a given marker. The dashed horizontal line indicates the Bonferroni significance cutoff ($p < 0.05$).	102

ABSTRACT

Since the development of next generation sequencing (NGS) technology, there has been an explosion in the diversity and availability of large-scale genetic data. In response, the field of plant breeding has shifted to adopt these new technologies and incorporate them into breeding programs. Specifically, the development of high-density genetic maps and crop-specific reference genomes has led to an increased capacity to obtain genome-wide marker information on many individuals. Two examples of the application of these resources are fine mapping approaches for gene discovery and assembling genomic selection models for earlier generation trait improvement. This dissertation seeks to apply both of these procedures to the Montana State University Spring Wheat Breeding Program for a multi-faceted approach to accelerate germplasm improvement. In Chapter 1, fine mapping was applied to a quantitative trait locus controlling productive tiller number, a major yield component in wheat which also contributes to phenotypic plasticity. In Chapters 2 and 3, genomic selection models were developed to predict twelve end-use quality traits (Chapter 2) and wheat stem sawfly resistance (Chapter 3). Together, these projects laid the foundation for a comprehensive adoption of NGS-based plant breeding methods into the MSU Spring Wheat Breeding Program. Adopting these methods will allow for increased genetic control and improved resource efficiency for the improvement of primary target traits, ensuring the development of the best possible wheat varieties for Montana growers.

CHAPTER ONE

INTRODUCTION

Following the release of the first completed draft genome by the human genome project (Lander et al. 2001) which relied on the relatively low throughput Sanger sequencing method, took multiple years to complete, and cost around US \$300 million dollars for the sequencing alone; it was clear that new technologies were needed to advance into the era of affordable and time-efficient whole genome sequencing (Shendure et al. 2004; Reis-Filho 2009). This need was met with the development of next generation sequencing (NGS) technology which allows for massively parallelized sequencing that produces hundreds of millions to billions of reads in one run for a much lower cost (Margulies et al. 2005; Shendure et al. 2005; Shendure and Ji 2008). This technological breakthrough revolutionized the landscape of scientific research by allowing for affordable and time-efficient whole genome and whole transcriptome projects (Dijk et al. 2014).

The development of NGS technology has likewise had a major effect on plant breeding, being the catalyst for entry into an entire new era of breeding approaches (Wallace et al. 2018; Sun et al. 2024). Specifically, NGS has led to the development of crop-specific reference genomes (Schnable et al. 2009; Kawahara et al. 2013; The International Wheat Genome Sequencing Consortium (IWGSC) et al. 2018), whole transcriptome RNA sequencing studies (Wang et al. 2009; Zhai et al. 2013; Upton et al. 2023), large scale gene mapping projects (He et al. 2017; Wang et al. 2023), and whole genome prediction models for rapid cycling and improved selection accuracy (Meuwissen et al. 2001; Heffner et al. 2010; Jannink et al. 2010).

By combining these new resources and tools, plant breeders have sought to achieve high efficiency and rapid trait improvement in their programs (Sun et al. 2024).

Hexaploid bread wheat (*Triticum aestivum* L., $2n = 6x = 42$) is one of the primary food crops produced globally with an annual production of 798 million tonnes in 2024 (FAOSTAT, 2026). Compared to other major food crops like maize (*Zea mays* L.) and rice (*Oryza sativa* L.) which are both diploid species, wheat has a substantially larger (Messing et al. 2004) and more complex genome due to its high repeat content and numerous homoeologous regions which have made it difficult to sequence (Aury et al. 2022). As a result, the adoption of some of these NGS technologies in wheat has lagged behind maize and rice as evidenced by the first published reference genome of wheat being released in 2018, nine years after the first maize reference genome and thirteen years after the first rice reference genome.

Genetic improvement of wheat cultivars is a multifaceted task as numerous traits including grain yield, grain protein content, abiotic and biotic stress tolerance, and end-use quality must be considered (Tadesse et al. 2019), with grain yield being the most critical. Total grain yield is determined by three different yield components including seed size, seed number per spike, and productive tiller number (PTN) (Yoshida 1972; Ma et al. 2007; Lai et al. 2025). Each of these yield component traits is genetically complex and also interacts with numerous environmental factors (Gupta et al. 2006; Quintero et al. 2018; Gasparis and Miłoszewski 2023; Zhang et al. 2025). Further, there are complex resource tradeoffs between these components which themselves can be influenced by environmental conditions (Slafer et al. 2014; Quintero et al. 2018; Jian et al. 2024). Taken together, it is no surprise that grain yield can be a challenging trait to improve in a breeding program.

It has been estimated that average grain yields for wheat need to increase by 2% each year in order to meet the anticipated future global food demand (Zhuang et al. 2024b); however, many studies report genetic gains for grain yield below this threshold (Dixon et al. 2009; Crespo-Herrera et al. 2018; Tadesse et al. 2019). Of particular interest to the projects presented herein, it was recently determined that annual rate of genetic gain for grain yield in the hard red spring wheat market class in the Northern Plains region was 0.61%, far below the 2% threshold (Gill et al. 2025a). Next generation sequencing technologies may provide the tools needed to increase these genetic gains to keep up with growing demands. Specifically, efficient and affordable development of high density genetic maps allows for fine mapping of known quantitative trait loci (QTL) to uncover causal variants and exert a more precise and reliable control on the trait of interest (Zhao et al. 2022; Qiao et al. 2025). Second, the availability and repeatability of whole genome marker data allows for the development of genomic selection models which can increase selection accuracy and decrease the selection interval by allowing for earlier generation selections (Jannink et al. 2010; Alahmad et al. 2022).

Nearly 5.3 million acres of wheat were planted in the state of Montana in 2025, with an approximately equal distribution of winter wheat and spring wheat acres (USDA NASS 2026), with the primary market classes of both types being hard red wheat. Montana is a semiarid growing region with primarily dryland acre production with water as a primarily limited resource (Lenssen et al. 2007). Conventionally, grain yield is the primary breeding target of the Montana State University Spring Wheat Breeding program, and work has been done in the program to map and fine tune genes controlling the different yield components (Hale et al. 2024; Wong et al. 2026 (under review)).

Specifically, productive tiller number (PTN) has been a trait of interest to the breeding program as it contributes to yield plasticity (Peltonen-Sainio et al. 2011; Sadras and Rebetzke 2013; Tokatlidis 2014). Wheat plants can adjust tiller numbers in response to different population densities (Lloveras et al. 2004) and different environmental conditions such as drought (Slafer et al. 2014). Therefore, increased tillering potential could be a solution to uncertain weather conditions, especially with respect to rainfall (Tokatlidis 2014). Further, planting at a lower population density corresponds to an increase in PTN (Casal et al. 1986; Jones et al. 2021), and recently, Hale et al. (2025) found that increased tiller numbers in durum wheat grown in Montana could lead to increased yields under drought conditions and lower planting densities. Therefore, an investigation into the genetic basis of tiller number in wheat could benefit Montana wheat breeding.

There have been numerous genes identified in bread wheat which control tiller number including the *tiller inhibition (tin)* genes, *fertile tiller inhibition (ftin)*, and *Tiller Number1 (TN1)* (Peng et al. 1998; Spielmeier and Richards 2004; Kuraparthi et al. 2007; Zhang et al. 2013; Hyles et al. 2017; Wang et al. 2022; Si et al. 2022; Dong et al. 2023; Schoen et al. 2023; Hou et al. 2026). Additionally, there have been multiple tiller number genes found in wheat which are orthologous to tiller or branching genes in other major crop species. These genes include *TaTB1*, *TaOTUB1*, *TaSPL14*, *TaD27*, and *TaMOC1* (Zhang et al. 2015; Liu et al. 2017; Dixon et al. 2018; Zhao et al. 2020; Jiang et al. 2021; Cao et al. 2023; Maqbool et al. 2025). Additionally, there remain many tiller number QTL which have not yet been fine mapped (Shang et al. 2021; Liu et al. 2025), demonstrating both the genetic complexity of tiller number in bread wheat and the need for continued work in this area.

One important QTL for PTN in Montana spring wheat is *Q_{Tn.mst-6B}* which was initially mapped in by Naruoka et al. (2011) in a recombinant inbred line (RIL) population derived from a cross between two major Montan spring wheat varieties: Reeder (PI 613586)/Conan (PI 607549) (WestBred, LLC). Eight pairs of near-isogenic lines (NILs) were then generated following the heterogenous inbred family (HIF) procedure described in Blake et al. (2011) between a cross of the varieties Reeder (high PTN)/Choteau (PI 633974; low PTN) (Lanning et al. 2004). The effect of *Q_{Tn.mst-6B}* in these NILs was then validated in 2010, in which it was found that the high tiller allele increased PTN by 13% across all NILs (Naruoka et al. 2011).

Additional NILs for *Q_{Tn.mst-6B}* were then created by crosses between Reeder (high PTN)/Hank (WestBred LLC; low PTN), and Vida (PI 642366; high PTN) (Lanning et al. 2006a)/McNeal (PI 574642; low PTN) (Lanning et al. 1994) and combined with the previous 8 NIL pair for a total of 13 NIL pairs (Nasseer et al. 2016). These NILs were tested in two years (2012-2013) in a total of nine RCBD trials with five trials conducted across two Montana locations, and four trials in Washington state between two locations with eight out nine trials being under dryland conditions (Nasseer et al. 2016). There was a significant difference in PTN between the *Q_{Tn.mst-6B}* alleles across all NIL pairs when averaging over all nine of the experiments, though not within all of the environments (Nasseer et al. 2016). There was a significant difference in early tiller number (ETN)--tiller number collected at the first node stage (Zadoks 3.1 (Zadoks et al. 1974))--in all five Montana experiments in which it was collected (Nasseer et al. 2016). Across all nine experiments, PTN was increased between the QTL alleles, but seed weight was significantly decreased, and there was a slight reduction in the number of seeds per head, which, when taken together, resulted in no difference in total grain yield between

the *Q_{Tn.mst-6B}* alleles (Nasseer et al. 2016). Given that *Q_{Tn.mst-6B}* has been robustly validated in Montana-adapted germplasm in different genetic backgrounds across multiple environments, fine mapping this QTL could allow for more reliable and precise control (Lande and Thompson 1990) over tiller number in the MSU Spring Wheat Breeding Program.

In addition to yield component traits like productive tiller number, another set of major breeding targets for wheat are the end-use quality traits. End-use quality is a catch all term to describe the properties and composition of a grain sample and its derived flour (milling traits), dough (dough traits), and baked products (baking traits). The primary classes of wheat grown in Montana (hard red winter, hard red spring) are known for their high grain protein content, high gluten strength, and desirable end-use quality; these traits have been shown to influence the market price of hard red wheat (Roberts et al. 2022; Subedi et al. 2023). Further, Montana wheat is marketed for its quality (<https://montanawbc.com/buyers/#Quality>), making end-use quality traits important targets for the MSU Spring Wheat Program.

End-use quality traits can be difficult to improve as they are polygenic traits (Battenfield et al. 2018; Gill et al. 2025b) and require multiple laborious, complex tests performed on large grain samples in order to measure them. Additionally, end-use quality phenotyping is often delayed until the late-stage trials of a breeding program (Battenfield et al. 2016). End-use quality traits are typically quite heritable (Souza et al. 2012; Laidig et al. 2017), and despite being polygenic, these traits are also influenced by major genes such as those representing the glutenin and gliadin classes of storage proteins (Peña et al. 2002; Li et al. 2020). Given these factors, genomic selection could be a highly productive tool for improving end-use quality.

Genomic selection (GS) was first implemented by Meuwissen et al. (2001) and represents a whole-genome prediction of trait performance that incorporates both large effect and small effect loci and has been shown to have superior performance to marker assisted selection (MAS) because of this fact (Lande and Thompson 1990; Bernardo and Yu 2007a; Jannink et al. 2010). Genomic selection involves using a training population of individuals which have been phenotyped for the trait of interest and genotyped with a set of genome-wide markers to predict breeding values for genotyped, but unphenotyped genetically related individuals (de los Campos et al. 2013; Isik et al. 2017; Bernardo 2020). Prediction accuracy in GS models describes the relationship between the predicted and true breeding values and is typically measured by correlation. Prediction accuracy is controlled by following factors: the size of the training population, the heritability of the trait, the number and distribution of genetic markers used, the number of QTL controlling the trait, and the relatedness between the training and predicting populations (Daetwyler et al. 2010; Bernardo 2016; Zhang et al. 2017). Therefore, an ideal GS model will have a large training population that is closely related to the prediction population of interest where both populations have been genotyped with a dense, genome-wide set of markers, and the traits of interest are consistently heritable. The MSU Spring Wheat Program naturally collects end-use quality data each year on the elite breeding lines present in the late-stage yield trials. Further, these lines are often used as crossing parents for the next generation of lines in the program. Therefore, NGS technology could be utilized to obtain dense genetic marker data on both past and up-and-coming breeding lines in the program. When combined with existing phenotypic quality data, the heritability of the end-use quality traits, and the known relatedness between historic and current breeding lines, it could be possible to implement GS for earlier

generation selection of end-use quality in the breeding program (Battenfield et al. 2016; Zhang-Biehn et al. 2021; Gill et al. 2023).

Genomic selection can also be implemented for traits related to biotic stresses. The most important insect pest of Montana wheat is the wheat stem sawfly (*Cephus cinctus* Norton) (hereafter sawfly) which has been estimated to cause an annual wheat yield loss of US \$400 million across the northern and central plains growing regions in North America (Beres et al. 2017). The historic regions of sawfly infestation, dating as far back as the early 1900s, include Montana, North and South Dakota, western Minnesota, and south-central Canada (Beres et al. 2011), though recently, the range of sawfly infestations has spread to the state of Washington, northern Idaho, Wyoming, Colorado, Nebraska, and Kansas (Beres et al. 2017; Weaver 2023). Yield loss due to sawfly is the result of the larvae feeding within the stems of the wheat plants and then cutting the stems in the late season before harvest resulting in severe lodging (Talbert et al. 2014). Resistance to the sawfly is another major breeding target of the MSU Spring Wheat Breeding Program, though it can be challenging to screen candidate lines for sawfly resistance due to the nongenetic, spatial variation of sawfly pressure within a field (Nansen et al. 2005) and the substantial inconsistencies in sawfly pressure year to year due to the influence of environmental conditions (Olfert et al. 2019; Ermatinger et al. 2025).

The primary mechanism of sawfly resistance is through the solid stem trait which describes the filling of pith within the wheat stem which creates a mechanical barrier to restrict larvae movement and potentially result in larvae death (Hayat et al. 1995; Hager et al. 2025). However, stem solidness does not entirely prevent stem cutting, with solid-stemmed cultivars still experiencing up to 30% stem cutting (Talbert et al. 2014; Wong et al. 2023). In addition, the

expression of stem solidness varies between environments and is affected by environmental conditions such as nitrogen availability and shade (Holmes 1984; Cook et al. 2004; Beres et al. 2012; Bathini et al. 2023). Further, stem solidness is a polygenic trait controlled by a major QTL *Qss.msub-3BL* (Cook et al. 2004; Nilsen et al. 2020) as well as many other minor effect QTL which may have epistatic interactions between them (McKenzie 1965; Lanning et al. 2006b; Sherman et al. 2010; Varella et al. 2015, 2019; Nilsen et al. 2017; Dhariwal et al. 2022).

Given all these features, resistance to the wheat stem sawfly could be another good candidate trait for genomic selection. As with end-use quality, the MSU Spring Wheat Breeding Program has historically collected sawfly resistance data as part of the late-stage trials. Further, the polygenic nature and complex genetic interactions associated with stem solidness could be well predicted by GS models which incorporate epistasis (McKinney et al. 2006; Howard et al. 2014; Jiang and Reif 2015), whereas a marker assisted selection approach may be insufficient (Lande and Thompson 1990; Jannink et al. 2010). Genomic selection models could both increase selection accuracy in replicated trials which did not get consistent or strong sawfly pressure, and they could allow for earlier generation selection for sawfly resistance at stages where phenotyping would not be possible (Isik et al. 2017).

The following chapters describe three research projects which sought to leverage NGS technologies to benefit the MSU Spring Wheat Breeding Program. The first project was to fine map the *QTn.mst-6B* for productive tiller number by utilizing genetic markers derived from an exome capture assay (Krasileva et al. 2017) and by leveraging RNA sequencing to identify differentially expressed genes within the QTL. The second project is to develop a genomic selection model for twelve primary end-use quality traits by leveraging historic quality data

collected by the breeding program and genome-wide genetic markers. The final project is to develop a genomic selection model and perform genome wide association mapping for wheat stem sawfly resistance in the breeding program. Together, these projects seek to improve the efficiency and reliability of improvement for many of the most important target traits in Montana Spring Wheat.

References

- Alahmad S, Rambla C, Voss-Fels KP, Hickey LT (2022) Accelerating Breeding Cycles. In: Reynolds MP, Braun H-J (eds) *Wheat Improvement: Food Security in a Changing Climate*. Springer International Publishing, Cham, pp 557–571
- Aury J-M, Engelen S, Istace B, et al (2022) Long-read and chromosome-scale assembly of the hexaploid wheat genome achieves high resolution for research and breeding. *Gigascience* 11:giac034. <https://doi.org/10.1093/gigascience/giac034>
- Bathini A, Mendu L, Singh NP, et al (2023) A “solid” solution for wheat stem sawfly (Hymenoptera: Cephidae) resistance: Genetics, breeding and development of solid stem wheat. *Crop Science* 63:1934–1948. <https://doi.org/10.1002/csc2.20998>
- Battenfield SD, Guzmán C, Gaynor RC, et al (2016) Genomic Selection for Processing and End-Use Quality Traits in the CIMMYT Spring Bread Wheat Breeding Program. *The Plant Genome* 9:plantgenome2016.01.0005. <https://doi.org/10.3835/plantgenome2016.01.0005>
- Battenfield SD, Sheridan JL, Silva LDCE, et al (2018) Breeding-assisted genomics: Applying meta-GWAS for milling and baking quality in CIMMYT wheat breeding program. *PLOS ONE* 13:e0204757. <https://doi.org/10.1371/journal.pone.0204757>
- Beres BL, Dosdall LM, Weaver DK, et al (2011) Biology and integrated management of wheat stem sawfly and the need for continuing research. *The Canadian Entomologist* 143:105–125. <https://doi.org/10.4039/n10-056>
- Beres BL, Hill BD, Cárcamo HA, et al (2017) An artificial neural network model to predict wheat stem sawfly cutting in solid-stemmed wheat cultivars. *Can J Plant Sci* 97:329–336. <https://doi.org/10.1139/cjps-2016-0364>
- Beres BL, McKenzie RH, Cárcamo HA, et al (2012) Influence of Seeding Rate, Nitrogen Management, and Micronutrient Blend Applications on Pith Expression in Solid-Stemmed Spring Wheat. *Crop Science* 52:1316–1329. <https://doi.org/10.2135/cropsci2011.05.0239>

- Bernardo R (2020) *Breeding for Quantitative Traits in Plants*, Third. Stemma Press
- Bernardo R (2016) Bandwagons I, too, have known. *Theor Appl Genet* 129:2323–2332. <https://doi.org/10.1007/s00122-016-2772-5>
- Bernardo R, Yu J (2007) Prospects for Genomewide Selection for Quantitative Traits in Maize. *Crop Science* 47:1082–1090. <https://doi.org/10.2135/cropsci2006.11.0690>
- Blake NK, Stougaard RN, Weaver DK, et al (2011) Identification of a quantitative trait locus for resistance to *Sitodiplosis mosellana* (Géhin), the orange wheat blossom midge, in spring wheat. *Plant Breeding* 130:25–30. <https://doi.org/10.1111/j.1439-0523.2010.01809.x>
- Cao L, Li T, Geng S, et al (2023) TaSPL14-7A is a conserved regulator controlling plant architecture and yield traits in common wheat (*Triticum aestivum* L.). *Front Plant Sci* 14:. <https://doi.org/10.3389/fpls.2023.1178624>
- Casal JJ, Sanchez RA, Deregibus VA (1986) The effect of plant density on tillering: The involvement of R/FR ratio and the proportion of radiation intercepted per plant. *Environmental and Experimental Botany* 26:365–371. [https://doi.org/10.1016/0098-8472\(86\)90024-9](https://doi.org/10.1016/0098-8472(86)90024-9)
- Cook JP, Wichman DM, Martin JM, et al (2004) Identification of Microsatellite Markers Associated with a Stem Solidness Locus in Wheat. *Crop Science* 44:1397–1402. <https://doi.org/10.2135/cropsci2004.1397>
- Crespo-Herrera LA, Crossa J, Huerta-Espino J, et al (2018) Genetic Gains for Grain Yield in CIMMYT's Semi-Arid Wheat Yield Trials Grown in Suboptimal Environments. *Crop Science* 58:1890–1898. <https://doi.org/10.2135/cropsci2018.01.0017>
- Daetwyler HD, Pong-Wong R, Villanueva B, Woolliams JA (2010) The Impact of Genetic Architecture on Genome-Wide Evaluation Methods. *Genetics* 185:1021–1031. <https://doi.org/10.1534/genetics.110.116855>
- de los Campos G, Hickey JM, Pong-Wong R, et al (2013) Whole-Genome Regression and Prediction Methods Applied to Plant and Animal Breeding. *Genetics* 193:327–345. <https://doi.org/10.1534/genetics.112.143313>
- Dhariwal R, Hiebert CW, Randhawa HS (2022) QTL analysis identified two major all-internodes solidness loci from a completely solid-stemmed spring wheat line. *Front Plant Sci* 13:. <https://doi.org/10.3389/fpls.2022.1035620>
- Dijk EL van, Auger H, Jaszczyszyn Y, Thermes C (2014) Ten years of next-generation sequencing technology. *Trends in Genetics* 30:418–426. <https://doi.org/10.1016/j.tig.2014.07.001>
- Dixon J, Braun HJ, Kosina P, Crouch JH (2009) *Wheat Facts and Futures 2009*. CIMMYT

- Dixon LE, Greenwood JR, Bencivenga S, et al (2018) TEOSINTE BRANCHED1 Regulates Inflorescence Architecture and Development in Bread Wheat (*Triticum aestivum*). *Plant Cell* 30:563–581. <https://doi.org/10.1105/tpc.17.00961>
- Dong C, Zhang L, Zhang Q, et al (2023) Tiller Number1 encodes an ankyrin repeat protein that controls tillering in bread wheat. *Nat Commun* 14:836. <https://doi.org/10.1038/s41467-023-36271-z>
- Ermatinger LS, Powell SL, Peterson RKD, Weaver DK (2025) Mapping Wheat Stem Sawfly (*Cephus cinctus* Norton) Infestations in Spring and Winter Wheat Fields via Multiway Modelling of Multitemporal Sentinel 2 Images. *Remote Sensing* 17:1950. <https://doi.org/10.3390/rs17111950>
- Gasparis S, Miłoszewski MM (2023) Genetic Basis of Grain Size and Weight in Rice, Wheat, and Barley. *International Journal of Molecular Sciences* 24:16921. <https://doi.org/10.3390/ijms242316921>
- Gill HS, Blecha S, Brault C, et al (2025a) Genetic gains from 60 years of spring wheat breeding in the Northern Plains of the United States. *Crop Science* 65:e70106. <https://doi.org/10.1002/csc2.70106>
- Gill HS, Conley E, Brault C, et al (2025b) Association mapping and genomic prediction for processing and end-use quality traits in wheat (*Triticum aestivum* L.). *The Plant Genome* 18:e20529. <https://doi.org/10.1002/tpg2.20529>
- Gupta PK, Rustgi S, Kumar N (2006) Genetic and molecular basis of grain size and grain number and its relevance to grain productivity in higher plants. *Genome* 49:565–571. <https://doi.org/10.1139/g06-063>
- Hager MS, Cook JP, Bothner B, Weaver DK (2025) Untargeted metabolomic and transcriptomic analysis in spring and durum wheat reveals potential mechanisms associated with the early stem solidness phenotype and resistance to wheat stem sawfly. *Front Plant Sci* 16:. <https://doi.org/10.3389/fpls.2025.1497732>
- Hale CO, Tillett BJ, Martin JM, et al (2024) The Grain Number Increase 1 alleles GNI-A1-105Y and -105K increase grain number in spring wheat. *Crop Science* 64:2212–2222. <https://doi.org/10.1002/csc2.21267>
- Hale CO, Volkman MM, Martin JM, et al (2025) A Teosinte Branched-B1 null mutation increases durum wheat tillering, increasing grain yield in certain environments. *Crop Science* 65:e70136. <https://doi.org/10.1002/csc2.70136>
- Hayat MA, Martin JM, Lanning SP, et al (1995) Variation for stem solidness and its association with agronomic traits in spring wheat. *Can J Plant Sci* 75:775–780. <https://doi.org/10.4141/cjps95-131>

- He J, Meng S, Zhao T, et al (2017) An innovative procedure of genome-wide association analysis fits studies on germplasm population and plant breeding. *Theor Appl Genet* 130:2327–2343. <https://doi.org/10.1007/s00122-017-2962-9>
- Heffner EL, Lorenz AJ, Jannink J-L, Sorrells ME (2010) Plant Breeding with Genomic Selection: Gain per Unit Time and Cost. *Crop Science* 50:1681–1690. <https://doi.org/10.2135/cropsci2009.11.0662>
- Holmes ND (1984) THE EFFECT OF LIGHT ON THE RESISTANCE OF HARD RED SPRING WHEATS TO THE WHEAT STEM SAWFLY, *CEPHUS CINCTUS* (HYMENOPTERA: CEPHIDAE). *The Canadian Entomologist* 116:677–684. <https://doi.org/10.4039/Ent116677-5>
- Hou S, Yu L, Guo W, et al (2026) Fine mapping and verification of a tiller inhibition gene in synthetic hexaploid wheat derivative. *Mol Breeding* 46:17. <https://doi.org/10.1007/s11032-026-01640-1>
- Howard R, Carriquiry AL, Beavis WD (2014) Parametric and Nonparametric Statistical Methods for Genomic Selection of Traits with Additive and Epistatic Genetic Architectures. *G3 Genes|Genomes|Genetics* 4:1027–1046. <https://doi.org/10.1534/g3.114.010298>
- Hyles J, Vautrin S, Pettolino F, et al (2017) Repeat-length variation in a wheat cellulose synthase-like gene is associated with altered tiller number and stem cell wall composition. *J Exp Bot* 68:1519–1529. <https://doi.org/10.1093/jxb/erx051>
- Isik F, Holland J, Maltecca C (2017) Genomic Selection. In: Isik F, Holland J, Maltecca C (eds) *Genetic Data Analysis for Plant and Animal Breeding*. Springer International Publishing, Cham, pp 355–384
- Jannink J-L, Lorenz AJ, Iwata H (2010) Genomic selection in plant breeding: from theory to practice. *Brief Funct Genomics* 9:166–177. <https://doi.org/10.1093/bfpg/elq001>
- Jian C, Pan Y, Liu S, et al (2024) The TaGW2-TaSPL14 module regulates the trade-off between tiller number and grain weight in wheat. *Journal of Integrative Plant Biology* 66:1953–1965. <https://doi.org/10.1111/jipb.13723>
- Jiang DJ, Hao XH, Zhang J, et al (2021) Reducing expression of TaOTUB1s decreases tiller number in wheat. *Plant Signaling & Behavior* 16:2018217. <https://doi.org/10.1080/15592324.2021.2018217>
- Jiang Y, Reif JC (2015) Modeling Epistasis in Genomic Selection. *Genetics* 201:759–768. <https://doi.org/10.1534/genetics.115.177907>
- Jones BH, Blake NK, Heo H-Y, et al (2021) Allelic response of yield component traits to resource availability in spring wheat. *Theor Appl Genet* 134:603–620. <https://doi.org/10.1007/s00122-020-03717-7>

- Kawahara Y, de la Bastide M, Hamilton JP, et al (2013) Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. *Rice* 6:4. <https://doi.org/10.1186/1939-8433-6-4>
- Krasileva KV, Vasquez-Gross HA, Howell T, et al (2017) Uncovering hidden variation in polyploid wheat. *Proceedings of the National Academy of Sciences* 114:E913–E921. <https://doi.org/10.1073/pnas.1619268114>
- Kuraparthi V, Sood S, Dhaliwal HS, et al (2007) Identification and mapping of a tiller inhibition gene (*tin3*) in wheat. *Theor Appl Genet* 114:285–294. <https://doi.org/10.1007/s00122-006-0431-y>
- Lai X, Zhu Z, Zhang Y, et al (2025) The selection and application of tiller number QTLs in modern wheat breeding. *Theor Appl Genet* 138:124. <https://doi.org/10.1007/s00122-025-04908-w>
- Laidig F, Piepho H-P, Rentel D, et al (2017) Breeding progress, environmental variation and correlation of winter wheat yield and quality traits in German official variety trials and on-farm during 1983–2014. *Theor Appl Genet* 130:223–245. <https://doi.org/10.1007/s00122-016-2810-3>
- Lande R, Thompson R (1990) Efficiency of Marker-Assisted Selection in the Improvement of Quantitative Traits. *Genetics* 124:743–756. <https://doi.org/10.1093/genetics/124.3.743>
- Lander ES, Linton LM, Birren B, et al (2001) Initial sequencing and analysis of the human genome. *Nature* 409:860–921. <https://doi.org/10.1038/35057062>
- Lanning S p., Carlson G r., Nash D, et al (2004) Registration of ‘Choteau’ Wheat. *Crop Science* 44:2264–2265. <https://doi.org/10.2135/cropsci2004.2264>
- Lanning SP, Carlson GR, Nash D, et al (2006a) Registration of “Vida” Wheat. *Crop Science* 46:2315–2316
- Lanning SP, Fox P, Elser J, et al (2006b) Microsatellite Markers Associated with a Secondary Stem Solidness Locus in Wheat. *Crop Science* 46:1701–1703. <https://doi.org/10.2135/cropsci2005.10-0379>
- Lanning SP, Talbert LE, McGuire CF, et al (1994) Registration of ‘McNeal’ Wheat. *Crop Science* 34:cropsci1994.0011183X003400040060x. <https://doi.org/10.2135/cropsci1994.0011183X003400040060x>
- Lenzen AW, Johnson GD, Carlson GR (2007) Cropping sequence and tillage system influences annual crop production and water use in semiarid Montana, USA. *Field Crops Research* 100:32–43. <https://doi.org/10.1016/j.fcr.2006.05.004>

- Li X, Li Y, Yu X, et al (2020) Genomics-Enabled Analysis of Puroindoline b2 Genes Identifies New Alleles in Wheat and Related Triticeae Species. *International Journal of Molecular Sciences* 21:. <https://doi.org/10.3390/ijms21041304>
- Liu J, Cheng X, Liu P, Sun J (2017) miR156-Targeted SBP-Box Transcription Factors Interact with DWARF53 to Regulate TEOSINTE BRANCHED1 and BARREN STALK1 Expression in Bread Wheat. *Plant Physiol* 174:1931–1948. <https://doi.org/10.1104/pp.17.00445>
- Liu J, Wang T, Lan Y, et al (2025) Fine-mapping and candidate gene identification for *QPtn.sau-4B* showing potential in increasing productive tiller number and yield in wheat. *The Crop Journal* 13:480–489. <https://doi.org/10.1016/j.cj.2025.01.014>
- Lloveras J, Manent J, Viudas J, et al (2004) Seeding Rate Influence on Yield and Yield Components of Irrigated Winter Wheat in a Mediterranean Climate. *Agronomy Journal* 96:1258–1265. <https://doi.org/10.2134/agronj2004.1258>
- Ma Z, Zhao D, Zhang C, et al (2007) Molecular genetic analysis of five spike-related traits in wheat using RIL and immortalized F2 populations. *Mol Genet Genomics* 277:31–42. <https://doi.org/10.1007/s00438-006-0166-0>
- Maqbool R, Nagarajan R, Mutti JS, Gill KS (2025) Wheat Tiller -1 (WT-1), a GRAS domain encoding gene, controls both tillering and spikelet number per spike in wheat. *Plant Mol Biol* 115:128. <https://doi.org/10.1007/s11103-025-01629-y>
- Margulies M, Egholm M, Altman WE, et al (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437:376–380. <https://doi.org/10.1038/nature03959>
- McKenzie H (1965) INHERITANCE OF SAWFLY REACTION AND STEM SOLIDNESS IN SPRING WHEAT CROSSES: STEM SOLIDNESS. *Can J Plant Sci* 45:591–600. <https://doi.org/10.4141/cjps65-111>
- McKinney BA, Reif DM, Ritchie MD, Moore JH (2006) Machine Learning for Detecting Gene-Gene Interactions. *Appl Bioinformatics* 5:77–88. <https://doi.org/10.2165/00822942-200605020-00002>
- Messing J, Bharti AK, Karlowski WM, et al (2004) Sequence composition and genome organization of maize. *Proc Natl Acad Sci U S A* 101:14349–14354. <https://doi.org/10.1073/pnas.0406163101>
- Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. *Genetics* 157:1819–1829. <https://doi.org/10.1093/genetics/157.4.1819>

- Nansen C, Macedo TB, Weaver DK, Peterson RKD (2005) Spatiotemporal distributions of wheat stem sawfly eggs and larvae in dryland wheat fields. *The Canadian Entomologist* 137:428–440. <https://doi.org/10.4039/n04-094>
- Naruoka Y, Talbert LE, Lanning SP, et al (2011) Identification of quantitative trait loci for productive tiller number and its relationship to agronomic traits in spring wheat. *Theor Appl Genet* 123:1043. <https://doi.org/10.1007/s00122-011-1646-0>
- Nasseer A m., Martin J m., Heo H y., et al (2016) Impact of a Quantitative Trait Locus for Tiller Number on Plasticity of Agronomic Traits in Spring Wheat. *Crop Science* 56:595–602. <https://doi.org/10.2135/cropsci2015.05.0325>
- Nilsen KT, N'Diaye A, MacLachlan PR, et al (2017) High density mapping and haplotype analysis of the major stem-solidness locus SSt1 in durum and common wheat. *PLOS ONE* 12:e0175285. <https://doi.org/10.1371/journal.pone.0175285>
- Nilsen KT, Walkowiak S, Xiang D, et al (2020) Copy number variation of TdDof controls solid-stemmed architecture in wheat. *Proceedings of the National Academy of Sciences* 117:28708–28718. <https://doi.org/10.1073/pnas.2009418117>
- Olfert O, Weiss RM, Catton H, et al (2019) Bioclimatic assessment of abiotic factors affecting relative abundance and distribution of wheat stem sawfly (Hymenoptera: Cephidae) in western Canada. *The Canadian Entomologist* 151:16–33. <https://doi.org/10.4039/tce.2018.46>
- Peltonen-Sainio P, Jauhiainen L, Sadras VO (2011) Phenotypic plasticity of yield and agronomic traits in cereals and rapeseed at high latitudes. *Field Crops Research* 124:261–269. <https://doi.org/10.1016/j.fcr.2011.06.016>
- Peña RJ, Trethowan R, Pfeiffer WH, Ginkel MV (2002) Quality (End-Use) Improvement in Wheat: Compositional, Genetic, and Environmental Factors. *Journal of Crop Production* 5:1–37. https://doi.org/10.1300/J144v05n01_02
- Peng ZS, Yen C, Yang JL (1998) Genetic control of oligo-culms in common wheat [*Triticum aestivum*]. Wheat Information Service (Japan)
- Qiao L, Zheng X, Zhao J, et al (2025) Genetic dissection of flag leaf morphology traits and fine mapping of a novel QTL (Qflw.sxau-6BL) in bread wheat (*Triticum aestivum* L.). *Theor Appl Genet* 138:21. <https://doi.org/10.1007/s00122-024-04802-x>
- Quintero A, Molero G, Reynolds MP, Calderini DF (2018) Trade-off between grain weight and grain number in wheat depends on GxE interaction: A case study of an elite CIMMYT panel (CIMCOG). *European Journal of Agronomy* 92:17–29. <https://doi.org/10.1016/j.eja.2017.09.007>

- Reis-Filho JS (2009) Next-generation sequencing. *Breast Cancer Res* 11:S12. <https://doi.org/10.1186/bcr2431>
- Roberts S, Brooks K, Nogueira L, Walters CG (2022) The role of quality characteristics in pricing hard red winter wheat. *Food Policy* 108:102246. <https://doi.org/10.1016/j.foodpol.2022.102246>
- Sadras VO, Rebetzke GJ (2013) Plasticity of wheat grain yield is associated with plasticity of ear number. *cpsc* 64:234–243. <https://doi.org/10.1071/CP13117>
- Schnable PS, Ware D, Fulton RS, et al (2009) The B73 Maize Genome: Complexity, Diversity, and Dynamics. *Science* 326:1112–1115. <https://doi.org/10.1126/science.1178534>
- Schoen A, Yadav I, Wu S, et al (2023) Identification and high-resolution mapping of a novel tiller number gene (*tin6*) by combining forward genetics screen and MutMap approach in bread wheat. *Funct Integr Genomics* 23:157. <https://doi.org/10.1007/s10142-023-01084-2>
- Shang Q, Wang Y, Tang H, et al (2021) Genetic, hormonal, and environmental control of tillering in wheat. *The Crop Journal* 9:986–991. <https://doi.org/10.1016/j.cj.2021.03.002>
- Shendure J, Ji H (2008) Next-generation DNA sequencing. *Nat Biotechnol* 26:1135–1145. <https://doi.org/10.1038/nbt1486>
- Shendure J, Mitra RD, Varma C, Church GM (2004) Advanced sequencing technologies: methods and goals. *Nat Rev Genet* 5:335–344. <https://doi.org/10.1038/nrg1325>
- Shendure J, Porreca GJ, Reppas NB, et al (2005) Accurate Multiplex Polony Sequencing of an Evolved Bacterial Genome. *Science* 309:1728–1732. <https://doi.org/10.1126/science.1117389>
- Sherman JD, Weaver DK, Hofland ML, et al (2010) Identification of Novel QTL for Sawfly Resistance in Wheat. *Crop Science* 50:73–86. <https://doi.org/10.2135/cropsci2009.03.0145>
- Si Y, Lu Q, Tian S, et al (2022) Fine mapping of the tiller inhibition gene *TIN5* in *Triticum urartu*. *Theor Appl Genet* 135:2665–2673. <https://doi.org/10.1007/s00122-022-04140-w>
- Slafer GA, Savin R, Sadras VO (2014) Coarse and fine regulation of wheat yield components in response to genotype and environment. *Field Crops Research* 157:71–83. <https://doi.org/10.1016/j.fcr.2013.12.004>
- Souza EJ, Sneller C, Guttieri MJ, et al (2012) Basis for Selecting Soft Wheat for End-Use Quality. *Crop Science* 52:21–31. <https://doi.org/10.2135/cropsci2011.02.0090>

- Spielmeyer W, Richards RA (2004) Comparative mapping of wheat chromosome 1AS which contains the tiller inhibition gene (*tin*) with rice chromosome 5S. *Theor Appl Genet* 109:1303–1310. <https://doi.org/10.1007/s00122-004-1745-2>
- Subedi M, Ghimire B, Bagwell JW, et al (2023) Wheat end-use quality: State of art, genetics, genomics-assisted improvement, future challenges, and opportunities. *Front Genet* 13:. <https://doi.org/10.3389/fgene.2022.1032601>
- Sun L, Lai M, Ghouri F, et al (2024) Modern Plant Breeding Techniques in Crop Improvement and Genetic Diversity: From Molecular Markers and Gene Editing to Artificial Intelligence—A Critical Review. *Plants* 13:2676. <https://doi.org/10.3390/plants13192676>
- Tadesse W, Sanchez-Garcia M, Assefa SG, et al (2019) Genetic Gains in Wheat Breeding and Its Role in Feeding the World. *Crop Breeding, Genetics and Genomics* 1:. <https://doi.org/10.20900/cbgg20190005>
- Talbert LE, Sherman JD, Hofland ML, et al (2014) Resistance to *Cephus cinctus* Norton, the wheat stem sawfly, in a recombinant inbred line population of wheat derived from two resistance sources. *Plant Breeding* 133:427–432. <https://doi.org/10.1111/pbr.12184>
- The International Wheat Genome Sequencing Consortium (IWGSC), Appels R, Eversole K, et al (2018) Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* 361:eaar7191. <https://doi.org/10.1126/science.aar7191>
- Tokatlidis I s. (2014) Addressing the yield by density interaction is a prerequisite to bridge the yield gap of rain-fed wheat. *Annals of Applied Biology* 165:27–42. <https://doi.org/10.1111/aab.12121>
- Upton RN, Correr FH, Lile J, et al (2023) Design, execution, and interpretation of plant RNA-seq analyses. *Front Plant Sci* 14:. <https://doi.org/10.3389/fpls.2023.1135455>
- USDA NASS (2026) USDA/NASS 2025 State Agriculture Overview for Montana. https://www.nass.usda.gov/Quick_Stats/Ag_Overview/stateOverview.php?state=MONTANA. Accessed 19 Apr 2026
- Varella AC, Weaver DK, Sherman JD, et al (2015) Association Analysis of Stem Solidness and Wheat Stem Sawfly Resistance in a Panel of North American Spring Wheat Germplasm. *Crop Science* 55:2046–2055. <https://doi.org/10.2135/cropsci2014.12.0852>
- Varella AC, Zhang H, Weaver DK, et al (2019) A Novel QTL in Durum Wheat for Resistance to the Wheat Stem Sawfly Associated with Early Expression of Stem Solidness. *G3 Genes|Genomes|Genetics* 9:1999–2006. <https://doi.org/10.1534/g3.119.400240>
- Wallace JG, Rodgers-Melnick E, Buckler ES (2018) On the Road to Breeding 4.0: Unraveling the Good, the Bad, and the Boring of Crop Quantitative Genomics. *Annual Review of Genetics* 52:421–444. <https://doi.org/10.1146/annurev-genet-120116-024846>

- Wang W, Guo W, Le L, et al (2023) Integration of high-throughput phenotyping, GWAS, and predictive models reveals the genetic architecture of plant height in maize. *Molecular Plant* 16:354–373. <https://doi.org/10.1016/j.molp.2022.11.016>
- Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* 10:57–63. <https://doi.org/10.1038/nrg2484>
- Wang Z, Wu F, Chen X, et al (2022) Fine mapping of the tiller inhibition gene TIN4 contributing to ideal plant architecture in common wheat. *Theor Appl Genet* 135:527–535. <https://doi.org/10.1007/s00122-021-03981-1>
- Weaver D (2023) Wheat stem sawfly (*Cephus cinctus* Norton). In: Eigenbrode S, Rashed A (eds) *Burleigh Dodds Series in Agricultural Science*. Burleigh Dodds Science Publishing, pp 93–134
- Wong ML, Bruckner PL, Berg JE, et al (2023) Evaluation of wheat stem sawfly-resistant solid stem Qss.msub-3BL alleles in hard red winter wheat. *Crop Science* 63:556–567. <https://doi.org/10.1002/csc2.20866>
- Yoshida S (1972) Physiological Aspects of Grain Yield. *Annual Review of Plant Biology* 23:437–464. <https://doi.org/10.1146/annurev.pp.23.060172.002253>
- Zadoks JC, Chang TT, Konzak CF (1974) A decimal code for the growth stages of cereals. *Weed Research* 14:415–421. <https://doi.org/10.1111/j.1365-3180.1974.tb01084.x>
- Zhai R, Feng Y, Wang H, et al (2013) Transcriptome analysis of rice root heterosis by RNA-Seq. *BMC Genomics* 14:19. <https://doi.org/10.1186/1471-2164-14-19>
- Zhang A, Wang H, Beyene Y, et al (2017) Effect of Trait Heritability, Training Population Size and Marker Density on Genomic Prediction Accuracy Estimation in 22 bi-parental Tropical Maize Populations. *Front Plant Sci* 8:. <https://doi.org/10.3389/fpls.2017.01916>
- Zhang B, Liu X, Xu W, et al (2015) Novel function of a putative MOC1 ortholog associated with spikelet number per spike in common wheat. *Sci Rep* 5:12211. <https://doi.org/10.1038/srep12211>
- Zhang J, Wu J, Liu W, et al (2013) Genetic mapping of a fertile tiller inhibition gene, *ftin*, in wheat. *Mol Breeding* 31:441–449. <https://doi.org/10.1007/s11032-012-9801-0>
- Zhang N, Liu Y, Gui S, Wang Y (2025) Regulation of tillering and panicle branching in rice and wheat. *Journal of Genetics and Genomics* 52:869–886. <https://doi.org/10.1016/j.jgg.2024.12.005>
- Zhao B, Wu TT, Ma SS, et al (2020) TaD27-B gene controls the tiller number in hexaploid wheat. *Plant Biotechnology Journal* 18:513–525. <https://doi.org/10.1111/pbi.13220>

- Zhao C, Liu X, Liu H, et al (2022) Fine mapping of QFlw-5B, a major QTL for flag leaf width in common wheat (*Triticum aestivum* L.). *Theor Appl Genet* 135:2531–2541.
<https://doi.org/10.1007/s00122-022-04135-7>
- Zhuang L, Liu H, Hou J, et al (2024) Genetic improvement of important agronomic traits in Chinese wheat breeding over the past 70 years. *BMC Plant Biol* 24:1151.
<https://doi.org/10.1186/s12870-024-05841-8>
- (2026) FAOSTAT. <https://www.fao.org/faostat/>. Accessed 19 Apr 2026

CHAPTER TWO

FINE MAPPING A QTL FOR PRODUCTIVE TILLER NUMBER

Introduction

Grain yield is the primary trait for improvement in spring wheat breeding. Grain yield is derived from three main components: productive tiller number (PTN), seed number per spike, and seed weight (Yoshida 1972; Ma et al. 2007; Lai et al. 2025). Productive tiller number specifically refers to the number of tillers produced by a wheat plant that yield fertile, grain-bearing, spikes. Wheat plants adjust tiller numbers in response to environmental conditions (Lloveras et al. 2004), which contributes to yield plasticity and helps maximize yield potential (Peltonen-Sainio et al. 2011; Sadras and Rebetzke 2013; Tokatlidis 2014; Nasseer et al. 2016). Further, there are complex interactions and trade-offs between PTN and other yield components (Slafer et al. 2014; Jian et al. 2024), meaning that precise control and tuning of PTN is necessary to achieve optimal yields.

Such control would be aided by the identification of major genes that control tiller number in wheat. Many contributing genes have been found, including seven *tiller inhibition* (*tin*) genes (Peng et al. 1998; Spielmeyer and Richards 2004; Kuraparthi et al. 2007; Hyles et al. 2017; Wang et al. 2022; Si et al. 2022; Schoen et al. 2023; Hou et al. 2026), *fertile tiller inhibition* (*ftin*) (Zhang et al. 2013), *Tiller Number1* (*TNI*) (Dong et al. 2023), and several orthologous genes from other crop species including *TaTBI* (Liu et al. 2017; Dixon et al. 2018), *TaOTUB1* (Jiang et al. 2021), *TaSPL14* (Cao et al. 2023), *TaD27* (Zhao et al. 2020), *TaMOCI* (Zhang et al. 2015; Maqbool et al. 2025) and others. Despite this significant progress, tiller

number is a highly complex trait (Zhang et al. 2025), and many important tiller number quantitative trait loci (QTLs) remain uncharacterized (Shang et al. 2021; Liu et al. 2025).

One PTN QTL of importance for Montana spring wheat germplasm is *QTn.mst-6B*. This QTL was first mapped by Naruoka et al. (2011) in recombinant inbred line (RIL) populations derived from crosses between major cultivars grown in Montana and from field trials conducted at the Arthur H. Post Research Farm in Bozeman, Montana. This QTL was then validated for its effect on both early tiller number (ETN) and PTN in nine field trials conducted across both Montana and Washington state (Nasseer et al. 2016). Given the consistency of this QTL across multiple studies and environments, and that currently, no major tiller number genes have been mapped to chromosome group 6, we sought to fine map *QTn.mst-6B* to identify the causal variant(s) therein. We found that the ability to detect differences in tiller number was highly dependent on environmental variables, especially supplemental lighting conditions. Further, we conducted RNA-seq analysis to interrogate genes within our QTL region as well as candidate orthologs from other crop species. Finally, we identify the most probable candidate genes based on all the collected evidence, and we describe the best path forward for further investigations and validations of this tillering locus.

Materials and Methods

Generating Plant Materials

The tiller number QTL, *QTn.mst-6B*, was first mapped in Naruoka et al. (2011). As part of that process, near isogenic lines (NILs) were developed from crosses between several pairs of released hard red spring wheat varieties which varied for their productive tiller number (PTN) phenotypes and the alleles at the respective *QTn.mst-6B* markers, wPt4716 and wPt3581

following the methods described in Blake et al. (2011). NIL pairs derived from the cross ‘Reeder’(PI 613586)(high PTN) / ‘Choteau’ (PI 633974)(low PTN) (R/C) were evaluated in field trials across multiple environments using RCBD experimental designs and found to differ significantly for both early tiller number (ETN) and PTN in all tested environments (5 and 10 respectively) between *Q_{Tn.mst-6B}* alleles (Naruoka et al. 2011; Nasseer et al. 2016).

Jones (2020) then fine-mapped *Q_{Tn.mst-6B}* from its original region spanning 126,136,002 bp to 475,658,707 bp (positions based on Chinese Spring wheat IWGSC RefSeq v1.0 reference genome (IWGSC et al. 2018)) to a ~20Mb region on the short arm spanning positions 226,868,798 bp through 247,209,591 bp (Figure 1.1). This was done by crossing sister lines within the one R/C NIL pair previously validated in the aforementioned field trials. F_{2-HIF} and F_{3-HIF} progeny were used to create a high resolution genetic map with markers identified from a 90K SNP assay (Wang et al. 2014) and an exome capture assay (Krasileva et al. 2017). F_{4-HIF} recombinant and non-recombinant progeny were then evaluated for ETN in replicated greenhouse experiments.

In this study, twelve F_{6-HIF} progeny lines from the R/C NIL pair were genotyped with eight KASP markers developed from the exome capture assay (Krasileva et al. 2017) and identified as either non-recombinant (haplotypes A and B) or recombinant between markers 237.1_7559 (237,416,343 bp) and 238.1_7559 (238,168,303 bp) (haplotypes C and D) (Figure 1.1). These lines, in addition to the original two NIL sister lines that were crossed to develop them, were used in a series of four greenhouse experiments to measure ETN with the goal to further fine map the QTL region.

A)	Entry	Generation		Haplotype	
	1	F6-HIF		A	
	2	F6-HIF		A	
	3	F6-HIF		A	
	4	F6-HIF		B	
	5	F6-HIF		B	
	6	F6-HIF		B	
	7	F6-HIF		C	
	8	F6-HIF		C	
	9	F6-HIF		C	
	10	F6-HIF		D	
	11	F6-HIF		D	
	12	F6-HIF		D	
	13	Control: NIL Sister Line		A	
	14	Control: NIL Sister Line		B	
B)	KASP Marker Name	226.8_7553	237.1_7559	238.1_7559	247.2_7560
	Haplotype	226,868,798	237,416,343	238,168,303	247,209,591
	A	2	2	2	2
	B	1	1	1	1
	C	1	1	2	2
	D	2	2	1	1
		KASP Marker Position (bp)			
		Genotype: Parent 1 or Parent 2			

Figure 1.1. A) Table of 14 entries evaluated in greenhouse trials and their respective haplotypes. Entries 1-12 consist of F₆-HIF lines that either non-recombinant (haplotypes A, B) or recombinant (haplotypes C, D). Entries 13 and 14 are the original NIL sister lines crossed to develop the F₆-HIF population and serve as the check lines for the experiments. B) Table of genotypes across eight different KASP markers spanning the narrow QTL region of interest. Genotypes are either “1” representing the allele of the high-ETN NIL sister line, or “2” for the allele of the low-ETN NIL sister line. Haplotypes A, B, C, and D represent non-recombinant (A,B) and recombinant (C,D) groupings.

Greenhouse Phenotyping Trials

Greenhouse phenotyping trials were conducted at the Montana State University Plant Growth Center (PGC) under 16 hour day lengths (with supplemental lighting), average temperatures at approximately 21°C (day) and 20°C (night), and fertilized weekly with Peter’s

Professional 20-20-20 fertilizer mixture at approximately 100 ppm. Seven-inch pots were filled with 50:50 PGC soil mix, which is an equal blend of a custom MSU soil mix (consisting of equal parts by volume of loam soil, washed concrete sand, and Canadian Sphagnum peat moss) and Sunshine® Mix #1 (<https://agriculture.montana.edu/pgc/specifications.html>). Pots were planted with three seeds of one entry and then were thinned to a total of two plants. Four trials in total were conducted: one in the summer of 2021, the fall of 2021, the winter of 2022, and the spring of 2022. An RCBD design was used in each trial with one pot of two plants per entry per block and 20 blocks in summer 2021 and 30 blocks total in the other three trials. Early tiller number (ETN) was measured as the total number of observable tillers (including the primary tiller) for a given plant and was collected on each plant in a trial at 4 weeks, 5 weeks, and 6 weeks after planting.

Greenhouse Data Analysis

The ETN counts collected on different weeks were analyzed separately. Data was combined across all four trials and analyzed with the following generalized linear mixed model:

$$ETN_{ijklm} \sim \text{Poisson}(\mu_{ijklm})$$

$$\log(\mu_{ijklm}) = \beta_0 + \beta_{1i} \text{Haplotype}_i + \beta_{2j} \text{Trial}_j + \beta_{3ij} \text{Haplotype} \times \text{Trial}_{ij} + \gamma_{jk} \quad (1.1)$$

$$\gamma_{jk} \stackrel{iid}{\sim} N(0, \sigma_{Block}^2)$$

where ETN_{ijklm} represents the early tiller number count of the m^{th} plant in the l^{th} pot in the k^{th} block of the j^{th} greenhouse trial with the i^{th} haplotype. β_0 is the global intercept, β_{1i} is the effect on the log mean ETN of i^{th} level of haplotype, β_{2j} is the effect on the log mean ETN of the j^{th} level greenhouse trial, β_{3ij} is the effect on the log mean ETN of the interaction between the i^{th}

haplotype level and the j^{th} greenhouse trial, and γ_{jk} is the random effect associated with the k^{th} block nested within the j^{th} greenhouse trial. The Poisson distribution was used because the response variable was a discrete count of tiller number, and therefore, the canonical link function for the Poisson, natural log, was used. This meant that all model outputs were interpreted on the natural log (hereafter “log”) scale.

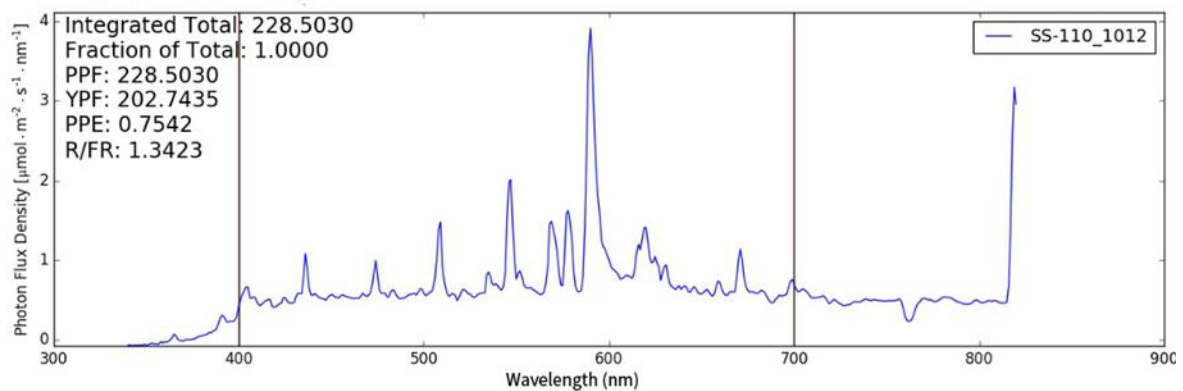
Models were fit using the `glmer` function from the `lme4` R package (Bates et al. 2015). Analysis of deviance based on Type II Wald chi-square tests was performed on fitted models using the `Anova` function from the `car` package (Fox and Weisberg 2019) (Table 1.1). Given significant interactions ($p\text{-value} < 0.05$) between trial and haplotype in all models, the `emmeans` R package (Lenth 2025) was then used to calculate estimated marginal log mean counts for each haplotype within each trial. The `multcomp` R package (Hothorn et al. 2008) was then used to assign compact letter display groupings of the haplotype levels within a trial based on pairwise contrasts and p -values corrected with Sidak’s method (Figure 1.4) which was selected for its stringent family-wise error control because it was critical to avoid type I errors which would cause us to fine map away from our true causal gene.

New Greenhouse Grow Lights Pilot Study

In the fall of 2023, new LED high-intensity grow lights, which had an increased red to far red (R/FR) ratio, were installed in the PGC greenhouse where previous phenotyping trials were conducted. In order to test the effect of the new grow lights on ETN and potentially observed stronger genetic effects, a pilot study was conducted using two $F_{7\text{-HIF}}$ recombinant entries genotyped with the same KASP markers and having haplotypes C and D respectively, as well as the NIL sister line checks with haplotypes A and B from the previous trials. In this study, eight

inch pots were used to minimize any resource competition between plants, two plants were kept per pot, temperature and day length were the same as previous trials, and the same fertilizer was applied twice per week at 100ppm to account for previously observed increased resource demands of spring wheat plants grown under the new lights (unpublished). An RCBD approach was used with twelve blocks in total. ETN was collected five and six weeks after planting, and data was analyzed in the same method as above but without the interaction term in the model. Specifically, the mean log early tiller numbers of the haplotype groups were compared in order to test whether any groups had different tiller numbers at either time point.

MVR1000/C/U Multi-Vapor bulbs



Arize Element L1000 Gen2 LED

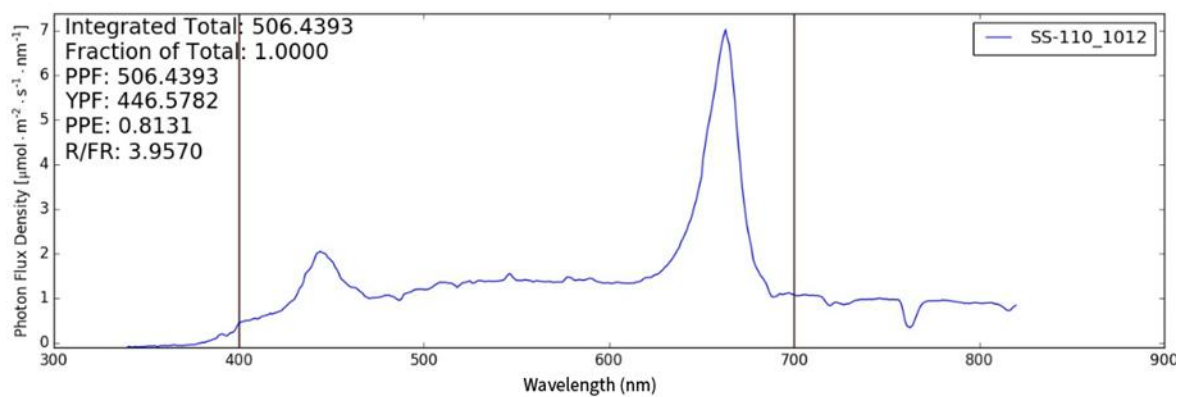


Figure 1.2. Light output metrics between the metal halide grow lights (MVR1000 Multi-Vapor; top chart) and the replacement LED lights (Arize Element L1000 Gen2; bottom chart). Graphs

show the measured photon flux density at each light wavelength from 300 to 900 nm. The total photon output (Integrated Total), fraction of total output captured, photosynthetic photon flux (PPF), yield photon flux (YPF), photon efficacy (PPE), and Red to Far Red light ratio are also reported for each grow light.

RNA-seq Analysis

RNA-seq was conducted in order to investigate differentially expressed genes occurring within *QTn.mst-6B*. Three biological replicates of the R/C NIL sister lines were grown in the greenhouse in the summer of 2020. Four weeks after planting, tiller bud tissue was extracted from the plants, and total RNA was extracted from the samples. Samples were then sent to the company GENEWIZ for sequencing. mRNA was isolated via PolyA selection and Illumina HiSeq 2x150 bp paired-end sequencing was performed.

Raw read files for each replicate were returned, and quality analysis was conducted using the FASTQC program v0.11.9 (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Adapter trimming and quality filtering was done using the *bbduk* tool v38.90 from the BBTools suite (Bushnell 2014). Reads were aligned to the Chinese Spring wheat IWGSC RefSeq v2.1 reference genome (Zhu et al. 2021) using the program STAR v2.7.9a (Dobin et al. 2013) and read counts were collected with the HTseq program v4.8.3. Finally, differential expression was analyzed using the R program DEseq2 (Love et al. 2014). Sample quality was assessed via principal component analysis (PCA) in which it was verified that samples from the same NIL parent clustered together. Then, differentially expressed genes (DEGs) were declared as those which had an adjusted p-value less than 0.01 when comparing expression between high-tiller samples to low-tiller samples. P-values were adjusted following the independent hypothesis weighting method (Ignatiadis et al. 2016) in which p-values were weighted by the mean normalized read counts for their respective gene. The false discovery rate was then controlled at

the default alpha value of 0.1. This more lenient p-value adjustment method was used as RNA-seq is a broad and exploratory method, so type 1 error control was less critical than in other analysis procedures we conducted.

RNA-seq Variant Calling with GATK

The “RNAseq short variant discovery” pipeline from The Genome Analysis Toolkit (GATK) website (<https://gatk.broadinstitute.org/hc/en-us/articles/360035531192-RNAseq-short-variant-discovery-SNPs-Indels>) was used as a guide for performing variant calling from the RNA-seq data files. All tools were from GATK v4.2.4.1 (McKenna et al. 2010; Van der Auwera and O’Connor 2020). Briefly, the STAR-aligned read files from the main RNA-seq analysis were first passed into the ‘MarkDuplicates’ function to identify and tag all reads that originate from the same DNA sequence. Next, the ‘SplitNCigarReads’ function was used to split reads that spanned splice events. Then, ‘HaplotypeCaller’ was used to identify SNPs and indels within each RNA-seq sample. Finally, the ‘VariantFiltration’ function was used to identify any variants with a Fisher Strand Bias (FS) value greater than 30 and a Quality by Depth (QD) less than 2.0. Variant files for each sample were then transferred into R. A variant between high/low tiller NIL parents was defined as a called variant that was present in all three samples for one tiller group but not present in all three samples for the other tiller group.

Orthologous Gene Identification

Tillering has been well studied in rice (*Oryza sativa*), with several major genes identified for the trait (Yan et al. 2023). Four major rice tillering genes, *MONOCULMI* (*MOCI*) (Li et al. 2003) GenBank: AY242058.1, *LAX PANICLE1* (*LAX1*) (Komatsu et al. 2001) GenBank: AB115668.1, *LAX PANICLE2* (*LAX2*) (Tabuchi et al. 2011) GenBank: AB669025.1, and

SQUAMOSA PROMOTER BINDING PROTEIN-LIKE14 (*OsSPL14*, or *IPAI*) (Jiao et al. 2010; Miura et al. 2010) GenBank: GU136674.1 were investigated for possible wheat orthologs. To do this, the CDS DNA sequences of each gene were compared against the wheat Chinese Spring IWGSC RefSeq v2.1 reference genome (Zhu et al. 2021) using the BLAST tool available from the T3 database website (Morales et al. 2022) with default parameters. The best results based on percent alignment and e-value were investigated in the JBrowse genome browser from the GrainGenes database (Yao et al. 2025) and in the Persephone genome browser (<https://persephonesoft.com/>) in order to view predicted genes from the Gnomon method (Souvorov et al. 2010). Genes were deemed candidate wheat orthologs by the combined evidence of the strength of the BLAST result and having a gene annotation in either the IWGSC high confidence gene annotations or in the Gnomon annotations that matched the same protein family as the original rice gene.

The same methods were used to investigate genes interacting with *MOC1* including *TILLERING AND DWARF1* (*TAD1*) (Xu et al. 2012) GenBank: AK058595.1, *MONOCULM 3* (*MOC3*) (Lu et al. 2015) GenBank: AB218894.1, *FLORAL ORGAN NUMBER1* (*FON1*) (Nagasawa et al. 1996) GenBank: AB182388.1, and *SLENDER RICE1* (*SLR1*) (Ogawa et al. 2000; Ikeda et al. 2001; Liao et al. 2019) GenBank: AB030956.1. For the *Arabidopsis* gene *MORE AXILLARY GROWTH2* (*MAX2*), the genomic DNA sequence was obtained from TAIR and subsequent analysis followed as before. For genes already described in wheat including the *SQUAMOSA PROMOTER BINDING PROTEIN-LIKE* genes (*TaSPL028*, *TaSPL029*, *TaSPL030*) (Zhu et al. 2020) and *TaSPL14* (sometimes called *TaSPL17*) (Cao et al. 2023), *TEOSINTE BRANCHEDI* (*TaTBI*) (Liu et al. 2017), *DWARF27* (*TaD27*) (Zhao et al. 2020), and

CRYPTOCHROME (*TaCRY1*, *TaCRY2*) (Yan et al. 2021) positions and gene names were updated to the newest version of the IWGSC reference genome (v2.1) using the genome browsers.

Results

Four greenhouse experiments were conducted between 2021 and 2022 encompassing all four seasons of external light and daylength patterns. In each trial, early tiller number (ETN) was counted for each single plant at four, five, and six weeks after planting. We found that the overall mean ETN across all haplotype groups was approximately 2.7 tillers at four weeks, 3.3 tillers at five weeks, and 3.8 tillers at six weeks after planting (Figure 1.3). We found that all entries in haplotype group A had mean ETNs equal to or below the overall mean, and all entries within haplotype group B had means greater than the overall mean ETN at all three time points across all four greenhouse trials (Figure 1.3). However, both haplotype groups C and D had entries with mean ETNs above and below the overall means at all timepoints across all trials (Figure 1.3).

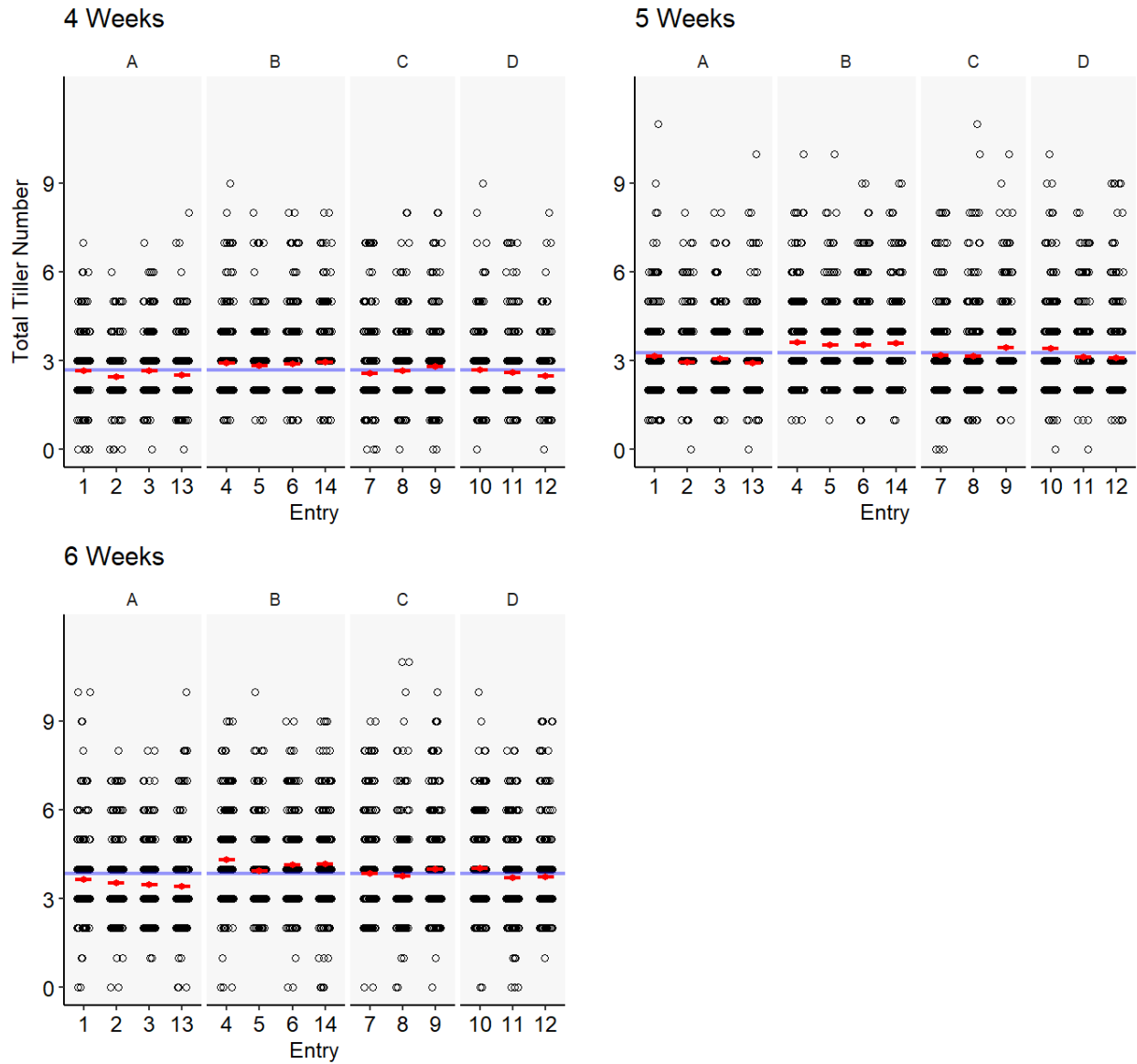


Figure 1.3. Early tiller number (ETN) counts are given for four, five, and six weeks after planting. Data is combined from four different greenhouse trials, and a single point indicates one plant-level observation within a pot, block, and trial. Different entries are denoted on the X-axis, and entries are grouped within shared haplotypes by the letters at the top of the plots. The blue horizontal line indicates the overall mean tiller number within the time after planting, across all trials. The red horizontal bars indicate the entry-level mean tiller numbers across all trials within the given time after planting.

After model fitting to test the log mean early tiller numbers between haplotype groups, we found significant interactions ($p\text{-value} < 0.05$) between haplotype groups and greenhouse

trials at all three timepoints, indicating that the time of year a greenhouse trial was conducted influenced the relationships between the haplotype groups (Table 1.1; Figure 1.4). Therefore, we performed pairwise contrast analysis between haplotype groups within each greenhouse trial separately. We detected a difference in the mean log early tiller counts (Sidak's adjusted p-value < 0.05) between haplotype groups A and B in the fall and summer trials at four weeks after planting; in the fall, spring, and summer trials at five weeks; and in fall, winter, and spring trials at six weeks (Figure 1.4). However, we were not able to detect a difference in mean log tiller counts between haplotypes C and D in any greenhouse trial at any timepoint (Figure 1.4).

Weeks After Planting	Model Term	Wald Chi-Square	Degrees of Freedom	P-value
4 Weeks	Haplotype	21.39	3	8.75E-05
4 Weeks	Trial	257.42	3	1.63E-55
4 Weeks	Haplotype x Trial	18.20	9	0.03288
5 Weeks	Haplotype	39.93	3	1.10E-08
5 Weeks	Trial	238.56	3	1.95E-51
5 Weeks	Haplotype x Trial	27.12	9	0.00133
6 Weeks	Haplotype	46.49	3	4.46E-10
6 Weeks	Trial	406.73	3	7.70E-88
6 Weeks	Haplotype x Trial	29.62	9	0.00051

Table 1.1. Results for Type II analysis of deviance for early tiller number collected at three different time points across four different experiments. Three separate generalized linear mixed models were fit; one for each time point of collection shown in column 1. Column 2 gives the model fixed effect term for the given time point, column 3 gives the associated Wald chi-square test statistic for the model term, column 4 gives the degrees of freedom for the chi-square test, and column 5 gives the unadjusted the p-values.

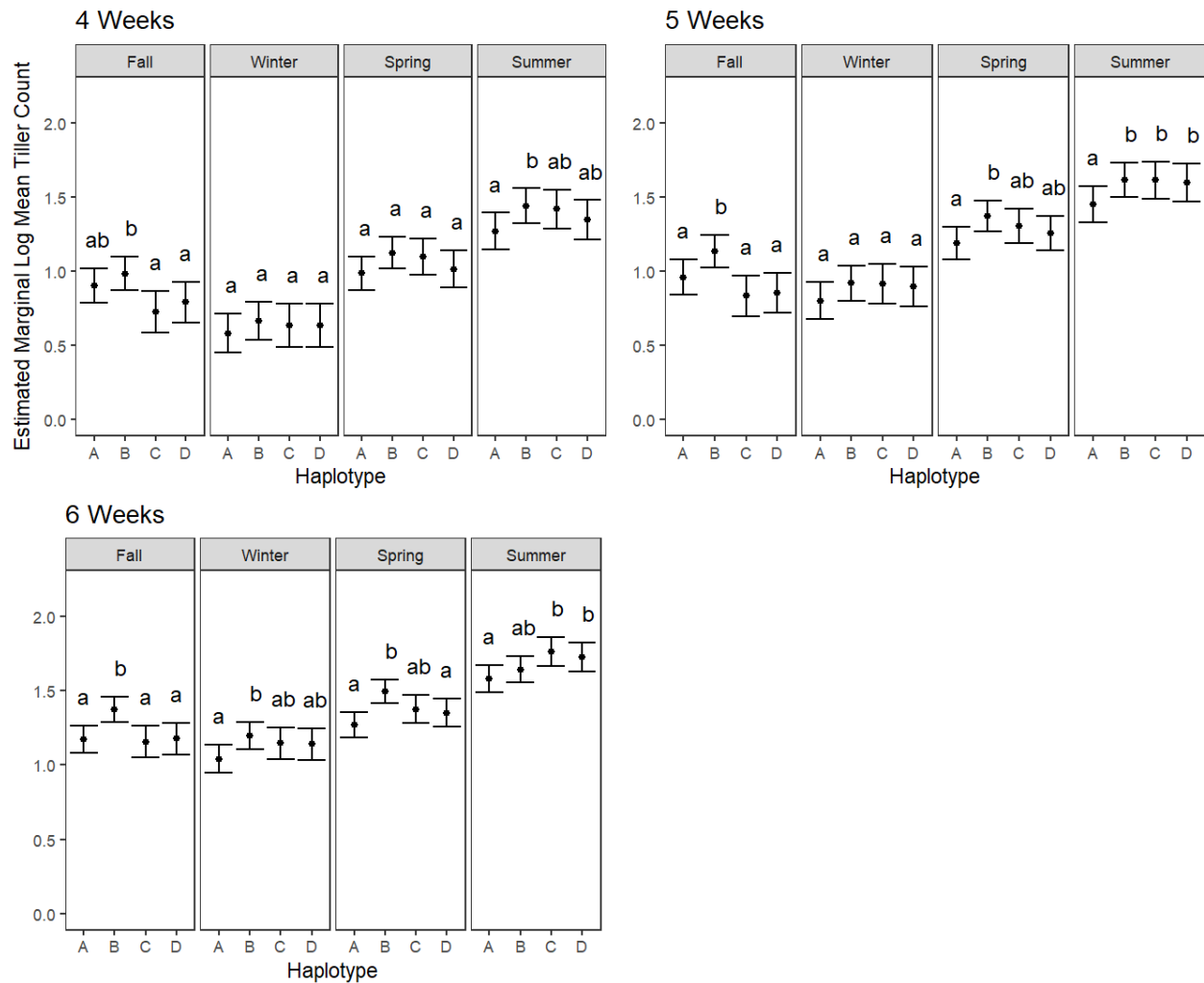


Figure 1.4. Estimated marginal log mean early tiller numbers are given between different haplotypes on the X-axis grouped within one of four greenhouse trials with data taken either four, five, or six weeks after planting. Bars indicate 95% confidence intervals around the estimated means and letters indicate grouping based on pairwise contrasts between haplotypes within a given trial with p-values adjusted by Sidak's method.

A later greenhouse study was conducted during the winter of 2023 using the new LED grow lights (Figure 1.2) and ETN was collected for each plant at five (Figure 1.5) and six (not shown) weeks after planting. Tiller counts were very similar between the two timepoints, and the overall mean ETN was 16.41 tillers at five weeks after planting (Figure 1.5). The mean tiller counts of each haplotype group at five weeks were 16.6 for A, 16.5 for B, 18.3 for C, and 14.25

for D (Figure 1.5). We were able to detect a difference in log mean tiller counts between haplotype groups C and D (Sidak's adjusted p-value < 0.05) but not between haplotype groups A and B (Figure 1.5).

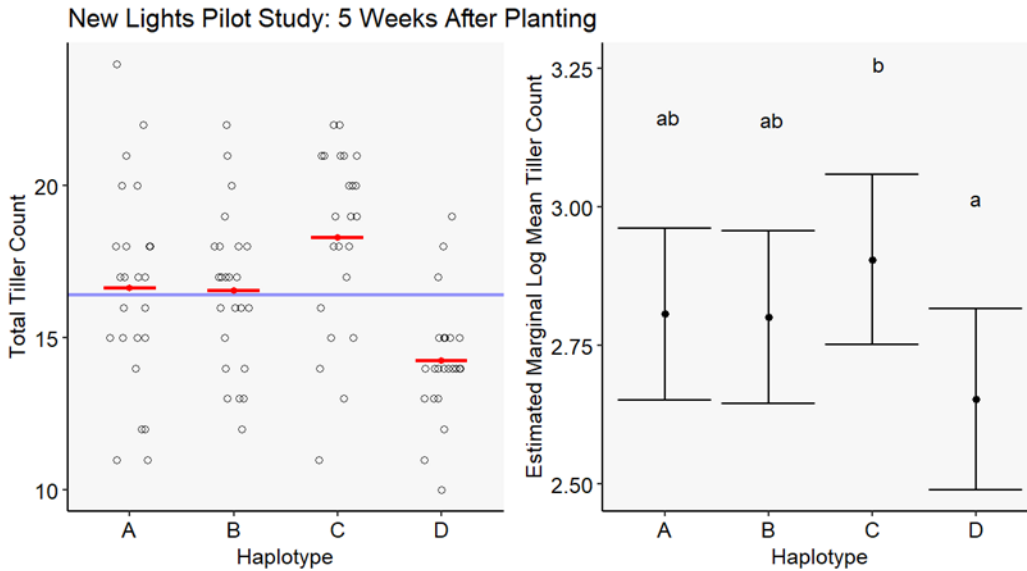


Figure 1.5. Results from the winter 2023 pilot study using the new LED grow lights. The left figure shows the raw ETN counts for each plant grouped by haplotype on the X-axis. The blue line indicates the overall mean ETN and individual red bars indicate the mean tiller counts within one haplotype. The right panel shows estimated log mean ETN values for the haplotype groups with bars representing 95% confidence intervals. Letters indicate grouping based on pairwise contrasts between haplotype groups with p-values adjusted by Sidak's method.

In order to investigate the expression patterns of genes located in and around our target QTL, we next conducted RNA sequencing on the NIL parents (haplotypes A and B) used as checks in the greenhouse studies. RNA-seq analysis revealed a total of 245 differentially expressed genes (DEGs) (adjusted p-value < 0.01) (Supplementary Table 2.1) across three samples each between the high and low tiller NIL parents. Of these, 75 DEGs were located on the target chromosome 6B, and four DEGs were located within the narrow region of *QTn.mst-6B* (Figure 1.6 A, B). Two DEGs in the narrow region had characterized functions including

TraesCS6B03G0506800 annotated as a S-adenosylmethionine synthetase and

TraesCS6B03G0522300 annotated as a putative F-box protein (Figure 1.6B).

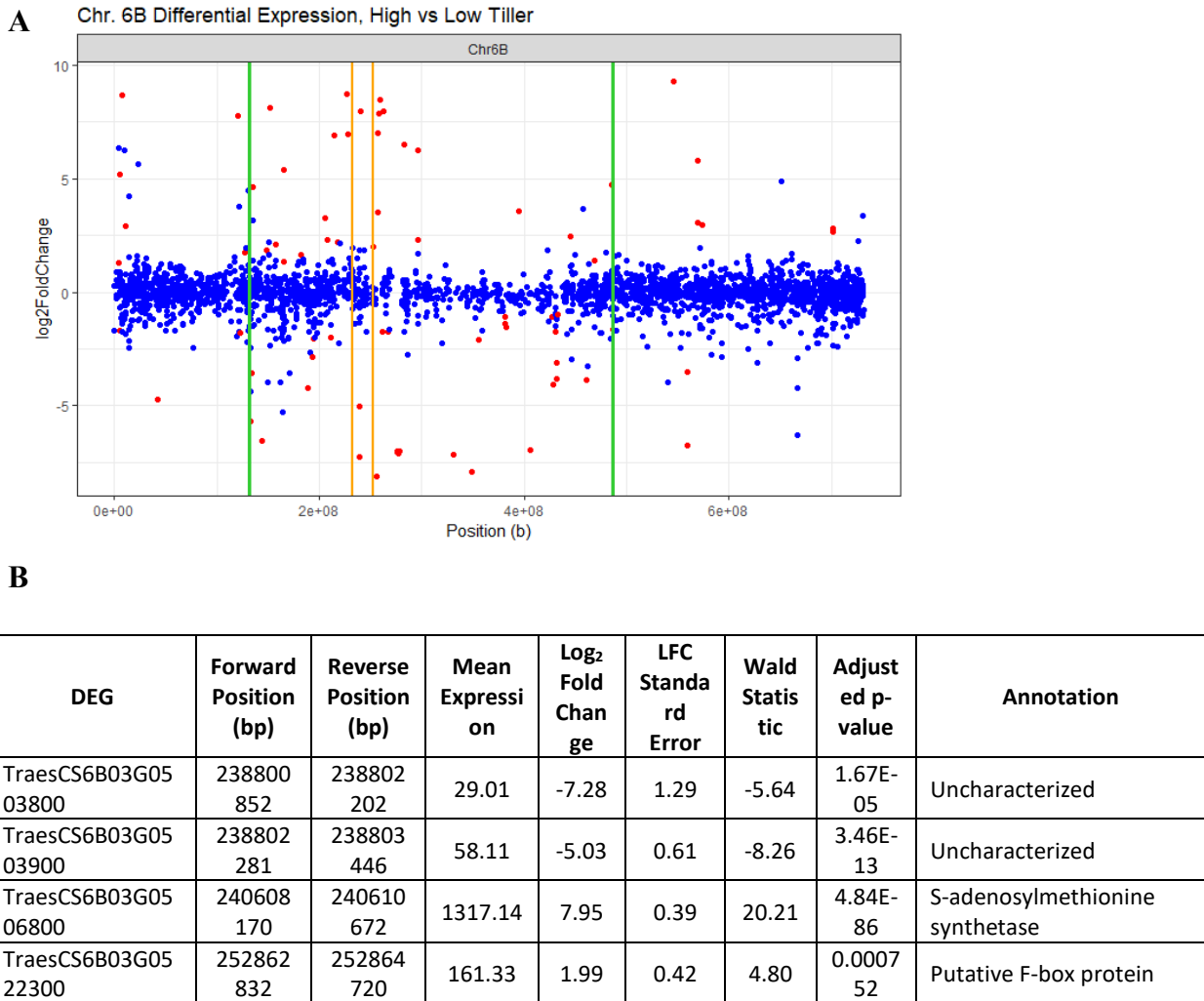


Figure 1.6. A) Differential expression analysis results from RNA-seq between high and low tiller NIL sister lines. Each point represents one gene on chromosome 6B of spring wheat. The Y-axis gives the log base 2 fold expression change of a gene in the high tiller line compared to the low tiller line. The X-axis gives the base pair position of the gene. A red point indicates the gene was differentially expressed (adjusted p-value > 0.01) whereas blue indicates the gene was not. The green vertical bars indicate the original range of *Qtn.mst-6B* where the orange, vertical lines indicate the fine-mapped narrow region of the QTL. B) Table of the four DEGs detected in the narrow QTL range, including their names, positions, associated statistics and functional annotations.

Major tillering genes in rice and *Arabidopsis* were aligned against the IWGSC wheat reference genome (v2.1) using BLAST. The top alignments based on percent identity and e-value were recorded as candidate orthologs along with the respective chromosomes, positions, and RNA-seq-derived results including expression patterns and detected variants (Table 1.2). In sum, two homeologous sets of candidate orthologs were detected for the rice gene *MOC1*, along with one homeologous set respectively for *LAX1*, *LAX2*, and *OsSPL14* (Table 1.2). Further investigation of *MOC1*-associated genes revealed two sets of candidates for *TAD1*, and one set for *FON1*, *MOC3*, and *SLR1* as well as one set for the *Arabidopsis* gene *MAX2* (Table 1.2). In all cases, the presence of a possible second set of orthologs was only considered if that set occurred on the candidate chromosome 6B. Additionally, the same information is provided for other, previously described orthologs in wheat, and across all investigated genes, none were differentially expressed between tiller groups, though many had possible sequence variants (Table 1.2).

Gene	Possible Wheat Orthologs	E-value	Chromosome	Forward Position (bp)	Base Mean Expression	Differential Expression Adjusted P-value	Number of Variants
Rice Tiller Genes							
MOC 1	TraesCS6B03G0494700	8.00E-96	6B	232245538	12.75	0.91	2
MOC 1	TraesCS6A03G0387600	5.00E-98	6A	154238173	15.70	0.98	0
MOC 1	TraesCS6D03G0340500	8.00E-96	6D	148584459	11.78	0.98	0
MOC 1	TraesCS7A03G0927500	0	7A	561458531	23.24	0.83	0
MOC 1	TraesCS7B03G0775400	0	7B	524362047	10.94	0.96	1
MOC 1	TraesCS7D03G0891300	0	7D	493162825	26.34	0.93	0
LAX 1	TraesCS3A03G0836000	3.00E-106	3A	599733908	3.31	0.99	0
LAX 1	TraesCS3D03G0767700	8.00E-102	3D	457365439	4.04	0.97	0
LAX 1	TraesCS3B03G0959300	1.00E-94	3B	617794265	1.05	NA	0
LAX 2	TraesCS2A03G0717000	7.00E-34	2A	492628649	41.52	0.93	0
LAX 2	TraesCS2D03G0662700	3.00E-32	2D	365768786	39.95	0.96	0
LAX 2	TraesCS2B03G0795200	3.00E-32	2B	440073002	35.82	0.70	0
OsSPL14,IPA 1	TraesCS7A03G0567100	3.00E-58	7A	230286572	562.18	0.92	4
OsSPL14,IPA 1	TraesCS7D03G0548900	2.00E-56	7D	215738730	251.05	0.63	0
OsSPL14,IPA 1	TraesCS7B03G0393600	7.00E-55	7B	201309179	124.39	0.87	0
<i>MOC1</i> Associated Gene (Rice)							
TAD1	TraesCS4D03G0793000	2.00E-50	4D	498056300	314.27	0.98	1

Table 1.2 continued

Gene	Possible Wheat Orthologs	E-value	Chromosome	Forward Position (bp)	Base Mean Expression	Differential Expression Adjusted P-value	Number of Variants
TAD1	LOC123105939	3.00E-48	5A	680361355	NA	NA	1
TAD1	TraesCS4B03G0900800	4.00E-47	4B	638971824	273.39	0.96	2
TAD1	LOC123158052	8.00E-39	7B	458597779	NA	NA	0
TAD1	LOC123153644	8.00E-39	7A	648498630	NA	NA	2
TAD1	LOC123139264	8.00E-39	6B	456384999	NA	NA	6
FON1	TraesCS7A03G1298300	0	7A	717685438	371.84	0.91	3
FON1	TraesCS7D03G1230700	0	7D	621095001	347.24	0.98	3
MOC 3	LOC123042297	2.00E-71	2B	723224185	NA	NA	0
MOC 3	LOC123050233	5.00E-68	2D	593111332	NA	NA	0
MOC 3	TraesCS2A03G1149400	2.00E-67	2A	728361298	0	NA	0
SLR 1	TraesCS4A03G0701300	0.00E+00	4A	581396439	5619.94	0.93	0
SLR 1	TraesCS4B03G0093100	0.00E+00	4B	33614435	6355.63	0.95	0
SLR 1	TraesCS4D03G0067100	0.00E+00	4D	19189660	2773.56	0.82	0
FON1	LOC123157457	0	7B	722812608	NA	NA	1

Table 1.2 continued

Gene	Possible Wheat Orthologs	E-value	Chromosome	Forward Position (bp)	Base Mean Expression	Differential Expression Adjusted P-value	Number of Variants
Arabidopsis Gene							
MAX2	TraesCS7B03G0019600	4.00E-18	7B	5391944	655.57	0.94	2
MAX2	TraesCS7D03G0240500	9.00E-15	7D	65749490	442.28	0.98	1
MAX2	TraesCS7A03G0262100	9.00E-15	7A	70993190	383.85	1.00	4
Wheat Gene	Wheat Gene Name						
TaSPL028	TraesCS6B03G0349300		6B	141662416	772.98	0.94	8
TaSPL029	TraesCS6B03G0457800		6B	206683594	2.48	NA	0
TaSPL030	TraesCS6B03G0465800		6B	211073058	192.92	0.99	6
TaTB1	TraesCS4A03G0702300		4A	581754927	82.30	0.32	0
TaTB1	TraesCS4B03G0092100		4B	33121434	61.65	0.11	0
TaTB1	TraesCS4D03G0066700		4D	18876554	142.05	0.98	2
TaD27	TraesCS7A03G1015100		7A	614161029	23.14	0.99	1
TaD27	TraesCS7B03G0855500		7B	573258279	31.30	0.97	0
TaD27	TraesCS7D03G0971000		7D	531954416	28.69	0.96	0
TaCRY1	TraesCS6B03G0620200		6B	344914861	638.08	0.48	0
TaCRY1	TraesCS6A03G0549000		6A	349034109	624.78	0.94	1
TaCRY1	TraesCS6D03G0463200		6D	271220502	446.69	0.99	0
TaCRY2	TraesCS6B03G0754600		6B	470319798	276.12	0.57	5

Table 1.2 continued

Gene	Possible Wheat Orthologs	E-value	Chromosome	Forward Position (bp)	Base Mean Expression	Differential Expression Adjusted P-value	Number of Variants
TaCRY2	TraesCS6A03G0617600		6A	422875249	418.97	0.96	0
TaCRY2	TraesCS6D03G0524500		6D	320691996	423.21	0.93	0

Table 1.2. Table of possible orthologous tillering genes from rice and *Arabidopsis* as well as potential tillering genes already identified in wheat. The gene names, chromosomes, and positions are provided from IWGSC RefSeq v2.1 as well as the mean expression, differential expression adjusted p-value, and detected variants from the RNA-seq data. Bolded rows indicate genes that are located on chromosome 6B.

Discussion

Greenhouse Phenotyping Trials

The expected result of the greenhouse phenotyping trials was that haplotypes A and B, corresponding to the low and high tiller parental check NILs respectively, would have significantly different ETNs at both four, five, and six weeks after planting for all trials given that the parental NILs had shown significantly different ETNs and PTNs in previous field studies (Naruoka et al. 2011; Nasseer et al. 2016; Jones et al. 2021). We also expected that recombinant haplotypes C and D would also have significantly different ETNs, given that they had contrasting genotypes at all tested markers in the QTL (Figure 1.1B). Finally, we expected that entries within a haplotype group would have similar ETNs due to being derived from NILs and having identical QTL haplotypes (Figure 1.1A).

We saw that overall ETNs increased between four and five weeks after planting and stayed consistent between five and six weeks after planting across all greenhouse trials (Figure 1.3). We also saw that entries within haplotypes A and B performed similarly to one another

relative to the overall means. However, entries within haplotypes C and D did not appear to perform consistently with entry 9 in haplotype C and 10 in haplotype D typically having seemingly contrasting phenotypes to the other entries in their groups (Figure 1.3), though this was not formally tested. This may suggest that the entries within the recombinant haplotypes are not genetically identical, though given the narrow tiller number ranges and probable confounding light stress, it is difficult to know for sure.

We observed that the timing of trials had a large effect on overall tiller numbers across haplotypes (Table 1.1), with the trials conducted during times with longer day lengths (spring and summer) having increased tiller numbers compared to those conducted in the fall and winter (Figure 1.4). Despite this, at six weeks after planting, we were able to detect differences in ETN between haplotypes A and B in all greenhouse trials except for the summer trial in which we detected differences at four and five weeks, but not six (Figure 1.4). However, there was not a single trial in which differences in tiller number could be detected between haplotypes C and D, possibly due to the inconsistency of the entries within those groups. Therefore, we were not able to fine map *QTn.mst-6B* any further from these greenhouse trials even with a high number of replicates.

Light Regulation of Tiller Number

The change in experimental conditions between the original metal halide grow lights and the new, high intensity, LED grow lights resulted in a dramatic increase in early tiller number (ETN) across all haplotypes. Specifically, at 5 weeks after planting, the overall mean ETN was around 4 tillers across four different greenhouse trials and across all haplotype groups using the original grow lights (Figure 1.3). In the pilot study using the new LED lights, at five weeks after

planting, the mean ETN across haplotypes was around 16 tillers, representing a 4-fold increase in mean ETN (Figure 1.5). This clearly demonstrated the effect of light conditions on tiller number in wheat, which was consistent with the findings of previous literature. Specifically, BOS and NEUTEBOOM (1998) found that increased light intensity (measured in $\mu\text{mol m}^{-2}\text{s}^{-1}$) resulted in an increase in tiller number. Further, multiple studies have shown that a low ratio of red to far red light (R/FR) is associated with a decrease in tiller number in wheat (Casal et al. 1986; EVERS et al. 2006; Ugarte et al. 2010; Xie et al. 2016). This finding is also related to the association between planting density and tiller number where studies have shown that greater planting densities are associated with a decrease in tiller number, possibly due to triggering shade response pathways (Casal et al. 1986; Jones et al. 2021). Comparing the two grow lights used, the LEDs had a much higher total light intensity (506 vs 229) and R/FR ratio (1.34 vs 3.96) (Figure 1.2), both of which could have contributed to the increase in observed ETN.

Despite the clear influence of light conditions on tiller number in this study, it is unlikely that light quality is related to the causal variant of *QTn.mst-6B*. The phytochrome R/FR light receptor genes as well as major phytochrome interacting factors (PIFs) have been described in wheat, and none fall within chromosome group 6 (Zhuang et al. 2024a; Yang et al. 2025). The blue light receptor gene *CRYPTOCHROME1* (*CRY1*) has been shown to repress hypocotyl elongation in *Arabidopsis* by inhibiting auxin (Xu et al. 2018; Mao et al. 2020) and brassinosteroid (BR) (Wang et al. 2018; He et al. 2019) signaling. Further, in both *Arabidopsis* and wheat, *CRY1* has been shown to repress gibberellic acid (GA) signaling and thus stabilize DELLA proteins and repress plant growth (Yan et al. 2021), and the wheat orthologs, *TaCRY1* and *TaCRY2*, are located on the long arms of chromosome group 6 (Yan et al. 2021; Table 1.2).

None of *TaCRY* genes we investigated were differentially expressed, though *TaCRY2b* did contain multiple sequence variants between the high and low tiller groups (Table 1.2).

RNA-seq Analysis

There was a total of 245 differentially expressed genes (DEGs) between the high and low tiller groups detected in the RNA-seq analysis (Supplementary Table 2.1). Of these DEGs, 75 were located on chromosome 6B and four fell within the narrow region of *QTn.mst-6B* (Figure 1.6 A,B). The first two DEGs, *TraesCS6B03G0503800* and *TraesCS6B03G0503900*, were located only 79 bp apart and though they were annotated as separate genes in the IWGSC high confidence gene annotations, the Gnomon model in the Persephone browser considered them one, unified gene. Regardless, both genes were functionally uncharacterized. The remaining two genes, *TraesCS6B03G0506800* and *TraesCS6B03G0522300*, were functionally annotated as “S-adenosylmethionine synthetase” and “Putative F-box protein” respectively (Figure 1.6B). *TraesCS6B03G0522300* was especially of interest because the F-box protein *MAX2* in *Arabidopsis* has been associated with axillary bud outgrowth (Stirnberg et al. 2002; Bennett et al. 2006). However, BLAST analysis suggested that *MAX2* is most strongly orthologous to chromosome group 7 (Table 1.2) and F-box proteins in general can be involved in many different biological pathways which may not be related to tillering (Kuroda et al. 2002).

Orthologous Gene Analysis

Tiller number has been comparatively well studied in rice (*Oryza sativa*), with many important genes identified (Yan et al. 2023). As such, we wanted to investigate if any major tillering-related genes in rice had possible orthologs to genes in or near our *QTn.mst-6B* locus. The major genes investigated were *MOC1*, *LAX1*, *LAX2*, and *IPA1* (*SPL14*). Of these, we found

one possible ortholog of *MOC1*, the gene *TraesCS6B03G0494700* located on chromosome 6B at forward position 232245538 bp (Table 1.2). Not only was this gene in very close proximity to the narrow QTL region (232965065 bp), it was also the only candidate ortholog of the four tested genes to land on chromosome 6B in general (Table 1.2). Notably, a wheat ortholog of *MOC1*, *TaMOC1*, has been described in previous literature (Zhang et al. 2015), but this refers to the orthologs located on chromosome group 7 (Table 1.2) which were initially found to be primarily associated with spikelet development (Zhang et al. 2015) though recently, Maqbool et al. (2025) has associated it with tiller number as well. While the group 7 orthologs had the strongest BLAST results, the next strongest hits were the group 6 orthologs which still had very small e-values (Table 1.2) and whose Persephone annotations still related them to *MOC1*. Therefore, we conclude that there may be two separate sets of *MOC1* orthologs in wheat.

MOC1 encodes a transcription factor in the plant GRAS gene family (Pysh et al. 1999; Li et al. 2003) which are involved in the signaling of plant phytohormones, especially gibberellic acid, as well as in many growth and development pathways (Khan et al. 2022). *MOC1* specifically is an ortholog of the *Lateral suppressor (Ls)* gene in tomato (Schumacher et al. 1999) and the *Arabidopsis Lateral Suppressor* gene (*LAS*) (Greb et al. 2003) in *Arabidopsis*. *MOC1* has been shown to regulate the initiation and outgrowth of axillary meristems with knockout mutations resulting in a monoculm phenotype (Li et al. 2003; Xu et al. 2012). Several relationships in rice between *MOC1* and other genes have been described. These include interactions with *MOC3* and *FON1* (Shao et al. 2019), *TAD1* (Xu et al. 2012), and *SLR1* (Liao et al. 2019). Therefore, we next investigated potential wheat orthologs of these genes to see if they demonstrated notable expression differences or contained detectable variants between the NIL

parent lines. We found potential orthologs for all of the above genes, and found that candidate orthologs of *FON1* and *TADI* as well as the known *SLR1* orthologs (the reduced-height *RHT* genes) were highly expressed in the NIL parents, though not differentially.

The *SPL* gene class has been characterized in numerous plant species (Cardon et al. 1999; Manning et al. 2006; Hou et al. 2013; Tripathi et al. 2017) and has been shown to contribute to multiple biological functions (Cardon et al. 1997; Chen et al. 2010). Within wheat, gene expression studies have shown that *TaSPL* genes are expressed in different tissues across all different growth stages, indicating that they also contribute to many biological processes (Zhu et al. 2020). The *SPL* gene *OsSPL14* has been shown to regulate tiller number in rice (Jiao et al. 2010; Miura et al. 2010), and its wheat ortholog, *TaSPL17* (also called *TaSPL14* or *TaIPAI*), has been shown to regulate tiller number in wheat (Cao et al. 2023; Jian et al. 2024). *TaSPL17-A* was strongly expressed in both NIL groups (base mean expression = 562), though not differentially (adj. p-value = 0.92), and it was the only homeolog of *TaSPL17* that had sequence variants between the NIL groups (Table 1.2).

It has been shown that some *TaSPL* genes regulate the expression of the major wheat tillering gene *TaTBI* (Liu et al. 2017). Zhu et al. (2020) described three *SPL* genes (*TaSPL028*, *TaSPL029*, and *TaSPL030*) found on chromosome 6B. Of these, *TaSPL028* and *TaSPL030* were both highly expressed, though not differentially, in all samples, and they both contained multiple sequence variants between high tiller and low tiller groups (Table 1.2). One possible hypothesis would be that these *SPL* genes are regulating *TaTBI* and thus causing the tillering phenotypes. Of the three *TaTBI* genes, there was some marginal evidence that the chromosome 4B homeolog (*TraesCS4B03G0092100*) was differentially expressed (adj. p-value = 0.11) between tiller

groups, though it had the lowest mean expression of the homeologs (Table 1.2). Also, the fold change indicated higher expression of the gene in the high tiller group (Supplementary Table 2.1), which would be counter to expectation as increased *TBI* expression decreases tiller number.

The plant hormones strigolactones (SLs) and brassinosteroids (BRs) have been shown to antagonistically regulate the expression of *TBI* in rice (Fang et al. 2020). Further, knockouts of the wheat SL synthesis gene, *TaD27*, have resulted in an increase in tiller number (Zhao et al. 2020; Awan et al. 2024). However, though tiller number was significantly changed, there was no difference detected in *TaTBI* expression (Zhao et al. 2020), suggesting that *TaD27* may not regulate *TBI* expression in wheat as it does in other plant species. We detected very low expression of *TaD27* in our NILs and found only one variant across the three homeologs, providing little evidence that this gene was relevant to our QTL (Table 1.2).

Conclusion

Across all our investigations, we deemed the mostly likely candidate gene to be the potential *MOCI* ortholog *TraesCS6B03G0494700* due to its close proximity to the narrow *QTh.mst-6B* locus, its having two exonic SNP variants between high and low tiller NIL groups, and its strong orthology to one of the most important tiller-regulating genes in rice. Future studies could more closely investigate the potential relationship between this gene and tiller number. Specifically, CRISPR knockout mutations could be introduced in the NIL tiller parents as they provide a nearly isogenic background and variable tiller phenotypes. Further, if new recombination events could be identified, the new LED grow lights could then allow for increased statistical power to detect differences in tillering phenotypes. This would allow for a

further fine mapping of the QTL region and provide more evidence for or against the causality of the *MOC1* locus.

References

- Awan MJA, Amin I, Rasheed A, et al (2024) Knockout mutation in TaD27 enhances number of productive tillers in hexaploid wheat. *Front Genome Ed* 6:. <https://doi.org/10.3389/fgeed.2024.1455761>
- Bates D, Mächler M, Bolker B, Walker S (2015) Fitting Linear Mixed-Effects Models Using **lme4**. *J Stat Soft* 67:. <https://doi.org/10.18637/jss.v067.i01>
- Bennett T, Sieberer T, Willett B, et al (2006) The Arabidopsis MAX Pathway Controls Shoot Branching by Regulating Auxin Transport. *Current Biology* 16:553–563. <https://doi.org/10.1016/j.cub.2006.01.058>
- Blake NK, Stougaard RN, Weaver DK, et al (2011) Identification of a quantitative trait locus for resistance to Sitodiplosis mosellana (Géhin), the orange wheat blossom midge, in spring wheat. *Plant Breeding* 130:25–30. <https://doi.org/10.1111/j.1439-0523.2010.01809.x>
- BOS H, NEUTEBOOM J (1998) Morphological Analysis of Leaf and Tiller Number Dynamics of Wheat (*Triticum aestivum*L.): Responses to Temperature and Light Intensity. *Annals of Botany* 81:131–139. <https://doi.org/10.1006/anbo.1997.0531>
- Bushnell B (2014) BBMap: A Fast, Accurate, Splice-Aware Aligner. Walnut Creek, CA
- Cao L, Li T, Geng S, et al (2023) TaSPL14-7A is a conserved regulator controlling plant architecture and yield traits in common wheat (*Triticum aestivum* L.). *Front Plant Sci* 14:. <https://doi.org/10.3389/fpls.2023.1178624>
- Cardon G, Höhmann S, Klein J, et al (1999) Molecular characterisation of the *Arabidopsis* SBP-box genes. *Gene* 237:91–104. [https://doi.org/10.1016/S0378-1119\(99\)00308-X](https://doi.org/10.1016/S0378-1119(99)00308-X)
- Cardon GH, Höhmann S, Nettlesheim K, et al (1997) Functional analysis of the *Arabidopsis thaliana* SBP-box gene SPL3: a novel gene involved in the floral transition. *The Plant Journal* 12:367–377. <https://doi.org/10.1046/j.1365-313X.1997.12020367.x>
- Casal JJ, Sanchez RA, Deregibus VA (1986) The effect of plant density on tillering: The involvement of R/FR ratio and the proportion of radiation intercepted per plant. *Environmental and Experimental Botany* 26:365–371. [https://doi.org/10.1016/0098-8472\(86\)90024-9](https://doi.org/10.1016/0098-8472(86)90024-9)

- Chen X, Zhang Z, Liu D, et al (2010) SQUAMOSA Promoter-Binding Protein-Like Transcription Factors: Star Players for Plant Growth and Development. *Journal of Integrative Plant Biology* 52:946–951. <https://doi.org/10.1111/j.1744-7909.2010.00987.x>
- Dixon LE, Greenwood JR, Bencivenga S, et al (2018) TEOSINTE BRANCHED1 Regulates Inflorescence Architecture and Development in Bread Wheat (*Triticum aestivum*). *Plant Cell* 30:563–581. <https://doi.org/10.1105/tpc.17.00961>
- Dobin A, Davis CA, Schlesinger F, et al (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29:15–21. <https://doi.org/10.1093/bioinformatics/bts635>
- Dong C, Zhang L, Zhang Q, et al (2023) Tiller Number1 encodes an ankyrin repeat protein that controls tillering in bread wheat. *Nat Commun* 14:836. <https://doi.org/10.1038/s41467-023-36271-z>
- EVERS JB, VOS J, ANDRIEU B, STRUIK PC (2006) Cessation of Tillering in Spring Wheat in Relation to Light Interception and Red : Far-red Ratio. *Ann Bot* 97:649–658. <https://doi.org/10.1093/aob/mcl020>
- Fang Z, Ji Y, Hu J, et al (2020) Strigolactones and Brassinosteroids Antagonistically Regulate the Stability of the D53–OsBZR1 Complex to Determine FC1 Expression in Rice Tillering. *Molecular Plant* 13:586–597. <https://doi.org/10.1016/j.molp.2019.12.005>
- Fox J, Weisberg S (2019) *An R Companion to Applied Regression, Third*. Sage, Thousand Oaks CA
- Greb T, Clarenz O, Schäfer E, et al (2003) Molecular analysis of the LATERAL SUPPRESSOR gene in *Arabidopsis* reveals a conserved control mechanism for axillary meristem formation. *Genes Dev* 17:1175–1187. <https://doi.org/10.1101/gad.260703>
- He G, Liu J, Dong H, Sun J (2019) The Blue-Light Receptor CRY1 Interacts with BZR1 and BIN2 to Modulate the Phosphorylation and Nuclear Function of BZR1 in Repressing BR Signaling in *Arabidopsis*. *Molecular Plant* 12:689–703. <https://doi.org/10.1016/j.molp.2019.02.001>
- Hothorn T, Bretz F, Westfall P (2008) Simultaneous Inference in General Parametric Models. *Biometrical Journal* 50:346–363
- Hou H, Li J, Gao M, et al (2013) Genomic Organization, Phylogenetic Comparison and Differential Expression of the SBP-Box Family Genes in Grape. *PLOS ONE* 8:e59358. <https://doi.org/10.1371/journal.pone.0059358>
- Hou S, Yu L, Guo W, et al (2026) Fine mapping and verification of a tiller inhibition gene in synthetic hexaploid wheat derivative. *Mol Breeding* 46:17. <https://doi.org/10.1007/s11032-026-01640-1>

- Hyles J, Vautrin S, Pettolino F, et al (2017) Repeat-length variation in a wheat cellulose synthase-like gene is associated with altered tiller number and stem cell wall composition. *J Exp Bot* 68:1519–1529. <https://doi.org/10.1093/jxb/erx051>
- Ikeda A, Ueguchi-Tanaka M, Sonoda Y, et al (2001) slender Rice, a Constitutive Gibberellin Response Mutant, Is Caused by a Null Mutation of the SLR1 Gene, an Ortholog of the Height-Regulating Gene GAI/RGA/RHT/D8. *Plant Cell* 13:999–1010. <https://doi.org/10.1105/tpc.13.5.999>
- Ignatiadis N, Klaus B, Zaugg JB, Huber W (2016) Data-driven hypothesis weighting increases detection power in genome-scale multiple testing. *Nat Methods* 13:577–580. <https://doi.org/10.1038/nmeth.3885>
- Jian C, Pan Y, Liu S, et al (2024) The TaGW2-TaSPL14 module regulates the trade-off between tiller number and grain weight in wheat. *Journal of Integrative Plant Biology* 66:1953–1965. <https://doi.org/10.1111/jipb.13723>
- Jiang DJ, Hao XH, Zhang J, et al (2021) Reducing expression of TaOTUB1s decreases tiller number in wheat. *Plant Signaling & Behavior* 16:2018217. <https://doi.org/10.1080/15592324.2021.2018217>
- Jiao Y, Wang Y, Xue D, et al (2010) Regulation of OsSPL14 by OsmiR156 defines ideal plant architecture in rice. *Nat Genet* 42:541–544. <https://doi.org/10.1038/ng.591>
- Jones BH (2020) Genetic dissection of grain yield and yield component traits in hexaploid spring wheat
- Jones BH, Blake NK, Heo H-Y, et al (2021) Allelic response of yield component traits to resource availability in spring wheat. *Theor Appl Genet* 134:603–620. <https://doi.org/10.1007/s00122-020-03717-7>
- Khan Y, Xiong Z, Zhang H, et al (2022) Expression and roles of GRAS gene family in plant growth, signal transduction, biotic and abiotic stress resistance and symbiosis formation—a review. *Plant Biology* 24:404–416. <https://doi.org/10.1111/plb.13364>
- Komatsu M, Maekawa M, Shimamoto K, Kyoizuka J (2001) The *LAX1* and *FRIZZY PANICLE 2* Genes Determine the Inflorescence Architecture of Rice by Controlling Rachis-Branch and Spikelet Development. *Developmental Biology* 231:364–373. <https://doi.org/10.1006/dbio.2000.9988>
- Krasileva KV, Vasquez-Gross HA, Howell T, et al (2017) Uncovering hidden variation in polyploid wheat. *Proceedings of the National Academy of Sciences* 114:E913–E921. <https://doi.org/10.1073/pnas.1619268114>

- Kuraparthi V, Sood S, Dhaliwal HS, et al (2007) Identification and mapping of a tiller inhibition gene (*tin3*) in wheat. *Theor Appl Genet* 114:285–294. <https://doi.org/10.1007/s00122-006-0431-y>
- Kuroda H, Takahashi N, Shimada H, et al (2002) Classification and Expression Analysis of Arabidopsis F-Box-Containing Protein Genes. *Plant Cell Physiol* 43:1073–1085. <https://doi.org/10.1093/pcp/pcf151>
- Lai X, Zhu Z, Zhang Y, et al (2025) The selection and application of tiller number QTLs in modern wheat breeding. *Theor Appl Genet* 138:124. <https://doi.org/10.1007/s00122-025-04908-w>
- Lenth RV (2025) emmeans: Estimated Marginal Means, aka Least-Squares Means
- Li X, Qian Q, Fu Z, et al (2003) Control of tillering in rice. *Nature* 422:618–621. <https://doi.org/10.1038/nature01518>
- Liao Z, Yu H, Duan J, et al (2019) SLR1 inhibits MOC1 degradation to coordinate tiller number and plant height in rice. *Nat Commun* 10:2738. <https://doi.org/10.1038/s41467-019-10667-2>
- Liu J, Cheng X, Liu P, Sun J (2017) miR156-Targeted SBP-Box Transcription Factors Interact with DWARF53 to Regulate TEOSINTE BRANCHED1 and BARREN STALK1 Expression in Bread Wheat. *Plant Physiol* 174:1931–1948. <https://doi.org/10.1104/pp.17.00445>
- Liu J, Wang T, Lan Y, et al (2025) Fine-mapping and candidate gene identification for *QPtn.sau-4B* showing potential in increasing productive tiller number and yield in wheat. *The Crop Journal* 13:480–489. <https://doi.org/10.1016/j.cj.2025.01.014>
- Lloveras J, Manent J, Viudas J, et al (2004) Seeding Rate Influence on Yield and Yield Components of Irrigated Winter Wheat in a Mediterranean Climate. *Agronomy Journal* 96:1258–1265. <https://doi.org/10.2134/agronj2004.1258>
- Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* 15:550. <https://doi.org/10.1186/s13059-014-0550-8>
- Lu Z, Shao G, Xiong J, et al (2015) *MONOCULM 3*, an Ortholog of *WUSCHEL* in Rice, Is Required for Tiller Bud Formation. *Journal of Genetics and Genomics* 42:71–78. <https://doi.org/10.1016/j.jgg.2014.12.005>
- Ma Z, Zhao D, Zhang C, et al (2007) Molecular genetic analysis of five spike-related traits in wheat using RIL and immortalized F2 populations. *Mol Genet Genomics* 277:31–42. <https://doi.org/10.1007/s00438-006-0166-0>

- Manning K, Tör M, Poole M, et al (2006) A naturally occurring epigenetic mutation in a gene encoding an SBP-box transcription factor inhibits tomato fruit ripening. *Nat Genet* 38:948–952. <https://doi.org/10.1038/ng1841>
- Mao Z, He S, Xu F, et al (2020) Photoexcited CRY1 and phyB interact directly with ARF6 and ARF8 to regulate their DNA-binding activity and auxin-induced hypocotyl elongation in *Arabidopsis*. *New Phytologist* 225:848–865. <https://doi.org/10.1111/nph.16194>
- Maqbool R, Nagarajan R, Mutti JS, Gill KS (2025) Wheat Tiller -1 (WT-1), a GRAS domain encoding gene, controls both tillering and spikelet number per spike in wheat. *Plant Mol Biol* 115:128. <https://doi.org/10.1007/s11103-025-01629-y>
- McKenna A, Hanna M, Banks E, et al (2010) The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 20:1297–1303. <https://doi.org/10.1101/gr.107524.110>
- Miura K, Ikeda M, Matsubara A, et al (2010) OsSPL14 promotes panicle branching and higher grain productivity in rice. *Nat Genet* 42:545–549. <https://doi.org/10.1038/ng.592>
- Morales N, Ogbonna AC, Ellerbrock BJ, et al (2022) Breedbase: a digital ecosystem for modern plant breeding. *G3 Genes|Genomes|Genetics* 12:jkac078. <https://doi.org/10.1093/g3journal/jkac078>
- Nagasawa N, Miyoshi M, Kitano H, et al (1996) Mutations associated with floral organ number in rice. *Planta* 198:627–633. <https://doi.org/10.1007/BF00262651>
- Naruoka Y, Talbert LE, Lanning SP, et al (2011) Identification of quantitative trait loci for productive tiller number and its relationship to agronomic traits in spring wheat. *Theor Appl Genet* 123:1043. <https://doi.org/10.1007/s00122-011-1646-0>
- Nasseer A m., Martin J m., Heo H y., et al (2016) Impact of a Quantitative Trait Locus for Tiller Number on Plasticity of Agronomic Traits in Spring Wheat. *Crop Science* 56:595–602. <https://doi.org/10.2135/cropsci2015.05.0325>
- Ogawa M, Kusano T, Katsumi M, Sano H (2000) Rice gibberellin-insensitive gene homolog, *OsGAI*, encodes a nuclear-localized protein capable of gene activation at transcriptional level. *Gene* 245:21–29. [https://doi.org/10.1016/S0378-1119\(00\)00018-4](https://doi.org/10.1016/S0378-1119(00)00018-4)
- Peltonen-Sainio P, Jauhiainen L, Sadras VO (2011) Phenotypic plasticity of yield and agronomic traits in cereals and rapeseed at high latitudes. *Field Crops Research* 124:261–269. <https://doi.org/10.1016/j.fcr.2011.06.016>
- Peng ZS, Yen C, Yang JL (1998) Genetic control of oligo-culms in common wheat [*Triticum aestivum*]. Wheat Information Service (Japan)

- Pysh LD, Wysocka-Diller JW, Camilleri C, et al (1999) The GRAS gene family in Arabidopsis: sequence characterization and basic expression analysis of the SCARECROW-LIKE genes. *The Plant Journal* 18:111–119. <https://doi.org/10.1046/j.1365-313X.1999.00431.x>
- Sadras VO, Rebetzke GJ (2013) Plasticity of wheat grain yield is associated with plasticity of ear number. *cpsc* 64:234–243. <https://doi.org/10.1071/CP13117>
- Schoen A, Yadav I, Wu S, et al (2023) Identification and high-resolution mapping of a novel tiller number gene (*tin6*) by combining forward genetics screen and MutMap approach in bread wheat. *Funct Integr Genomics* 23:157. <https://doi.org/10.1007/s10142-023-01084-2>
- Schumacher K, Schmitt T, Rossberg M, et al (1999) The Lateral suppressor (*Ls*) gene of tomato encodes a new member of the VHIID protein family. *Proceedings of the National Academy of Sciences* 96:290–295. <https://doi.org/10.1073/pnas.96.1.290>
- Shang Q, Wang Y, Tang H, et al (2021) Genetic, hormonal, and environmental control of tillering in wheat. *The Crop Journal* 9:986–991. <https://doi.org/10.1016/j.cj.2021.03.002>
- Shao G, Lu Z, Xiong J, et al (2019) Tiller Bud Formation Regulators MOC1 and MOC3 Cooperatively Promote Tiller Bud Outgrowth by Activating *FON1* Expression in Rice. *Molecular Plant* 12:1090–1102. <https://doi.org/10.1016/j.molp.2019.04.008>
- Si Y, Lu Q, Tian S, et al (2022) Fine mapping of the tiller inhibition gene *TIN5* in *Triticum urartu*. *Theor Appl Genet* 135:2665–2673. <https://doi.org/10.1007/s00122-022-04140-w>
- Slafer GA, Savin R, Sadras VO (2014) Coarse and fine regulation of wheat yield components in response to genotype and environment. *Field Crops Research* 157:71–83. <https://doi.org/10.1016/j.fcr.2013.12.004>
- Souvorov A, Kapustin Y, Kiryutin B, et al (2010) Gnomon: NCBI Eukaryotic Gene Prediction Tool. National Center for Biotechnology Information, Bethesda, MD
- Spielmeyer W, Richards RA (2004) Comparative mapping of wheat chromosome 1AS which contains the tiller inhibition gene (*tin*) with rice chromosome 5S. *Theor Appl Genet* 109:1303–1310. <https://doi.org/10.1007/s00122-004-1745-2>
- Stirnberg P, van de Sande K, Leyser HMO (2002) *MAX1* and *MAX2* control shoot lateral branching in Arabidopsis. *Development* 129:1131–1141. <https://doi.org/10.1242/dev.129.5.1131>
- Tabuchi H, Zhang Y, Hattori S, et al (2011) *LAX PANICLE2* of Rice Encodes a Novel Nuclear Protein and Regulates the Formation of Axillary Meristems. *Plant Cell* 23:3276–3287. <https://doi.org/10.1105/tpc.111.088765>

- The International Wheat Genome Sequencing Consortium (IWGSC), Appels R, Eversole K, et al (2018) Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* 361:eaar7191. <https://doi.org/10.1126/science.aar7191>
- Tokatlidis I s. (2014) Addressing the yield by density interaction is a prerequisite to bridge the yield gap of rain-fed wheat. *Annals of Applied Biology* 165:27–42. <https://doi.org/10.1111/aab.12121>
- Tripathi RK, Goel R, Kumari S, Dahuja A (2017) Genomic organization, phylogenetic comparison, and expression profiles of the SPL family genes and their regulation in soybean. *Dev Genes Evol* 227:101–119. <https://doi.org/10.1007/s00427-017-0574-7>
- Ugarte CC, Trupkin SA, Ghiglione H, et al (2010) Low red/far-red ratios delay spike and stem growth in wheat. *J Exp Bot* 61:3151–3162. <https://doi.org/10.1093/jxb/erq140>
- Van der Auwera GA, O'Connor BD (2020) *Genomics in the Cloud: Using Docker, GATK, and WDL in Terra*, 1st edn.
- Wang S, Wong D, Forrest K, et al (2014) Characterization of polyploid wheat genomic diversity using a high-density 90 000 single nucleotide polymorphism array. *Plant Biotechnology Journal* 12:787–796. <https://doi.org/10.1111/pbi.12183>
- Wang W, Lu X, Li L, et al (2018) Photoexcited CRYPTOCHROME1 Interacts with Dephosphorylated BES1 to Regulate Brassinosteroid Signaling and Photomorphogenesis in Arabidopsis. *Plant Cell* 30:1989–2005. <https://doi.org/10.1105/tpc.17.00994>
- Wang Z, Wu F, Chen X, et al (2022) Fine mapping of the tiller inhibition gene TIN4 contributing to ideal plant architecture in common wheat. *Theor Appl Genet* 135:527–535. <https://doi.org/10.1007/s00122-021-03981-1>
- Xie Q, Mayes S, Sparkes DL (2016) Optimizing tiller production and survival for grain yield improvement in a bread wheat × spelt mapping population. *Ann Bot* 117:51–66. <https://doi.org/10.1093/aob/mcv147>
- Xu C, Wang Y, Yu Y, et al (2012) Degradation of MONOCULM 1 by APC/CTAD1 regulates rice tillering. *Nat Commun* 3:750. <https://doi.org/10.1038/ncomms1743>
- Xu F, He S, Zhang J, et al (2018) Photoactivated CRY1 and phyB Interact Directly with AUX/IAA Proteins to Inhibit Auxin Signaling in Arabidopsis. *Molecular Plant* 11:523–541. <https://doi.org/10.1016/j.molp.2017.12.003>
- Yan B, Yang Z, He G, et al (2021) The blue light receptor CRY1 interacts with GID1 and DELLA proteins to repress gibberellin signaling and plant growth. *Plant Communications* 2:100245. <https://doi.org/10.1016/j.xplc.2021.100245>

- Yan Y, Ding C, Zhang G, et al (2023) Genetic and environmental control of rice tillering. *The Crop Journal* 11:1287–1302. <https://doi.org/10.1016/j.cj.2023.05.009>
- Yang Z, Kan W, Wang Z, et al (2025) Genome-wide identification and expression analysis of phytochrome gene family in Aikang58 wheat (*Triticum aestivum* L.). *Front Plant Sci* 15:. <https://doi.org/10.3389/fpls.2024.1520457>
- Yao E, Blake VC, Sharma PK, et al (2025) GrainGenes: genetics, genomes, and pangenomes. *GENETICS* iyaf270. <https://doi.org/10.1093/genetics/iyaf270>
- Yoshida S (1972) Physiological Aspects of Grain Yield. *Annual Review of Plant Biology* 23:437–464. <https://doi.org/10.1146/annurev.pp.23.060172.002253>
- Zhang B, Liu X, Xu W, et al (2015) Novel function of a putative MOC1 ortholog associated with spikelet number per spike in common wheat. *Sci Rep* 5:12211. <https://doi.org/10.1038/srep12211>
- Zhang J, Wu J, Liu W, et al (2013) Genetic mapping of a fertile tiller inhibition gene, *ftin*, in wheat. *Mol Breeding* 31:441–449. <https://doi.org/10.1007/s11032-012-9801-0>
- Zhang N, Liu Y, Gui S, Wang Y (2025) Regulation of tillering and panicle branching in rice and wheat. *Journal of Genetics and Genomics* 52:869–886. <https://doi.org/10.1016/j.jgg.2024.12.005>
- Zhao B, Wu TT, Ma SS, et al (2020) TaD27-B gene controls the tiller number in hexaploid wheat. *Plant Biotechnology Journal* 18:513–525. <https://doi.org/10.1111/pbi.13220>
- Zhu T, Liu Y, Ma L, et al (2020) Genome-wide identification, phylogeny and expression analysis of the SPL gene family in wheat. *BMC Plant Biol* 20:420. <https://doi.org/10.1186/s12870-020-02576-0>
- Zhu T, Wang L, Rimbart H, et al (2021) Optical maps refine the bread wheat *Triticum aestivum* cv. Chinese Spring genome assembly. *The Plant Journal* 107:303–314. <https://doi.org/10.1111/tpj.15289>
- Zhuang H, Guo Z, Wang J, Chen T (2024) Genome-wide identification and comprehensive analysis of the phytochrome-interacting factor (PIF) gene family in wheat. *PLOS ONE* 19:e0296269. <https://doi.org/10.1371/journal.pone.0296269>

CHAPTER THREE

EVALUATING REACTION NORM AND FIXED EFFECT MODELS FOR GENOMIC SELECTION OF END-USE QUALITY IN WHEAT

Contribution of Authors and Co-Authors

Manuscript in Chapter 3

Author: Jared Lile

Contributions: Formal data analysis, including designing and validating prediction models, figure design and visualization, writing the original draft and editing the manuscript.

Co-Author: Jennifer Lachowicz

Contributions: Provided analysis methodological review and supervision, and contributed to manuscript writing and revision.

Co-Author: Jason D. Fiedler

Contributions: Curated all genetic data for the study and contributed to manuscript writing and review

Co-Author: Jason P. Cook

Contributions: Secured research funding, curated all phenotypic data in this study, provided analysis review and supervision, and contributed to manuscript review and writing.

Manuscript Information

Jared Lile, Jennifer Lachowiec, Jason D. Fiedler, Jason P. Cook

Status of Manuscript: [Double click to put an X in one of the options below, then delete instruction in brackets. If you cannot double click to add an X, search for “Check Box Content Control.”]

- ☐ Prepared for submission to a peer-reviewed journal
- ☒ Officially submitted to a peer-reviewed journal
- ☐ Accepted by a peer-reviewed journal
- ☐ Published in a peer-reviewed journal

Submitted manuscript – 02/07/2026

Abstract

Exceptional end-use quality in hexaploid bread wheat is an important factor for selling wheat in the global marketplace. However, many end-use quality traits are time-consuming and expensive to test, meaning their measurement is often delayed until late-stage trials of a breeding program. Genomic selection (GS) is one tool that has been shown to allow for earlier generation improvement of many traits in wheat breeding. Further, advances in GS methods may improve prediction accuracies, such as incorporating known loci that control traits as fixed effects and utilizing single-stage reaction norm models to control for genotype by environment interactions. Therefore, we evaluated conventional methods as well as these two advances to perform GS for twelve important end-use quality traits collected by the Montana State University (MSU) Spring Wheat Breeding Program from 2010-2022 across 842 elite breeding lines. We found that the conventional GS models performed well in 5-fold cross validation with prediction accuracies ranging from 0.4-0.7 and in leave-one-year-out validation with most trait-year combinations exceeding an accuracy threshold of 0.3. We did not observe any difference in prediction accuracy for the inclusion of fixed effect loci, but we did observe strong within-year correlations for the single-step reaction norm models in both validation methods. We concluded that both conventional, two-step GS and single-step reaction norm GS offer viable methods for earlier generation improvement of major end-use quality traits in the MSU Spring Wheat Breeding Program.

Introduction

Genomic selection (GS) is a modern breeding technology that has seen wide and diverse adoption over the last decade in both plant and animal breeding (Crossa et al. 2017a; Alemu et al.

2024). One primary driver for this adoption is the ability of GS to increase genetic gain over time for many different traits. This is achieved by the incorporation of genome-wide information that can simultaneously model both large and small effect loci, leading to a more comprehensive prediction of trait performance compared to single locus methods like marker assisted selection (MAS), especially for low heritability traits (Bernardo and Yu 2007a). Additionally, there have been many recent developments to improve GS models that enable prediction with more complicated large multi-environmental datasets.

In hexaploid bread wheat (*Triticum aestivum*), GS has been utilized for numerous traits including yield, grain protein content, and disease resistance (Bonnett et al. 2022; Michel et al. 2019; Steiner et al. 2019; Merrick et al. 2022). One of the major classes of wheat grown in the United States is hard red spring wheat (HRSW) which accounted for 25.5% of the total US wheat production in 2024 (USDA NASS 2024). Hard red wheat, in general, is known for its high grain protein content, gluten strength, and desirable end-use quality which influences its market price (Roberts et al. 2022; Subedi et al. 2023). End-use quality can be broadly split into three categories: milling traits, dough traits, and baking traits. Milling traits refer to the properties and composition of a grain sample and its derived flour, while dough traits and baking traits refer to the properties of the dough and final baked products derived from a flour sample.

End-use quality traits are ideal targets for genomic selection because they are often highly heritable (Souza et al. 2012; Laidig et al. 2017), though also polygenic (Battenfield et al. 2018; Gill et al. 2025b). This means end-use quality traits are consistently controlled by genetics, but there are likely too many important loci for a single locus method like MAS to encompass, indicating that a genome-wide method like GS is needed instead. In addition, end-use quality

traits are typically measured late in a breeding cycle due to the labor, complexity of tests, and large grain samples required for the measurements (Battenfield et al. 2016). Genomic selection enables end-use quality improvement at earlier stages of the breeding pipeline, increasing the rate of genetic gain over time.

Multiple studies have investigated the use of GS for trait improvement in different wheat classes, leading to several key advances to make better predictions. Two of these advances include incorporating known QTL markers as fixed effects, (Zhao et al. 2014; Michel et al. 2018; Gao et al. 2025) and modeling of genotype by environment (GE) interactions (Burgueño et al. 2012; Sandhu et al. 2022; Brault et al. 2025), though few studies have incorporated these advances into end-use quality prediction.

End-use quality traits are known to exhibit varying levels of GE (Prasad et al. 2003; Aoun et al. 2021; Shahi et al. 2024) which can lower prediction accuracy (Burgueño et al. 2012). One method to overcome this is the single-step reaction norm model described in Jarquín et al. (2014) which is an extension of the GBLUP model and performs genomic prediction in one step using the raw phenotypic data as a response. It also models GE interactions between additive genetic values and environmental covariates using prebuilt covariance matrices (Jarquín et al. 2014; Crossa et al. 2017a; Persa et al. 2020). This single-step method directly predicts phenotypic values within a specific environment, and it has been successfully utilized in genomic selection for grain yield in wheat (Jarquín et al. 2017; Pérez-Rodríguez et al. 2017).

In this study, we conduct genomic selection for twelve end-use quality traits in the HRSW market class using twelve years of historic quality data from the Montana State University Spring Wheat Breeding Program. We examine two advances in GS: the incorporation

of known fixed effect QTL and the modeling of GE interactions with single-step reaction norm models. We utilize two validation schemes (k-fold and leave-one-year-out) to investigate the effectiveness of these methods on their ability to perform genomic selection for end-use quality traits in hard red spring wheat.

Materials and Methods

Phenotypic Data Collection

End-use quality data was obtained from the Montana State University Spring Wheat Breeding Program yield trials conducted from the years 2010-2022. Trials were conducted at multiple locations each year, and within each location, two trials consisting of the preliminary yield trial (PYT) and advanced yield trial (AYT) were conducted with different sets of experimental lines and a set of common check varieties (Supplementary Figure S2.1). The check varieties also varied between years with some checks consistently included and others replaced overtime. Each year, a subset of lines from the PYT were selected to advance to the AYT, resulting in an unbalanced dataset with better performing lines being evaluated in more location-years. Different experimental designs, primarily lattice and randomized complete block, were used over the tested years. One replicate for each line from a trial was submitted for quality testing with one check variety, Duclair, having two replicates for consistency. In total, 842 lines were phenotyped, including 24 check varieties and 818 elite experimental lines from the program. Notably, experimental lines from the 2021 PYT were phenotyped but not genotyped, so that trial only had training data for the check varieties (Supplementary Figure S2.1).

End-use quality data was collected by the Montana State University Cereal Quality Laboratory with protocols following the American Association of Cereal Chemists methods (AACC, 2000). In brief, moisture and grain protein content were first collected on a cleaned grain sample using a Foss Infratec 1241 Near Infrared Reflectance (NIR) instrument (Infratech 1241 Grain Analyzer; Foss North America) calibrated via AACC Method 46-30 (AACC, 2000). The sample was then tempered to 15.5% moisture and milled using a Brabender Quadrumat Sr. Mill (C.W. Brabender Instruments Inc.) with flour yield describing the percentage of pure flour compared to total milled product. After two weeks of stabilization, flour protein was measured using the same Infratec 1241 instrument. Total ash content (wheat ash) and flour ash content were also collected following AACC Method 08-01 adjusted to 14% moisture content (AACC, 2000). Dough properties including mix time, mix water absorption, and mix tolerance were collected using a Mixograph following AACC Method 54-40A with a 10g sample corrected to a 14% moisture basis (AACC, 2000). Finally, bake quality traits including bake mixing time, bake water absorption, and loaf volume were collected following AACC Method 10-10B (AACC, 2000). Single kernel hardness was measured separately via a single-kernel characterization system (SKCS) (Perten Instruments, Springfield, IL, U.S.A.).

Broad-sense heritability was calculated for each end-use quality trait using the Cullis method with the following formula:

$$H_{Cullis}^2 = 1 - \frac{\bar{v}_{\Delta}^{BLUP}}{2 * \sigma_g^2} \quad (2.1)$$

Where σ_g^2 is the genotypic variance and $\bar{v}_{\Delta}^{BLUP}$ is the mean variance of a difference between two genotypic Best Linear Unbiased Predictors (BLUPs) (Cullis et al. 2006). The Cullis method was

selected due its performance in unbalanced datasets (Schmidt et al. 2019) which was the case in this study as the check varieties had many more observations than the elite lines.

Genetic Data Collection

Genotyping-by-sequencing (GBS) was performed on the 842 phenotyped lines at the USDA-ARS Genotyping Lab (Fargo, North Dakota) following a protocol adapted from Poland et al. (2012), which uses the double restrictions enzymes, *Pst*I and *Mse*I. GBS libraries were sequenced with Illumina NextSeq 2000 system with 150 bp paired ends (Illumina, San Diego, USA). Sequence data were quality filtered and trimmed for adapters with *bbduk* (BBTools, <https://jgi.doe.gov/data-and-tools/bbtools>). The trimmed paired-end sequences were aligned to the Chinese Spring wheat IWGSC RefSeq v1.0 reference genome (IWGSC et al. 2018) using Bowtie2 (Langmead and Salzberg 2012). The aligned reads were then sorted and indexed using SAMtools (Danecek et al. 2011, 2021a), and vcf files were generated with mapping quality score of at least 30, using BCFtools (Danecek et al. 2021a). The variant file was further filtered using vcftools, with a max-missing of 50%, a minor allele frequency (MAF) of 0.01, and a minimum depth coverage of 4. Chromosome UN was removed, and missing data was imputed using Beagle v5.4 (Browning et al. 2018a) using default settings, no reference panel, and the Synthetic W7984 x Opata M85 RIL 906 genetic map (Gutierrez-Gonzalez et al. 2019) to inform recombination frequencies.

In total, there were 21,958 SNPs after filtering and imputation. All SNPs were used for all genomic selection models. For genome wide association studies (GWAS), SNPs with a minor allele frequency (MAF) less than 0.01 were removed.

Two-Step Analysis Method

Two methods of genomic prediction were utilized in this study. The first method, hereafter the “two-step method” (Garrick et al. 2009), is an approach where the analysis is split into two separate steps. In the first step, summary values for each phenotyped line are calculated across all environments. Then, in the second step, the summary values are used as the response variable in a new model with the genetic markers to calculate genomic estimated breeding values (GEBVs). In this study, best linear unbiased estimators (BLUEs) (Holland and Piepho 2024) were calculated for each line from the following model:

$$Trait_{ijk} = \mu + Line_i + Environment_j + Line * Environment_{ij} + \varepsilon_{ijk} \quad (2.2)$$

Where $Trait_{ijk}$ is the k^{th} observation of an end-use quality trait measured on the i^{th} line in the j^{th} environment where an environment denotes a single combination of a year, location, and trial (AYT or PYT) within a location. $Line_i$ is fixed effect of the i^{th} line, $Environment_j$ is the random effect of j^{th} environment following a $\mathcal{N}(0, \sigma_{Env}^2)$ distribution, $Line * Environment_{ij}$ is the random interaction of the i^{th} line with the j^{th} environment following a $\mathcal{N}(0, \sigma_{Line * Env}^2)$ distribution, and ε_{ijk} is the random error term following $\mathcal{N}(0, \sigma^2)$. These models were fit using the lme4 package (Bates et al. 2015), and BLUEs were extracted using the emmeans package (Lenth 2024) in the data analysis software R (R Core Team 2024).

The second step of the analysis was to fit standard genomic selection models including RR-BLUP with the rrBLUP R package (Endelman 2011) and Bayesian Ridge Regression (BRR), Bayesian LASSO (B.Lasso), BayesA, BayesB, and BayesC models using the BGLR R package (Pérez and de los Campos 2014). Reproducing kernel Hilbert space (RKHS) models were also fit with the BGLR package using a Gaussian kernel computed as follows:

$$\mathbf{K}(x_i, x_j) = \exp\left(-h\left(\frac{d_{ij}}{k}\right)\right) \quad (2.3)$$

where d_{ij} is the squared Euclidean distance based on genetic markers between individuals i and j , k is the mean squared Euclidean distance across all individuals, and h is a bandwidth parameter set equal to one. All Bayesian models were fit using 10,000 iterations, with the first 1,000 used as burn-in and the default thinning value of 5.

GWAS

Genome wide association studies (GWAS) were conducted for each end-use quality trait using GAPIT version 3 (Wang and Zhang 2021) with the BLINK model (Huang et al. 2019). A minor allele frequency threshold of 0.01 for SNP inclusion was used for all GWAS, as well as two principal components to control for population structure. Significant markers for each trait were determined using the default Bonferroni correction thresholds (Wang and Zhang 2021) (Supplementary Table S2.1). GWAS-informed genomic selection models (Bernardo 2014) were identical to the models described above except that the significant GWAS markers for a given trait were removed from the random effect marker matrix and treated instead as unique fixed effect variables. During cross-validation, GWAS was performed using only the training data in each fold to avoid so called “inside trading” effect of using fixed GWAS markers identified with the full phenotypic dataset (Arruda et al. 2016). To illustrate this effect, separate models were tested with the inclusion of GWAS markers identified on the full dataset.

Known QTL as Fixed Effects

Genes corresponding to major wheat seed proteins including the glutenin, gliadin, puroindoline, purothionin, avenin-like, purinin, and triticin families, as well as amylase/trypsin

inhibitors, were considered potential known loci that could influence end-use quality (Li et al. 2020). Markers were first filtered to those that fell within 1 Mb of any known seed protein genes, and in cases where multiple markers were close to a gene, the marker with the highest minor allele frequency was selected. Selected markers were treated as fixed effects in two-step GS models in the same manner as the GWAS-informed models above.

In addition to selecting markers based on minor allele frequency, we also selected all genetic markers within 1 Mb of any known seed protein gene and performed principal component analysis on the resulting marker matrix. Two-step GS models were then cross-validated with the inclusion of the first 1, 3, 20, or 50 principal components as fixed effects, with the selected markers removed from the random effect marker matrix as before.

Single-Step Analysis Method

The second method of genomic prediction used in this study is hereafter referred to as the “single-step” method, which combines the raw phenotypic data and genetic marker data together into one prediction model. These models were based on the reaction norm models described in Jarquín et al. (2014) and can be generally written as follows:

$$y_{ijk} = \mu + E_i + L_j + EL_{ij} + \epsilon_{ijk} \quad (2.4)$$

where y_{ijk} is the k^{th} observation of an end-use quality trait measured on the j^{th} line in the i^{th} environment where, as before, an environment denotes a combination of a year, location, and trial within a location (AYT or PYT). μ is the overall trait mean, E_i is the fixed effect of the i^{th} environment, and L_j is the random effect of the j^{th} line with $\mathbf{L} \sim \mathcal{N}(0, \mathbf{Z}_L \mathbf{K} \mathbf{Z}_L' \sigma_{Line}^2)$. EL_{ij} is the interaction between the i^{th} environment and the j^{th} line with $\mathbf{EL} \sim \mathcal{N}(0, [\mathbf{Z}_L \mathbf{K} \mathbf{Z}_L'] \circ [\mathbf{Z}_E \mathbf{Z}_E'] \sigma_{EL}^2)$ where $\circ$ is the Hadamard product, $\mathbf{Z}_L$ is an $(n \times j)$ incidence matrix for phenotypic observations

of each line, $\mathbf{Z}_E$ is an $(n \times i)$ incidence matrix relating observations to an environment, $\mathbf{K}$ is either the $(j \times j)$ Gaussian kernel matrix described in equation 2.3 for the RKHS models, or for all other genomic selection models, $\mathbf{K}$ is the $(j \times j)$ genomic relationship $\mathbf{G}$ matrix calculated using the *a.mat* function from the rrBLUP R package (Endelman 2011) using the GBS markers. Finally, ϵ_{ijk} is the random error term following $\epsilon_{ijk} \sim i.i.d. \mathcal{N}(0, \sigma^2)$. Code to implement these models was based on the framework provided in Montesinos-López et al. 2021.

Cross-Validation

Two methods of cross-validation were used to assess the predictive ability of tested models. The first method was a five-fold cross-validation which differed slightly in implementation between the two-step and single-step methods. In the two-step method, the line BLUEs were randomly divided into five equal folds and each fold was sequentially removed and then predicted with all of the candidate GS models. Pearson's correlation between the predicted and observed values within a fold was measured and reported as the prediction accuracy. The root mean squared error (RMSE) and mean average error (MAE) were also recorded (Supplementary Table S2.2). Twenty allocations of five folds were created for a total of 100 cycles of cross-validation for each model. The mean correlation, RMSE, and MAE were taken across all cycles as well as the standard deviation of the correlations to give a sense of the range of accuracies between cycles.

For five-fold cross-validation of the single-step method, the raw phenotypic data points were randomly assigned into five equal folds. One fold was then removed and the candidate model was used to directly predict the removed phenotypic values. The correlation, RMSE, and MAE were taken both at the location-year level and across all location-years (Supplementary

Table S2.4). Both methods were then averaged across 100 cycles of cross-validation for all metrics, and the standard deviation of the correlation values was also taken.

The second method of cross-validation was a leave-one-year-out (LOYO) validation. For the two-step approach, all non-check lines phenotyped in a given year were identified and their BLUEs were removed. The remaining lines were then used to predict GEBVs for the removed lines, and the correlation, RMSE, and MAE were calculated (Supplementary Table S2.3). The removal of the most recent year in the dataset (2022) was technically a forward prediction as all previous years of data were used to predict the newest year. Similar methods were used for the single-step approach, but specifically the phenotypic values of non-check lines within a year were removed, and the metrics correlation, RMSE, and MAE were taken both at the location-trial level within a year and across all locations (Supplementary Table S2.5). Check line phenotypes were retained within a year as the check line end-use quality data is always available to the breeding program before genomic predictions must be made.

Results

Predictive ability in genomic selection (GS) depends on, among other factors, the heritability of the predicted trait (Bernardo 2020). Broad-sense heritability for the twelve end-use quality traits tested in this study was determined (Figure 2.1a) using the Cullis method (Cullis et al. 2006). Wheat ash content had the lowest heritability with a value of 0.39, with all other traits exceeding a broad-sense heritability of 0.60. Pairwise Pearson correlation coefficients between the BLUEs of all 12 end-use quality traits were also calculated (Figure 2.1b). The pairs of traits with high correlation (above 0.85) were flour protein content and grain protein content, bake water absorption and mix water absorption, and bake time and mix time. Grain protein content is

commonly used as an indicator proxy trait for overall end-use quality but was only moderately, positively correlated with four out of the eleven traits and strongly correlated only with flour protein content

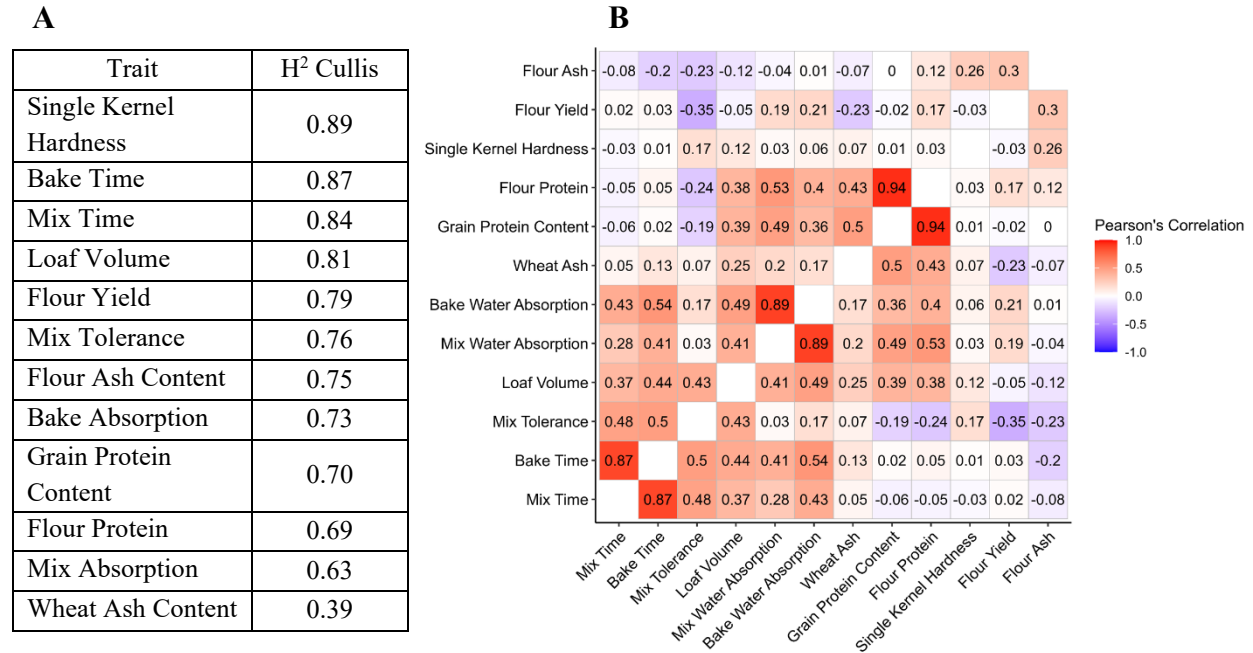


Figure 2.1. A) Table of broad-sense heritabilities for twelve end-use quality traits calculated using the Cullis method. B) Pearson correlation matrix between the BLUEs of the same twelve end-use quality traits with darker red indicating a stronger, positive correlation and darker blue indicating a stronger, negative correlation.

Two-Step Analysis Results

Seven genomic selection models were tested for each end-use quality trait as the second stage of the two-step analysis. One hundred cycles of five-fold cross-validation for each model within each quality trait were performed (Figure 2.2). The mean accuracy across all cycles ranged from 0.4 to 0.7 with standard deviations between 0.04 to 0.06. There was no appreciable difference between the mean prediction accuracies of the different genomic selection models within each trait. However, prediction accuracy did vary across traits with mix tolerance and

flour yield having the highest accuracy on average, and the trait with the lowest heritability, wheat ash content, having the lowest accuracy.

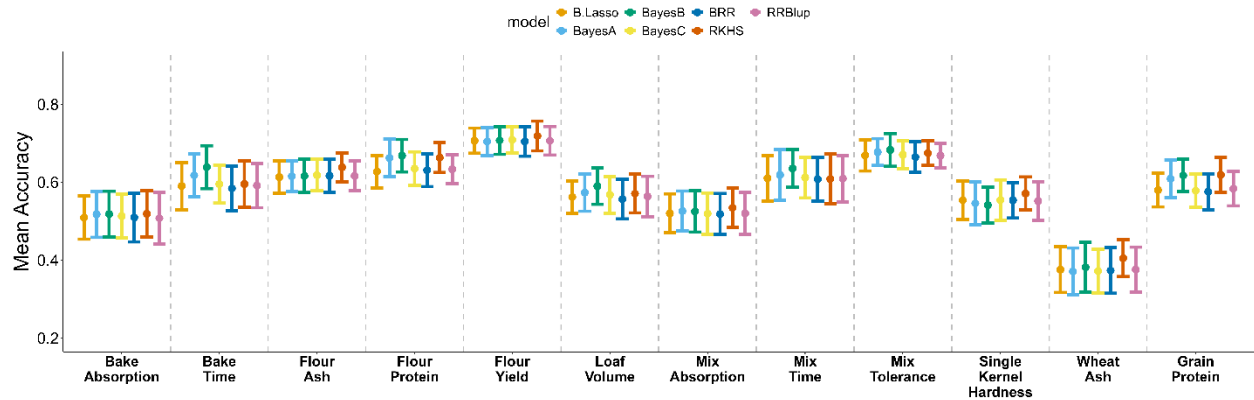


Figure 2.2. Five-fold cross-validation results ($n=100$) for twelve end-use quality traits. The mean prediction accuracy is given on the y-axis and error bars denote one standard deviation. Colors represent results from one of seven different genomic selections that were tested.

Given that most of the end-use quality traits had a high broad-sense heritability (Figure 2.1a) and there are known major genes controlling them (Eagles et al. 2002), we next investigated the effect of removing certain markers from the random effect marker matrix and treating them as fixed effects to exert a stronger influence on trait predictions. Using the BRR genomic selection model within the two-step method framework for consistency, we found no difference in the average prediction accuracy, across 100 cycles of five-fold cross-validation, between the baseline model with no added fixed effects compared to models with included fixed effect markers (Figure 2.3). The only case where an increase in mean prediction accuracy was observed, around one standard deviation higher for all traits, was when we included fixed effect markers identified by a GWAS analysis of the full phenotypic BLUES dataset (Figure 2.3). This

method problematically included both training and validation sets of data, which led to data leakage and inflated accuracies (Supplementary Table S2.1).

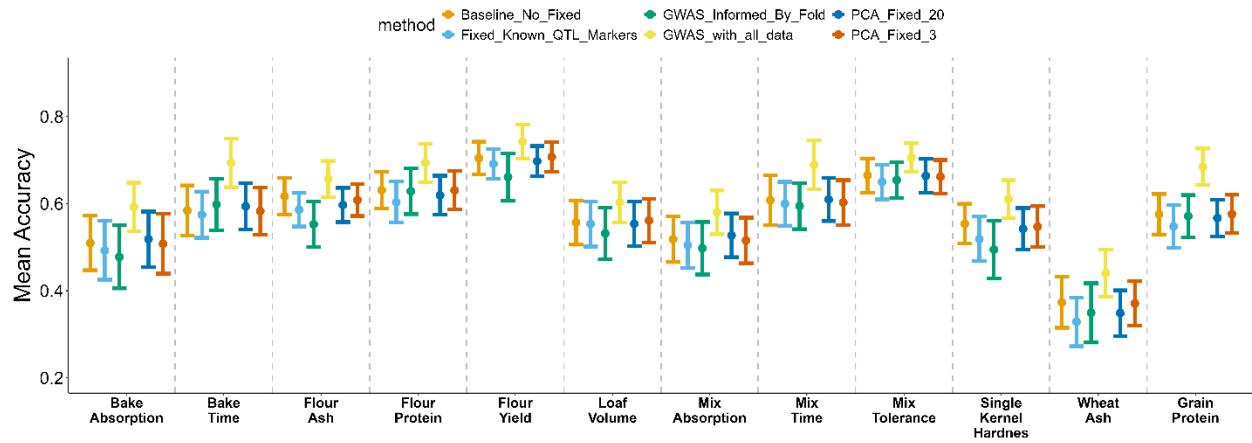


Figure 2.3. Five-fold cross-validation results (n=100) for the two-step method showing the mean prediction accuracy for twelve different end-use quality traits. The BRR genomic selection model was used in all cases with different colors representing different fixed effect specifications. The first, orange bars are the baseline model with no added fixed effect terms which are compared against the inclusion of markers associated to potential known QTL, markers identified by GWAS on each training fold, markers identified by GWAS on the full phenotypic BLUEs dataset, and 20 or 3 principal components of a PCA across all markers associated with a potential known QTL.

Leave-One-Year-Out (LOYO) validation was performed for the same seven GS models as the five-fold cross-validation (Figure 2.4). A cutoff of 0.3 was used to delineate a useful prediction versus an uninformative or even detrimental one (Heffner et al. 2010; Marulanda et al. 2021). The traits bake time, loaf volume, mix time, mix tolerance, and single kernel hardness had prediction accuracies above the 0.3 cutoff for all 13 years. Other traits varied in accuracy depending on the year left out, and wheat ash content was the only trait that consistently failed to meet the cutoff for most years. For all traits, performance of different models was nearly identical, though notably, in the forward prediction year 2022, the traits grain protein content and

flour protein content had discernably higher prediction accuracies for the RKHS model compared to all other models.

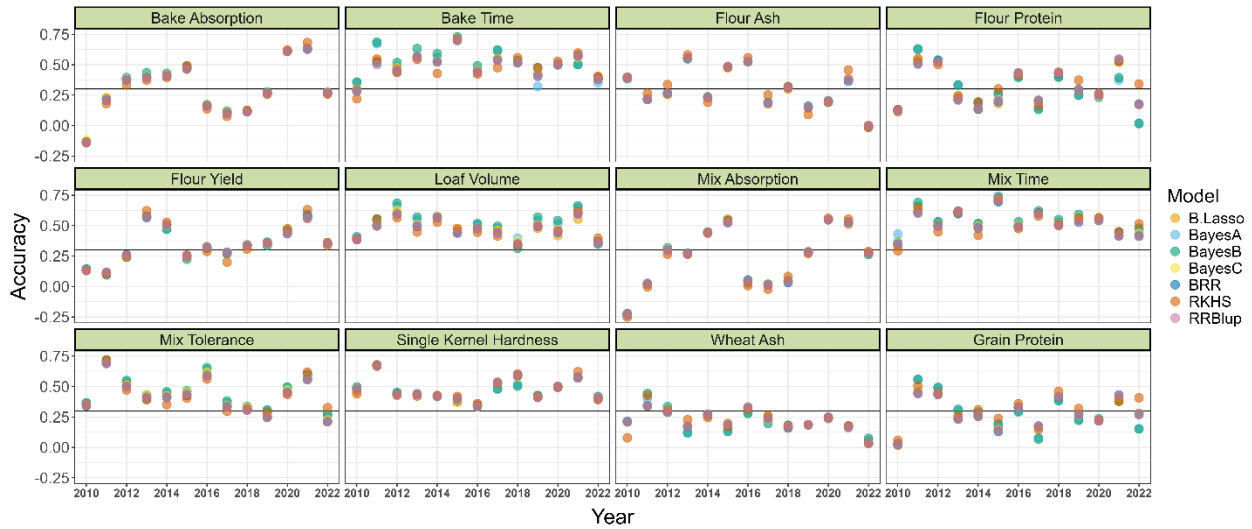


Figure 2.4. Leave-one-year-out (LOYO) validation results for the two-step analysis approach. The x-axis denotes the year in which the experimental lines were left out of the training data and predicted. The y-axis shows the accuracy as the correlation between the predicted breeding values and known BLUES for the experimental lines in the left-out year. Each box represents one end-use quality trait and for each trait, results are given for seven different GS models denoted with different colors.

Single-Step Analysis Results

Five-fold cross-validation for the same six GS models (without RR-BLUP, see Methods) was also performed with the single-step analysis method (Figure 2.5). Prediction accuracy in this context refers to the correlation between observed phenotypic values and predicted phenotypic values within a location-year combination across the preliminary and advanced yield trials conducted at that site. The mean prediction accuracy refers to the average correlation within a location-year across 100 cycles of cross-validation. There was little discernible difference between the different models tested. The traits with the highest median of the means accuracy

across all models and location-years were bake time, flour yield, mix time, and single kernel hardness with values at nearly 0.8, while the trait with the lowest median accuracy was wheat ash content at a little under 0.5. Interestingly, these results corresponded almost exactly to the broad-sense heritabilities (Figure 2.1a) in terms of the relative trait rankings.

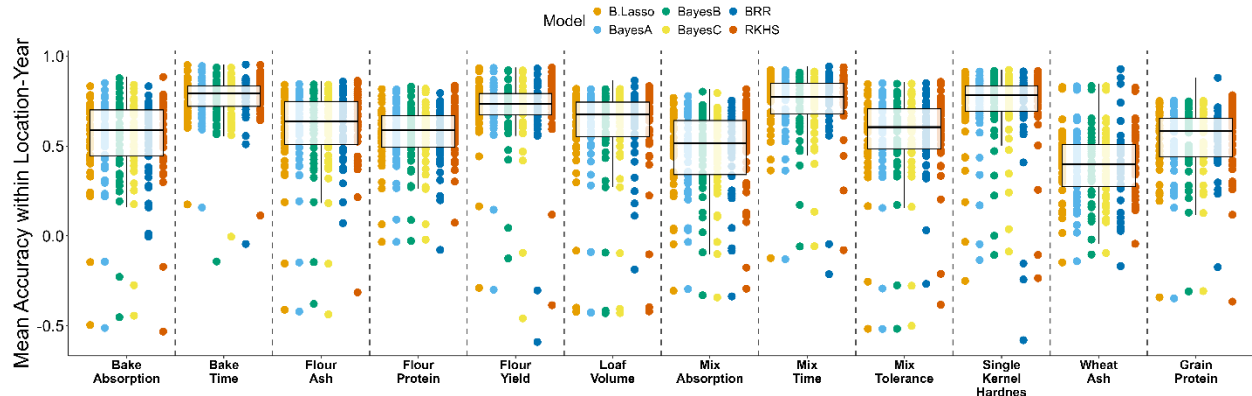


Figure 2.5. Five-fold cross-validation results ($n=100$) for the single-step analysis method for twelve end-use quality traits. The mean prediction accuracy across all cycles of cross-validation, within a location-year, represented by each point. The x-axis denotes the quality trait tested and each color represents a different genomic selection model. Box plots show the overall median of the means prediction accuracy for each trait across all location-years and models.

Leave-One-Year-Out (LOYO) validation was conducted on the single-step analysis method in the same manner as the two-step except that prediction accuracy was taken as the correlation between predicted and observed phenotypic values within a location-year and then the median accuracy was calculated across locations within the left-out year (Figure 2.6). As before, a minimum accuracy of 0.3 was considered the threshold for a useful model. Traits such as mix time, bake time, and loaf volume had very consistent accuracy between different locations within the same year. For all traits however, this consistency varied between years with 2018 and 2021 seemingly being more variable between locations across most tested traits (Figure 2.6).

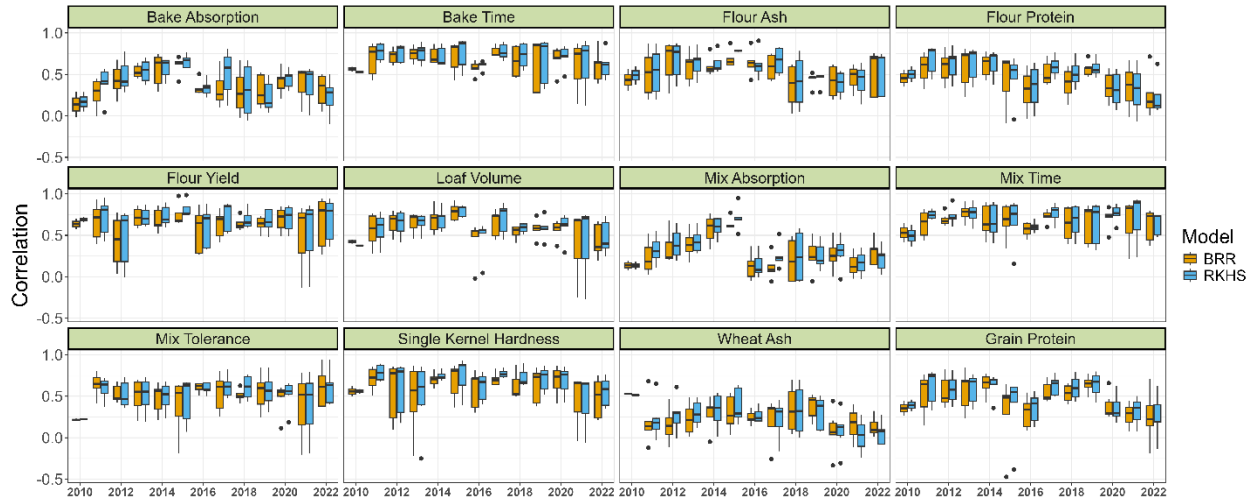


Figure 2.6. Leave-one-year-out (LOYO) validation results for the single-step method of analysis. The correlation between predicted and observed phenotypic values within a location-trial are shown on the y-axis with boxplots showing the median and range of correlations between different locations within the left-out year, denoted by the x-axis. Two representative genomic selection models (BRR and RKHS) are shown in different colors.

Discussion

The goal of this project was to perform genomic selection (GS) for end-use quality traits in hard red spring wheat. Two broad approaches for GS were tested: the two-step method in which summary BLUEs were precalculated before performing GS and the single-step method in which genetic markers were used in tandem with raw phenotypic data. Both methods were tested under two types of cross-validation, five-fold and Leave-One-Year-Out (LOYO). Further, within the five-fold cross-validation of the two-step method, the inclusion of different types of fixed effects was assessed.

Broad-sense heritabilities for the twelve end-use quality traits were generally high, ranging from 0.4 to 0.9 (Figure 2.1a) which was consistent with expectations and matched well with other studies in bread wheat that tested overlapping traits with the same method of

heritability calculation (Sandhu et al. 2022; Gill et al. 2023). Correlations between trait BLUEs revealed that they could be sorted into roughly three (overlapping) groups with group one consisting of mix time, bake time, mix tolerance, loaf volume, mix water absorption, and bake water absorption; group two consisting of loaf volume, mix water absorption, bake water absorption, total wheat ash content, grain protein, and flour protein; and group 3 consisting of flour ash content, flour yield, and single kernel hardness (Figure 2.1b). Most notably, grain protein content, which is widely used as an indicator of overall quality, was only positively correlated with five out of eleven traits and was weakly, negatively correlated with the important gluten strength trait, mix tolerance (Figure 2.1b). This was also consistent with the literature as many studies have shown that grain protein content is uncorrelated with other important quality traits such mix time, bake time, flour yield, and gluten strength (Battenfield et al. 2016; Ibba et al. 2020; Zhang-Biehn et al. 2021; Sandhu et al. 2021; Gill et al. 2025b) showing that predicting these traits directly with GS could be a superior method of improving them compared to selecting based on protein content.

The two-step method showed promising results for all quality traits except for total wheat ash content which, consistent with its lower broad-sense heritability, under performed in both cross-validation methods (Figure 2.2, 2.4). Mean prediction accuracies in the five-fold cross-validation for all the other traits met or exceeded 0.5 (Figure 2.2) which was 0.1-0.2 units higher than the k-fold validation results for the same models and traits reported in some studies (Zhang-Biehn et al. 2021; Fradgley et al. 2023; Gill et al. 2023) though consistent with the results from other studies (Battenfield et al. 2016; Sandhu et al. 2022). This was not surprising as the training dataset used in this study was large relative to the size of the breeding program. Prediction

accuracy is expected to increase as the training population size increases and as the relatedness between lines increases (Bernardo 2020). The individuals in a smaller breeding program, such as the one here, are likely to be more closely related than those from larger, more diverse programs. Further, twelve years of multi-environmental training data is substantial, and nearly 22,000 genetic markers would be expected to provide excellent, whole-genome coverage. All of these factors in tandem suggest that somewhat higher prediction accuracies, compared to some other studies, would be observed.

In the more rigorous LOYO validation, most quality traits had a prediction accuracy at or above the threshold for successful predictions of 0.3 for all years (Figure 2.4). The exception being wheat ash content, and to a lesser extent, mix water absorption. This indicated that the majority of end-use quality traits could be consistently predicted with a useful level of accuracy in spite of the year-to-year variation. Interestingly, the forward predictions, which correspond to the most recent year, 2022, in the LOYO figure, were generally equal to or less than the values for the same trait-model forward predictions reported in other studies (Battenfield et al. 2016; Gill et al. 2025b). The results from Gill et al. (2025) are especially of note as, to our knowledge, this is the only other study to investigate genomic selection of end-use quality traits within the hard red spring wheat market class. While our five-fold validation results tended to be equal or higher than their ten-fold validation results, our forward predictions were generally quite similar. A driving reason for this was likely the lack of available genotypic data for the 2021 PYT lines, which meant that the number of training lines in the 2022 AYT was quite small, only consisting of lines from two or more years prior and the usual check varieties. The AYT lines having

phenotypic data from previous years surely helped the model predict the AYT and PYT lines in a new year, and reducing the AYT sample size would likely reduce overall prediction accuracies.

End-use quality in hexaploid bread wheat is known to be controlled by several major QTL (Payne 1987; Eagles et al. 2002; Rai and Han 2023); further, it has been shown in wheat that giving markers associated with major QTL more weight in a genomic selection model, such as by treating them as fixed effects, can improve prediction accuracy (Rutkoski et al. 2014; Zhao et al. 2014; Bentley et al. 2014). However, the method of identifying which markers to treat as fixed is also important. Notably, including markers that were identified by association mapping on the combined training and testing datasets for GS can lead to artificially inflated prediction accuracies in a phenomenon termed the “inside trading effect”(Arruda et al. 2016). We demonstrated this effect in the five-fold cross-validation of the two-step method in which we observed a clear increase in prediction accuracy when markers detected by GWAS on the full BLUEs dataset were treated as fixed effects (Figure 2.3), similar to findings in Michel et al. (2018). Additionally, Michel et al. (2018) found a slight increase in prediction accuracy when markers associated with known high molecular weight glutenin genes were treated as fixed effects, which avoids the inside trading issue. Building on this, we identified genetic markers associated with the known major seed protein genes and treated those markers as fixed effects. Because many of these genes had multiple markers within 1 Mb, in addition to selecting just one marker per gene, we also performed principal component analysis on all candidate markers and then included different numbers of principal components as fixed effects in GS models. Finally, we incorporated de novo GWAS markers identified by performing GWAS on each training fold of cross-validation rather than the full dataset as fixed effects (Spindel et al. 2016). We did not

find a discernable improvement to prediction accuracy with any of these approaches, which suggested that other methods of improving model predictions should be explored in this population.

Genotype by environment (GE) interactions can reduce prediction accuracy and are known to be present in many end-use quality traits (Aoun et al. 2021). One method to account for these interactions is the single-step reaction norm model described in Jarquín et al. (2014). We investigated if these GE models could provide superior performance to the conventional two-step methods we tested previously. The median of these prediction accuracies across all location-years tended to be quite high, corresponding very closely to the broad-sense heritability values for the respective quality trait (Figure 2.1a, 2.5). The LOYO validation showed consistently strong performance for the traits mix tolerance, mix time, bake time, flour yield, and loaf volume, though other traits including grain protein, flour protein, wheat ash, and mix absorption had lower and less consistent accuracies (Figure 2.6). Further, both validation approaches had variability in prediction accuracy between different location-years, with a few location-years having notably low accuracy (Figure 2.5, 2.6). This was probably in part due to the same issue with the 2021 PYT lines discussed in the two-step section. However, it also is possible that there are some outlier environments in the training data, and optimization of the training data by removing any outlier environments could lead to higher and more consistent prediction accuracy (Heslot et al. 2013; Aoun et al. 2021).

The results from the two-step and single-step methods are not directly comparable in this study as the two-step method predicted genomic estimated breeding values (GEBVs) for each line whereas the single-step method was used to directly predict phenotypic values for a line

grown in a specific environment. Both outputs could be of use to a breeder as the GEBV represents the performance of a line as a potential parent across all tested environments whereas the single-step method gives the predicted performance for a line itself within a specific environment of interest (Jarquín et al. 2017; Persa et al. 2020). Both methods performed well; specifically, in the LOYO validation, both methods consistently gave prediction accuracies above the threshold of 0.3 (Figure 2.4, 2.6) for the majority of traits.

We showed that two methods of genomic selection, two-step BLUEs and single-step reaction norm models with a marker by environment interaction term, could be successfully used to predict important end-use quality traits in the hard red spring wheat breeding program at Montana State University. We expect that implementing these methods will allow for improvement of many end-use quality traits at earlier stages in the breeding program before phenotyping is possible. This would increase the selection intensity for these traits and lead to increased genetic gain over time for end-use quality; ensuring that breeding targets for these traits are met and that important markets, which depend on quality standards, will be maintained for producers in years to come.

References

- AACC (2000) Approved methods of the American association of cereal chemists. American Association of Cereal Chemists
- Alemu A, Åstrand J, Montesinos-López OA, et al (2024) Genomic selection in plant breeding: Key factors shaping two decades of progress. *Molecular Plant* 17:552–578. <https://doi.org/10.1016/j.molp.2024.03.007>
- Aoun M, Carter A, Thompson YA, et al (2021) Environment characterization and genomic prediction for end-use quality traits in soft white winter wheat. *The Plant Genome* 14:e20128. <https://doi.org/10.1002/tpg2.20128>

- Arruda MP, Lipka AE, Brown PJ, et al (2016) Comparing genomic selection and marker-assisted selection for Fusarium head blight resistance in wheat (*Triticum aestivum* L.). *Mol Breeding* 36:84. <https://doi.org/10.1007/s11032-016-0508-5>
- Bates D, Mächler M, Bolker B, Walker S (2015) Fitting Linear Mixed-Effects Models Using **lme4**. *J Stat Soft* 67:. <https://doi.org/10.18637/jss.v067.i01>
- Battenfield SD, Guzmán C, Gaynor RC, et al (2016) Genomic Selection for Processing and End-Use Quality Traits in the CIMMYT Spring Bread Wheat Breeding Program. *The Plant Genome* 9:plantgenome2016.01.0005. <https://doi.org/10.3835/plantgenome2016.01.0005>
- Battenfield SD, Sheridan JL, Silva LDCE, et al (2018) Breeding-assisted genomics: Applying meta-GWAS for milling and baking quality in CIMMYT wheat breeding program. *PLOS ONE* 13:e0204757. <https://doi.org/10.1371/journal.pone.0204757>
- Bentley AR, Scutari M, Gosman N, et al (2014) Applying association mapping and genomic selection to the dissection of key traits in elite European wheat. *Theor Appl Genet* 127:2619–2633. <https://doi.org/10.1007/s00122-014-2403-y>
- Bernardo R (2014) Genomewide Selection when Major Genes Are Known. *Crop Science* 54:68–75. <https://doi.org/10.2135/cropsci2013.05.0315>
- Bernardo R (2020) *Breeding for Quantitative Traits in Plants*, Third. Stemma Press
- Bernardo R, Yu J (2007) Prospects for Genomewide Selection for Quantitative Traits in Maize. *Crop Science* 47:1082–1090. <https://doi.org/10.2135/cropsci2006.11.0690>
- Bonnett D, Li Y, Crossa J, et al (2022) Response to Early Generation Genomic Selection for Yield in Wheat. *Front Plant Sci* 12:. <https://doi.org/10.3389/fpls.2021.718611>
- Brault C, Conley EJ, Read AC, et al (2025) Improving genomic prediction for plant disease using environmental covariates. *Plant Methods* 21:114. <https://doi.org/10.1186/s13007-025-01418-0>
- Browning BL, Zhou Y, Browning SR (2018) A One-Penny Imputed Genome from Next-Generation Reference Panels. *The American Journal of Human Genetics* 103:338–348. <https://doi.org/10.1016/j.ajhg.2018.07.015>
- Burgueño J, de los Campos G, Weigel K, Crossa J (2012) Genomic Prediction of Breeding Values when Modeling Genotype $\times$ Environment Interaction using Pedigree and Dense Molecular Markers. *Crop Science* 52:707–719. <https://doi.org/10.2135/cropsci2011.06.0299>
- Crossa J, Pérez-Rodríguez P, Cuevas J, et al (2017) Genomic Selection in Plant Breeding: Methods, Models, and Perspectives. *Trends in Plant Science* 22:961–975. <https://doi.org/10.1016/j.tplants.2017.08.011>

- Cullis BR, Smith AB, Coombes NE (2006) On the design of early generation variety trials with correlated data. *JABES* 11:381–393. <https://doi.org/10.1198/108571106X154443>
- Danecek P, Auton A, Abecasis G, et al (2011) The variant call format and VCFtools. *Bioinformatics* 27:2156–2158. <https://doi.org/10.1093/bioinformatics/btr330>
- Danecek P, Bonfield JK, Liddle J, et al (2021) Twelve years of SAMtools and BCFtools. *GigaScience* 10:giab008. <https://doi.org/10.1093/gigascience/giab008>
- Eagles HA, Hollamby GJ, Gororo NN, Eastwood RF (2002) Estimation and utilisation of glutenin gene effects from the analysis of unbalanced data from wheat breeding programs. *Aust J Agric Res* 53:367–377. <https://doi.org/10.1071/ar01074>
- Endelman JB (2011) Ridge Regression and Other Kernels for Genomic Selection with R Package rrBLUP. *The Plant Genome* 4:. <https://doi.org/10.3835/plantgenome2011.08.0024>
- Fradgley NS, Bentley AR, Gardner KA, et al (2023) Maintenance of UK bread baking quality: Trends in wheat quality traits over 50 years of breeding and potential for future application of genomic-assisted selection. *The Plant Genome* 16:e20326. <https://doi.org/10.1002/tpg2.20326>
- Gao Y, Dhillon GS, Joshi P, et al (2025) Enhancing genomic predictive ability of yield and yield-related traits in spring wheat by integrating major plant adaptation genes as a fixed effect. *Theor Appl Genet* 138:290. <https://doi.org/10.1007/s00122-025-05075-8>
- Garrick DJ, Taylor JF, Fernando RL (2009) Deregressing estimated breeding values and weighting information for genomic regression analyses. *Genet Sel Evol* 41:55. <https://doi.org/10.1186/1297-9686-41-55>
- Gill HS, Brar N, Halder J, et al (2023) Multi-trait genomic selection improves the prediction accuracy of end-use quality traits in hard winter wheat. *The Plant Genome* 16:e20331. <https://doi.org/10.1002/tpg2.20331>
- Gill HS, Conley E, Brault C, et al (2025) Association mapping and genomic prediction for processing and end-use quality traits in wheat (*Triticum aestivum* L.). *The Plant Genome* 18:e20529. <https://doi.org/10.1002/tpg2.20529>
- Gutierrez-Gonzalez JJ, Mascher M, Poland J, Muehlbauer GJ (2019) Dense genotyping-by-sequencing linkage maps of two Synthetic W7984×Opata reference populations provide insights into wheat structural diversity. *Sci Rep* 9:1793. <https://doi.org/10.1038/s41598-018-38111-3>
- Heffner EL, Lorenz AJ, Jannink J-L, Sorrells ME (2010) Plant Breeding with Genomic Selection: Gain per Unit Time and Cost. *Crop Science* 50:1681–1690. <https://doi.org/10.2135/cropsci2009.11.0662>

- Heslot N, Jannink J-L, Sorrells ME (2013) Using Genomic Prediction to Characterize Environments and Optimize Prediction Accuracy in Applied Breeding Data. *Crop Science* 53:921–933. <https://doi.org/10.2135/cropsci2012.07.0420>
- Holland JB, Piepho H-P (2024) Don't BLUP Twice. *G3 Genes|Genomes|Genetics* 14:jkae250. <https://doi.org/10.1093/g3journal/jkae250>
- Huang M, Liu X, Zhou Y, et al (2019) BLINK: a package for the next level of genome-wide association studies with both individuals and markers in the millions. *GigaScience* 8:giy154. <https://doi.org/10.1093/gigascience/giy154>
- Ibba MI, Crossa J, Montesinos-López OA, et al (2020) Genome-based prediction of multiple wheat quality traits in multiple years. *The Plant Genome* 13:e20034. <https://doi.org/10.1002/tpg2.20034>
- Jarquín D, Crossa J, Lacaze X, et al (2014) A reaction norm model for genomic selection using high-dimensional genomic and environmental data. *Theor Appl Genet* 127:595–607. <https://doi.org/10.1007/s00122-013-2243-1>
- Jarquín D, Lemes da Silva C, Gaynor RC, et al (2017) Increasing Genomic-Enabled Prediction Accuracy by Modeling Genotype × Environment Interactions in Kansas Wheat. *The Plant Genome* 10:plantgenome2016.12.0130. <https://doi.org/10.3835/plantgenome2016.12.0130>
- Laidig F, Piepho H-P, Rentel D, et al (2017) Breeding progress, environmental variation and correlation of winter wheat yield and quality traits in German official variety trials and on-farm during 1983–2014. *Theor Appl Genet* 130:223–245. <https://doi.org/10.1007/s00122-016-2810-3>
- Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9:357–359. <https://doi.org/10.1038/nmeth.1923>
- Lenth RV (2024) emmeans: Estimated Marginal Means, aka Least-Squares Means
- Li X, Li Y, Yu X, et al (2020) Genomics-Enabled Analysis of Puroindoline b2 Genes Identifies New Alleles in Wheat and Related Triticeae Species. *International Journal of Molecular Sciences* 21:. <https://doi.org/10.3390/ijms21041304>
- Marulanda JJ, Mi X, Utz HF, et al (2021) Optimum breeding strategies using genomic and phenotypic selection for the simultaneous improvement of two traits. *Theor Appl Genet* 134:4025–4042. <https://doi.org/10.1007/s00122-021-03945-5>
- Merrick LF, Lozada DN, Chen X, Carter AH (2022) Classification and Regression Models for Genomic Selection of Skewed Phenotypes: A Case for Disease Resistance in Winter Wheat (*Triticum aestivum* L.). *Front Genet* 13:. <https://doi.org/10.3389/fgene.2022.835781>

- Michel S, Kummer C, Gallee M, et al (2018) Improving the baking quality of bread wheat by genomic selection in early generations. *Theor Appl Genet* 131:477–493. <https://doi.org/10.1007/s00122-017-2998-x>
- Michel S, Löschenberger F, Ametz C, et al (2019) Simultaneous selection for grain yield and protein content in genomics-assisted wheat breeding. *Theor Appl Genet* 132:1745–1760. <https://doi.org/10.1007/s00122-019-03312-5>
- Montesinos-López A, Montesinos-López OA, Montesinos-López JC, et al (2021) A guide for kernel generalized regression methods for genomic-enabled prediction. *Heredity* 126:577–596. <https://doi.org/10.1038/s41437-021-00412-1>
- Payne PI (1987) Genetics of Wheat Storage Proteins and the Effect of Allelic Variation on Bread-Making Quality. *Annual Review of Plant Biology* 38:141–153. <https://doi.org/10.1146/annurev.pp.38.060187.001041>
- Pérez P, de los Campos G (2014) Genome-Wide Regression and Prediction with the BGLR Statistical Package. *Genetics* 198:483–495. <https://doi.org/10.1534/genetics.114.164442>
- Pérez-Rodríguez P, Crossa J, Rutkoski J, et al (2017) Single-Step Genomic and Pedigree Genotype × Environment Interaction Models for Predicting Wheat Lines in International Environments. *The Plant Genome* 10:plantgenome2016.09.0089. <https://doi.org/10.3835/plantgenome2016.09.0089>
- Persa R, Iwata H, Jarquin D (2020) Use of family structure information in interaction with environments for leveraging genomic prediction models. *The Crop Journal* 8:843–854. <https://doi.org/10.1016/j.cj.2020.06.004>
- Poland JA, Brown PJ, Sorrells ME, Jannink J-L (2012) Development of High-Density Genetic Maps for Barley and Wheat Using a Novel Two-Enzyme Genotyping-by-Sequencing Approach. *PLOS ONE* 7:e32253. <https://doi.org/10.1371/journal.pone.0032253>
- Prasad M, Kumar N, Kulwal P, et al (2003) QTL analysis for grain protein content using SSR markers and validation studies using NILs in bread wheat. *Theor Appl Genet* 106:659–667. <https://doi.org/10.1007/s00122-002-1114-y>
- R Core Team (2024) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria
- Rai A, Han SS (2023) Recent developments on the contribution of glutenin and puroindoline proteins to improve wheat grain quality. *Cereal Chemistry* 100:56–71. <https://doi.org/10.1002/cche.10607>
- Roberts S, Brooks K, Nogueira L, Walters CG (2022) The role of quality characteristics in pricing hard red winter wheat. *Food Policy* 108:102246. <https://doi.org/10.1016/j.foodpol.2022.102246>

- Rutkoski JE, Poland JA, Singh RP, et al (2014) Genomic Selection for Quantitative Adult Plant Stem Rust Resistance in Wheat. *The Plant Genome* 7:plantgenome2014.02.0006. <https://doi.org/10.3835/plantgenome2014.02.0006>
- Sandhu KS, Aoun M, Morris CF, et al (2021) Genomic Selection for End-Use Quality and Processing Traits in Soft White Winter Wheat Breeding Program with Machine and Deep Learning Models. *Biology* 10:. <https://doi.org/10.3390/biology10070689>
- Sandhu KS, Patil SS, Aoun M, Carter AH (2022) Multi-Trait Multi-Environment Genomic Prediction for End-Use Quality Traits in Winter Wheat. *Frontiers in Genetics* 13:
- Schmidt P, Hartung J, Rath J, Piepho H-P (2019) Estimating Broad-Sense Heritability with Unbalanced Data from Agricultural Cultivar Trials. *Crop Science* 59:525–536. <https://doi.org/10.2135/cropsci2018.06.0376>
- Shahi D, Guo J, Pradhan S, et al (2024) Genome-Wide Association Study and Genomic Prediction of Soft Wheat End-Use Quality Traits Under Post-Anthesis Heat-Stressed Conditions. *Biology* 13:. <https://doi.org/10.3390/biology13120962>
- Souza EJ, Sneller C, Guttieri MJ, et al (2012) Basis for Selecting Soft Wheat for End-Use Quality. *Crop Science* 52:21–31. <https://doi.org/10.2135/cropsci2011.02.0090>
- Spindel JE, Begum H, Akdemir D, et al (2016) Genome-wide prediction models that incorporate de novo GWAS are a powerful new tool for tropical rice improvement. *Heredity* 116:395–408. <https://doi.org/10.1038/hdy.2015.113>
- Steiner B, Michel S, Maccaferri M, et al (2019) Exploring and exploiting the genetic variation of Fusarium head blight resistance for genomic-assisted breeding in the elite durum wheat gene pool. *Theor Appl Genet* 132:969–988. <https://doi.org/10.1007/s00122-018-3253-9>
- Subedi M, Ghimire B, Bagwell JW, et al (2023) Wheat end-use quality: State of art, genetics, genomics-assisted improvement, future challenges, and opportunities. *Front Genet* 13:. <https://doi.org/10.3389/fgene.2022.1032601>
- The International Wheat Genome Sequencing Consortium (IWGSC), Appels R, Eversole K, et al (2018) Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* 361:eaar7191. <https://doi.org/10.1126/science.aar7191>
- USDA NASS (2024) Small Grains 2024 Summary 09/30/2024. United States Department of Agriculture
- Wang J, Zhang Z (2021) GAPIT Version 3: Boosting Power and Accuracy for Genomic Association and Prediction. *Genomics, Proteomics & Bioinformatics* 19:629–640. <https://doi.org/10.1016/j.gpb.2021.08.005>

- Zhang-Biehn S, Fritz AK, Zhang G, et al (2021) Accelerating wheat breeding for end-use quality through association mapping and multivariate genomic prediction. *The Plant Genome* 14:e20164. <https://doi.org/10.1002/tpg2.20164>
- Zhao Y, Mette MF, Gowda M, et al (2014) Bridging the gap between marker-assisted and genomic selection of heading time and plant height in hybrid wheat. *Heredity* 112:638–645. <https://doi.org/10.1038/hdy.2014.1>

CHAPTER FOUR

GENOMIC SELECTION FOR WHEAT STEM SAWFLY

RESISTANCE

Introduction

Wheat stem sawfly (*Cephus cinctus* Norton) (hereafter sawfly) is a major pest of hexaploid bread wheat (*Triticum aestivum* L.) in the central growing region of North America. The historic regions of sawfly infestation include Montana, North and South Dakota, western Minnesota, and south-central Canada (Beres et al. 2011), though recently, the range of sawfly infestations has spread to the state of Washington, northern Idaho, Wyoming, Colorado, Nebraska, and Kansas (Beres et al. 2017; Weaver 2023). Sawfly larvae feed within the stems of wheat plants and then cut the stems in the late season causing lodging and yield loss (Talbert et al. 2014). It has been estimated that the total economic loss due to sawfly damage could reach 400 million dollars annually (Beres et al. 2017). Sawfly pressure is also heavily controlled by environmental conditions, making it a challenge to anticipate the exact regions of heavy sawfly pressure in a given year (Olfert et al. 2019; Ermatinger et al. 2025).

Historically, management of sawfly has been achieved through the deployment of wheat cultivars containing the solid-stem trait, meaning that their stems fill with pith during development, resulting in restricted larvae movement and potential death (Hayat et al. 1995; Hager et al. 2025). While it has been shown that solid-stemmed cultivars can substantially reduce yield losses due to lodging under heavy sawfly pressure (Beres et al. 2007), even solid-stemmed cultivars can experience up to 30% stem cutting (Talbert et al. 2014; Wong et al. 2023; Hager et

al. 2025). Also, the degree of stem solidness varies substantially between environments (Cook et al. 2004), and it has been shown that abiotic factors such as increased shading and increased planting density can both decrease stem solidness (Holmes 1984; Beres et al. 2012). Further, stem solidness may vary between different internodes within a single stem and between different tillers of the same plant (Bainsla et al. 2020) giving sawfly “weak points” to target for egg laying. Given these current limitations of stem solidness, general improvement of sawfly resistance, meaning either an improved solid stem or entirely new mechanism of sawfly repulsion, is still a primary breeding target for wheat cultivar development in the affected growing regions.

Stem solidness has long been known to be a polygenic trait controlled by one major gene and multiple minor effect loci with probable epistatic interactions between them (McNeal 1955; McKenzie 1965). The major locus was first identified by Cook et al. (2004) as *Qss.msub-3BL* which mapped to the long arm of chromosome 3B and explained 76% of the solid stem variation in the double haploid line population it was mapped in. A second QTL was mapped to the long arm of chromosome 3D by Lanning et al. (2006) in the background of *Qss.msub-3BL*, which explained 31% of the variation in stem solidness. Numerous other minor QTL for stem solidness have since been identified in wheat (Sherman et al. 2010; Varella et al. 2015, 2019; Nilsen et al. 2017; Dhariwal et al. 2022). Due to the large effect of *Qss.msub-3BL*, marker assisted selection (MAS) has historically been deployed to select for solid-stemmed varieties. However, MAS becomes more challenging to implement as more unique loci are discovered and subsequently required to make further gains.

Given the polygenic nature of stem solidness, the desire to improve total sawfly resistance beyond just the solid stem trait, and the difficulty of consistently anticipating and phenotyping for sawfly resistance year to year, we sought to investigate the use of genomic selection (GS) for improving sawfly resistance in wheat. Genomic selection was first described in Meuwissen et al. (2001) and constitutes a whole genome prediction of individual performance, including both large effect and small effect loci as well as potentially nonadditive gene relationships, resulting in superior predictions to MAS, especially for low heritability traits (Bernardo and Yu 2007b; Crossa et al. 2017b). Genomic selection has been successfully adopted for numerous important traits in wheat (Bonnett et al. 2022; Gill et al. 2023; Brault et al. 2025), though only one study has applied it to total sawfly resistance (Graham 2025).

We utilized a historic agronomic dataset, collected by the Montana State University Spring Wheat Breeding Program from the years 2013-2025 combined genetic data collected from an efficient and repeatable SNP assay platform in order to conduct GS for sawfly resistance, meaning decreased sawfly-induced stem lodging, in the breeding program. We investigated seven popular GS models spanning parametric, Bayesian, semi-parametric, and machine learning approaches in two validation scenarios. Also, we investigated the genetic basis for sawfly resistance in our population and uncovered complex, epistatic gene relationships that were likely influencing our trait. Overall, we found that GS methods could be a viable and important breeding tool for improving sawfly resistance in the breeding program.

Methods

Phenotypic Data Collection

Historic agronomic data was collected by the Montana State University Spring Wheat Breeding Program between the years 2013-2025. Specifically, sawfly resistance was measured on a 0-100 scale describing the visually assessed, total percentage of lodged stems in a given three-row plot. Sawfly resistance was only collected in locations where sawfly insects were observed at some point in the growing season and other lodging-inducing factors like heavy irrigation and late-season hail were not present. A more resistant line would have a lower percentage of sawfly-induced lodging. Reduced lodging values could be attributed to the solid stem trait or any other mechanism available to a plant that would repel or combat sawfly cutting. Three different breeding pipeline stages of data were considered. The F6 generation trials consisted of between 350 to 420 unique breeding lines planted in three-row plots in an augmented experimental design with one replication of each breeding line planted per trial across multiple blocks and six released varieties planted as replicated checks within each block. The F6 trials were planted in one or two locations each year: always at the Arthur H. Post Research Farm in Bozeman, MT, and in 2024 and 2025, also at a leased farm near the city of Fort Benton, MT (Figure 3.1). The other two types of trials were the preliminary yield trials (PYT) containing either 81 or 121 unique elite breeding lines and the advanced yield trials (AYT) containing anywhere from 20 to 64 elite breeding lines. The AYT and PYT were both planted as three row plots in randomized complete block experimental designs at numerous locations across the state of Montana. Selection occurred between each stage, so lines that succeeded to the AYT stage had more total observations in more unique environments compared to lines that were removed at prior stages. Therefore, the phenotypic dataset used in this project was highly unbalanced with

many breeding lines, especially those in the F6 trials, only having one phenotypic observation in total (Figure 3.1).

The phenotypic dataset was made even more unbalanced by the focus on the sawfly resistance trait. Wheat stem sawfly pressure is highly regional (Ermatinger et al. 2025) and many trials did not have any sawfly pressure in a given year. Therefore, only trials in which sawfly pressure did occur could be kept, resulting in some years having only one or two locations of data and other years having up to nine locations (Figure 3.1).

To investigate the trends in sawfly resistance and the extent of the unbalancedness of the data, summary statistics of sawfly resistance were generated for each year using the “favstats” function of the mosaic R package (Pruim et al. 2017) (Table 3.1). Broad-sense heritability was estimated using the Cullis method (Table 3.1) as follows:

$$H_{Cullis}^2 = 1 - \frac{\bar{v}_{\Delta}^{BLUP}}{2 * \sigma_g^2} \quad (3.1)$$

where σ_g^2 is the genotypic variance and $\bar{v}_{\Delta}^{BLUP}$ is the mean variance of a difference between two genotypic Best Linear Unbiased Predictors (BLUPs)(Cullis et al. 2006). The Cullis method was selected because it has been shown to have superior performance in cases of unbalanced data compared to the traditional method of calculating broad-sense heritability (Schmidt et al. 2019).

The fully random model used to calculate line BLUPs in each year was the following:

$$y_{ijkl} = \mu + line_i + location_j + line * location_{ij} + trial_k + rep_l + \varepsilon_{ijkl} \quad (3.2)$$

where y_{ijkl} is a single observation of sawfly resistance and μ is the grand mean, $line_i$ is the random effect of the i^{th} line following a $\mathcal{N}(0, \sigma_{line}^2)$, $location_j$ is the random effect of the j^{th} location nested within year following a $\mathcal{N}(0, \sigma_{location}^2)$, $line * location_{ij}$ is the random

interaction between the i^{th} line and the j^{th} location nested within year following a $\mathcal{N}(0, \sigma_{line*location}^2)$, $trial_k$ is the random effect of the k^{th} trial nested within location nested within year following a $\mathcal{N}(0, \sigma_{trial}^2)$, rep_l is the random effect of the l^{th} replicate nested within a trial:location:year combination and following a $\mathcal{N}(0, \sigma_{rep}^2)$, and finally, ε_{ijkl} is the residual term following a $\mathcal{N}(0, \sigma^2)$. In cases where there was only one location tested within a year, a reduced model was fit excluding the location and location*line terms. All of these models were fit using the lme4 package in R (Bates et al. 2015; R Core Team 2024).



Figure 3.1. Incidence plot of each individual line phenotyped for sawfly resistance on the X-axis and the phenotyping trials in which a line was observed on the Y-axis for each year of data collected. The top bar indicates if a line was observed only one time across all trials (grey color), if the line was observed more than one time (green color), or if the line was a released variety used as a check in some trials (orange color).

Best linear unbiased estimators (BLUEs) were estimated for each line to summarize performance across all tested environments using the following model:

$$Y_{ijklmn} = \mu + G_i + Year_j + L_k + LYG_{ijk} + T_{jkl} + R_{jklm} + \varepsilon_{ijklmn} \quad (3.3)$$

$$BLUE_i = \hat{\mu} + \hat{G}_i \quad (3.4)$$

where Y_{ijklmn} is the n^{th} phenotypic observation of resistance in the m^{th} replicate nested within the l^{th} trial nested within the k^{th} location occurring in the j^{th} year taken on the i^{th} genotype. μ is the overall grand mean. G_i , $Year_j$, and L_k are the fixed effect terms associated with the i^{th} genotype, the j^{th} year, and the k^{th} location respectively. LYG_{ijk} is the random interaction term between the location-year combination of the k^{th} location with the j^{th} year interacting with the i^{th} genotype. LYG_{ijk} is distributed as $\sim \mathcal{N}(0, \sigma_{\text{location-year*genotype}}^2)$. T_{jkl} is the random intercept associated with the l^{th} trial (AYT, PYT, or F6) nested within location-year and follows a $\mathcal{N}(0, \sigma_{\text{Trial}}^2)$, and R_{jklm} is the random intercept of the m^{th} replicate following a $\mathcal{N}(0, \sigma_{\text{Rep}}^2)$. Finally, ε_{ijklmn} is the residual term following a $\mathcal{N}(0, \sigma_{jk}^2)$ with a unique residual term estimated for each location-year. This model was estimated using the sommer R package (Giovanny 2016). BLUEs were then calculated following equation 3.4 in which the estimated fixed effect associated with the i^{th} genotype, averaging across all levels of year and location, was added to the estimated grand mean.

Genetic Data Collection

Genome-wide marker data was obtained using a USDA-3K genotypic array containing 3420 variant sites mapped to the Chinese Spring Wheat IWGSC RefSeq v2.1 reference genome (Zhu et al. 2021). Genotyping was conducted in four batches between 2022-2025. Each batch was combined in R using the vcfR package (Knaus and Grünwald 2016) to load genotype files, and a custom R script to combine them by aligning marker names. Some batches had multiple replicates of the same line. In such cases, concordance between replicates was measured as the

ratio of matching genotype calls between reps to total genotype calls. Replicates with concordance to the majority of other replicates less than 0.9 were removed. Then, a majority voting method was used to combine replicated samples into one consensus result in which the most common genotype between replicates was taken for each marker. In cases of a tie vote, one of the tied genotypes was selected at random.

After the separate genotype files were combined, bcftools v1.22 (Danecek et al. 2021b) was used for further preprocessing steps. First, chromosome unknown was removed for downstream compatibility. Then, the reference alleles were normalized back against the IWGSC v2.1 reference genome, and any multiallelic sites were split into separate rows. This was necessary because some variants could either be a SNP or a deletion, and individuals were genotyped with having both outcomes. However, downstream processes required only biallelic markers, so to avoid any information loss, the variants were split and kept. Finally, markers with over 80 percent missing data were removed. This brought the total number of variants in the genotype file to 3750.

Beagle v5.5 (Browning et al. 2018b, 2021) was then used to impute missing genotype calls with a burnin parameter set to 10, total iterations set to 20, and all other parameters as their default values. The Synthetic W7984 x Opata M85 RIL 993 monotonic genetic map (Gutierrez-Gonzalez et al. 2019) was provided to estimate the recombination rate to improve imputations. Not every genotyped line had available sawfly resistance data. Therefore, the imputed marker dataset was filtered to include only lines with phenotypic data, and after filtering, any monomorphic markers were removed.

Genomic Selection Models

The BWGS R package (Charney et al. 2020; Tran and Charney 2021) was used to fit all genomic selection (GS) models. These models included Genomic Best Linear Unbiased Predictor (GBLUP) (VanRaden 2008), Epistatic GBLUP (EGBLUP) (Jiang and Reif 2015), Bayesian Ridge Regression (BRR) (de los Campos et al. 2013), Bayesian LASSO (BL) (Tibshirani 1996; Park and Casella 2008), Reproducing Kernel Hilbert Space (RKHS) using a Gaussian kernel (Gianola et al. 2006), Random Forest (RF) (Breiman 2001), and Support Vector Machine (SVM) (Ogutu et al. 2011). These models were selected in order to represent the major classes of models used in genomic selection namely; parametric (GBLUP, EGBLUP), Bayesian (BRR, BL), semiparametric (RKHS), and machine learning (RF, SVM) (Crossa et al. 2017b). The EGBLUP, RKHS, RF, and SVM models were specifically selected due to their ability account for nonadditive effects like epistasis (McKinney et al. 2006; Howard et al. 2014; Morota and Gianola 2014; Jiang and Reif 2015)

Model Validation

Two methods of validation were used to assess model performance, both of which were implemented in the BWGS package. The first method was five-fold cross validation in which the training data, consisting of the sawfly resistance line BLUEs calculated from model 4, was randomly divided into five equal folds. Each fold was removed in turn from the training data, with the remaining folds being used to fit the tested GS models. Accuracy was measured as the correlation between the genomic estimated breeding values (GEBVs) from the model for the removed fold against the known phenotypic BLUEs. A total of 100 cycles of cross validation were performed for each model with the mean and standard deviation taken across all cycles representing the model's overall performance.

The second method of validation used was a forward validation in which the BLUEs estimated for 349 experimental breeding lines grown in the 2025 F6 field trial were used as the testing set. All of the phenotypic data collected in 2025 was removed, and BLUEs were recalculated for all lines in the remaining training data. The same GS models were then used to predict GEBVs for the held out test lines, and accuracy was reported as the correlation between the predicted GEBVs and the observed phenotypic BLUEs. Then, based on the GEBVs, a theoretical selection was made where the worst 25% of lines (lines with the largest GEBV values) were removed. Then, the average BLUE (true resistance value) was taken of the selected lines and compared relative to the overall average of the BLUEs of all the 349 predicted lines. A positive value indicated that the model's selected lines were on average less resistant to sawfly compared to the full population. Next, we defined an "elite line" among the 349 F6 lines to have a BLUE value in the bottom (most resistant) 25% of the group. With this, we then counted the number of elite lines that were selected for removal by each model. This method was meant to test a model's ability to predict the newest generation of genotyped lines when no phenotypic records are available and to measure how useful the predictions of those lines might be to a breeding program.

Genome Wide Association Study

Genome Wide Association Studies (GWAS) were conducted using the rMVP R package (Yin et al. 2021, 2026) with the phenotypic BLUEs and the marker dataset as inputs. The mixed linear model (MLM) method was used. Seven principal components were included to account for population structure, with the number chosen based on the jointing in a scree plot, and markers were filtered to a minor allele frequency of 0.01. The default significance cutoff was used which

is a Bonferroni-adjusted threshold from $p \leq 0.05$ based on the total number of markers evaluated.

Epistatic Analyses

The ‘interactionPval’ function of the SNPAssoc R package (González et al. 2007; Moreno et al. 2024) was used to investigate all pairwise epistatic SNP interactions. The function performs a log-likelihood ratio test between a full interaction model and a reduced additive SNP model for each pair of markers. The response was the vector of resistance BLUEs, and there were no additional covariates beyond the marker pair. The p-value threshold for a significant interaction was a Bonferroni-adjusted threshold using the total number of tests (5,891,028) performed and a starting threshold of $p \leq 0.05$.

Removal of the F6 Data

Due to the large number of F6 lines in the phenotypic dataset which only had one observation of sawfly resistance (Figure 3.1), we wanted to investigate the effect on the results of all the above analysis methods (GS, GWAS, epistasis) on whether we included the F6 lines (total number of phenotyped lines = 1310) or if we only included lines from the replicated AYT and PYT (total number of phenotyped lines = 624). In both scenarios, all of the above procedures were identical except that the number of input BLUEs varied as well as the values of the BLUEs themselves.

Results

We first investigated the per year summary statistics of the raw phenotypic data for sawfly resistance, measured as the percentage of lodged stems in a plot, of the full phenotypic

data. We found that the values for the maximum and average lodging percentage varied substantially between years with 2014 having the lowest maximum at 15% and the lowest average at 1.27% while the years 2021, 2022, 2024, and 2025 all shared the highest maximum at 90% with 2025 also having the highest average lodging percentage at 56.35% (Table 3.1). We also found that sample size also varied by year with 2016 having the fewest number of phenotypic observations at 54 and 2025 having the most observations at 1033 (Table 3.1). Likewise, the Cullis broad-sense heritabilities ranged from 0.06 in 2015 to 0.85 in 2016 (Table 3.1), and the overall Cullis broad-sense heritability for the full dataset was 0.66.

Year	Minimum Lodging %	Maximum Lodging %	Mean Lodging %	Sample Size	Cullis H^2
2013	0	20	2.89	66	0.60
2014	0	15	1.27	162	0.49
2015	0	30	2.61	297	0.06
2016	1	25	10.80	54	0.85
2017	0	50	11.53	183	0.75
2018	0	40	2.68	549	0.21
2019	1	70	11.06	576	0.68
2020	0	70	12.13	660	0.48
2021	0	90	23.18	660	0.69
2022	0	90	15.17	636	0.77
2023	0	55	10.43	452	0.32
2024	0	90	33.38	988	0.50
2025	1	90	56.35	1033	0.50

Table 3.1. Summary statistics and Cullis broad-sense heritability values for each year of the full phenotypic sawfly resistance dataset (including the F6 trials).

Five-fold cross validation (n=100) was then performed on BLUEs calculated from the full dataset and from the reduced dataset not containing any F6 trials. For the full dataset,

average prediction accuracies ranged from 0.63 to 0.66 with the RF, EGBLUP, and RKHS models having slightly higher mean accuracies compared to the other models (Figure 3.2). In the reduced dataset, mean prediction accuracies were lower for all models and ranged from 0.53 to 0.61 with the RF model having the highest prediction accuracy, followed by EGBLUP, RKHS, and SVM being grouped and then BRR, BL, and GBLUP grouped below them (Figure 3.2).

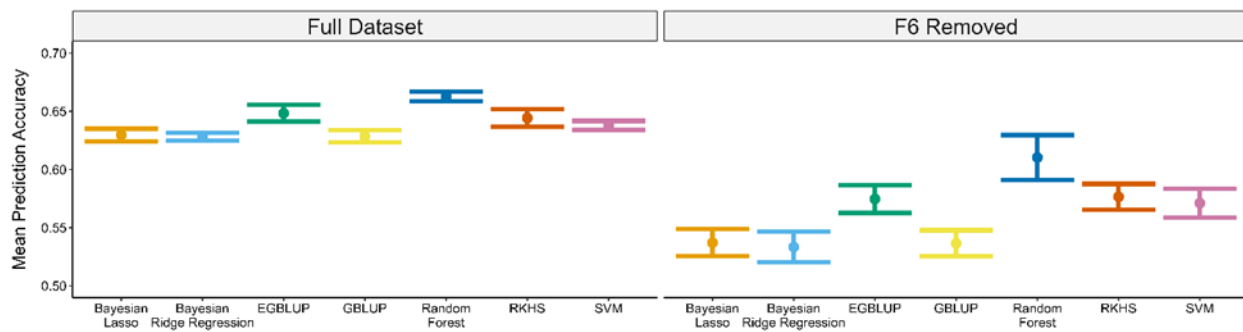


Figure 3.2. Five-fold cross validation results for seven different genomic selection models evaluated by their ability to predict sawfly resistance BLUEs. The model tested is on the X-axis and the mean prediction accuracy ($n=100$) $\pm$ one standard deviation is shown on the Y-axis. Results are shown for models fit using the full phenotypic dataset and using the phenotypic dataset without the included F6 trials.

Forward validation results demonstrated slightly higher prediction accuracies for all models using the full dataset (0.33-0.40) rather than the reduced, F6 removed dataset (0.29-0.37) (Table 3.2). However, the relative average sawfly resistance of the selected population compared to the full population of predicted lines varied by model between the two datasets where the GBLUP and BL models especially had a higher relative average using the reduced dataset compared to the full (Table 3.2). Across both datasets the relative average ranged from 4.88 to 8.86 percentage units higher of a mean in the selected population compared to the overall

population (Table 3.2). The number of elite lines selected for removal ranged from 5 to 13 out of 86 with all models except for RF removing fewer elite lines with the reduced dataset (Table 3.2).

Full Dataset				F6 Data Removed			
Model	Prediction Accuracy	Relative Average Removed (%)	Number of Elites Removed	Model	Prediction Accuracy	Relative Average Removed (%)	Number of Elites Removed
GBLUP	0.39	+6.70	13	GBLUP	0.35	+8.07	7
EGBLUP	0.40	+8.75	9	EGBLUP	0.37	+7.95	7
BRR	0.39	+7.61	12	BRR	0.35	+7.95	7
BL	0.40	+7.95	11	BL	0.34	+8.86	5
RKHS	0.39	+6.93	12	RKHS	0.34	+7.95	5
RF	0.33	+7.27	10	RF	0.29	+4.88	12
SVM	0.38	+8.29	9	SVM	0.33	+7.95	7

Table 3.2. Forward validation results for predicting sawfly resistance using seven genomic selection models. The first five columns denote results from the full dataset while the last five columns are results from the reduced dataset with the F6 trials removed. The first column in each section gives the GS model tested followed by the prediction accuracy in the second column. The third column gives the relative sawfly resistance in percent of the most resistant line that was selected for removal by the model compared to the mean resistance of all the predicted lines. The fourth column gives the mean resistance of all the lines selected for removal by the model relative to the overall mean of the predicted lines. Finally, the last column in each section denotes the number of lines selected for removal by the model that have true resistance values in the best 25% of all the predicted lines (designated “elites” relative to the group).

Because of the tendency for models that account for epistasis (EGBLUP) to outperform models that do not account for it (GBLUP) (Figure 3.2; Table 3.2), we next wanted to investigate marker pairs for potential interactions. There were numerous significant pairwise marker to marker interactions associated with sawfly resistance detected in both the full and F6 removed datasets, though substantially more interactions were detected with the reduced data (Figure 3.3). The p-values associated with the interactions were smaller with the full dataset, though all interactions shown in Figure 3.3 were below the significance threshold.

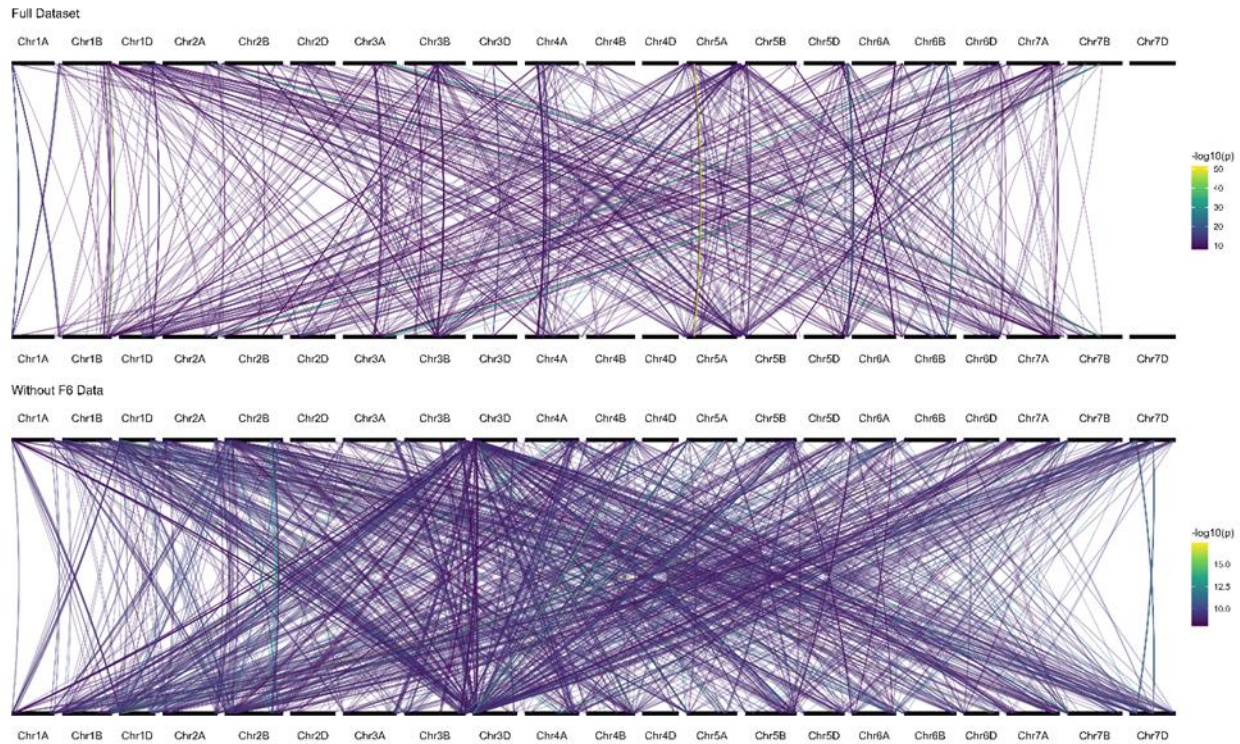


Figure 3.3. Pairwise significant marker interactions (Bonferroni corrected p-value < 0.05) that were detected in the full dataset (top panel) and in the F6 removed dataset (bottom panel). Each line represents one significant marker-marker interaction with the color of the line denoting the relative size of the p-value associated with that marker test.

Mixed linear model (MLM) GWAS results detected three significant marker-sawfly resistance associations using the full dataset BLUES; two markers clustered together on chromosome 4AL and one marker on chromosome 4BS (Figure 3.4). The alternate alleles of the two markers on chromosome 4A were estimated to have opposite effects in which the nonreference allele associated with marker ExcapSNP-4A_158 was predicted to increase sawfly resistance (negative effect on percent lodging) while the nonreference allele associated with marker ExcapSNP-4A_159 was estimated to decrease sawfly resistance (Table 3.3). The minor allele frequency associated with ExcapSNP-4A_159 was substantially lower than ExcapSNP-4A_158 (0.1 vs 0.5) (Table 3.3). There were six significant marker-trait associations detected

using the reduced, F6 removed, dataset, none of which were the same as the full dataset (Figure 3.4). These markers occurred on chromosomes 1D, 4B, 5D, and 7D with the largest estimated effect being associated with ExcapSNP-5D_87 of which the alternate allele was estimated to increase percent lodging by 5.88% on average and the minor allele frequency was 0.03 (Table 3.3).

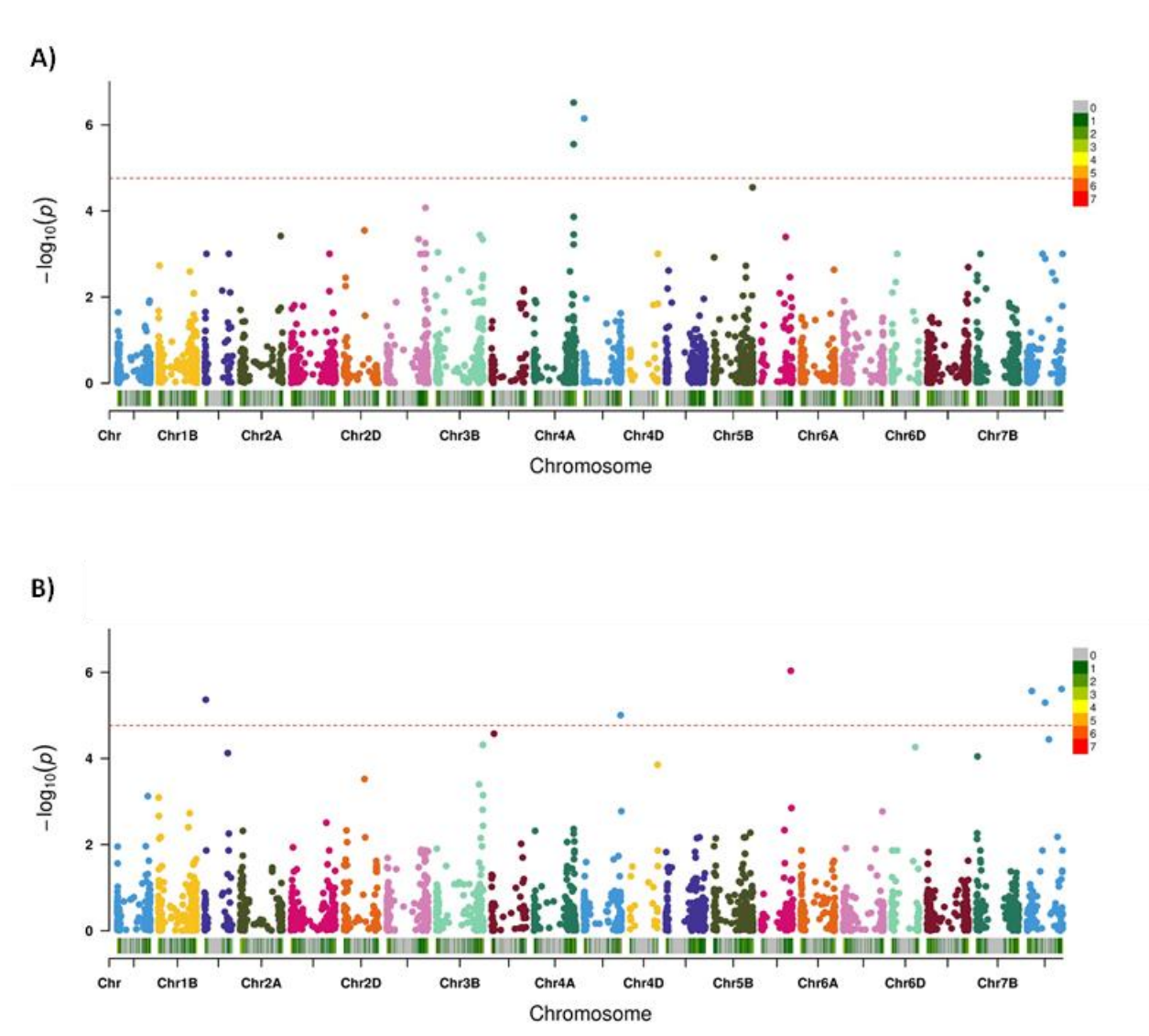


Figure 3.4. MLM GWAS results for A) models fit using the full dataset and B) models fit with the F6 trials removed. In both panels, the X-axis denotes the physical position (bp) a marker as

well as the chromosome on which it sits. The Y-axis denotes the $-\log_{10}$ of the p-value associated with a given marker. The dashed horizontal line indicates the Bonferroni significance cutoff ($p < 0.05$).

Marker	Chromosome	Position (bp)	MAF	Effect (Percent Lodging)	SE	P-value
Full Dataset						
ExcapSNP-4A_158	Chr4A	713025031	0.50	-4.78	0.93	3.061e-07
ExcapSNP-4A_159	Chr4A	713398602	0.10	4.67	0.99	2.840e-06
ExcapSNP-4B_9	Chr4B	6555519	0.29	-8.58	1.72	7.170e-07
F6 Data Removed						
ExcapSNP-1D_35	Chr1D	21816923	0.14	-3.58	0.77	4.339e-06
ExcapSNP-4B_105	Chr4B	659228747	0.09	-4.08	0.92	9.886e-06
ExcapSNP-5D_87	Chr5D	532064744	0.03	5.88	1.19	9.228e-07
ExcapSNP-7D_54	Chr7D	89630263	0.20	3.66	0.77	2.721e-06
ExcapSNP-7D_87b	Chr7D	327797338	0.09	3.89	0.85	5.067e-06
ExcapSNP-7D_128	Chr7D	623689198	0.19	-3.18	0.69	2.445e-06

Table 3.3. MLM GWAS results for the full dataset and the F6 removed dataset. Significant markers (Bonferroni threshold $p < 0.05$) are given in the first column with their corresponding chromosomes, positions, minor allele frequencies (MAF), estimated effect of the alternate allele on percent lodging, standard error of the effect (SE) and raw p-value.

Discussion

The range and average of stem lodging percentage, indicating sawfly resistance, varied drastically between years along with the sample sizes and Cullis broad-sense heritabilities (Table 3.1). Beres et al. (2007) performed a similar study by testing six hollow-stemmed and two solid-stemmed cultivars in one location across three years (2000-2002). They likewise found a large range in stem cutting percentages between different cultivars and also that stem cutting

percentage varied significantly by year (Beres et al. 2007). These findings are expected as wheat stem sawfly distributions are largely driven by climate variables, and weather conditions within the same location between different years can affect sawfly pressure and behavior (Wallace and McNeal 1966; Beres et al. 2011; Olfert et al. 2019). This is one major driver of the differing sample sizes between years, as some locations would have sawfly pressure one year and not have them in another. This further explains the differences in heritability as sawfly pressure may be influenced by environmental factors such as wind speed which could confound results.

Both training population size and heritability are a major factors in determining prediction accuracy in genomic selection (Daetwyler et al. 2010), and it has been suggested that between heritability, training population size, and marker density, heritability has the greatest impact on GS accuracies (Zhang et al. 2017). Therefore, removing the lowest heritability year ($H^2 = 0.06$) may improve accuracies; however, further testing would need to be performed to confirm this.

To investigate the question of quantity versus quality of training data, two different training sets were considered: one which included all available sawfly data (full dataset), and one in which the two, large F6 trials in 2024 and 2025 were removed (F6 removed). There were several motivating reasons for this investigation. First, the inclusion of those two trials more than doubled the set of line BLUEs resulting in a training population size of 1310 lines compared to 642 without the F6 trials, which should increase prediction accuracies and provide more statistical power for other methods like genome wide association study (GWAS) (Wang and Xu 2019). It was not possible to obtain the heritability of the F6 trials as, at the time of analysis, only

the augmented-design-corrected values were available for those trials, meaning that only one observation was available for each line and that variance components could not be estimated.

However, there was a concern with the F6 data. Specifically, that wheat stem sawfly is known to exhibit a strong “edge effect” meaning that the exterior plants of a field will often experience higher sawfly pressure than the interior plants (Nansen et al. 2005). This could result in differential sawfly pressures across an experiment, and though this should be accounted for by the block correction, it is contingent that there was consistent pressure within a block. Having only one data point per line furthers this concern as any bias or error not corrected by the experimental design will carry forward to downstream analyses.

Looking at the five-fold cross validation results, it was clear that all models using the full dataset had a higher mean prediction accuracy compared to the same models using the F6 removed dataset (Figure 3.2). Graham (2025) conducted wheat GS for sawfly traits including stem solidness, stem cutting, and sawfly infestation. The most similar trait to our sawfly resistance phenotype is stem cutting. Graham (2025) reported average three-fold cross validation accuracies around 0.43 for all models tested, which was lower than our result here, likely due to the lower Cullis heritability (0.25) for stem cutting and a smaller training population.

The random forest (RF) model had the highest average prediction accuracy with both datasets, followed by EGBLUP, RKHS, and SVM (Figure 3.2). These models have all been shown to be able to incorporate nonadditive epistatic interactions into their predictions (McKinney et al. 2006; Howard et al. 2014; Morota and Gianola 2014; Jiang and Reif 2015), whereas the other three models tested do not. This suggested that important epistatic interactions

were present in both datasets, and the models that were able to capture those interactions were making better predictions.

However, this idea of epistatic interactions was not as clear in the forward validation results. The EGBLUP and SVM models performed the best overall when the full dataset was used; however, in the reduced dataset, the Bayesian LASSO model performed the best, and with regard to relative average and fewest number of elite lines removed, it outperformed all models that used the full dataset (Table 3.2). While the RKHS model also performed well with the reduced dataset, the RF model had worst values in all metrics with the reduced dataset (Table 3.2). This could have been due to the BWGS package prioritizing speed and computational efficiency over iteration number and tuning parameters. Perhaps increasing parameters such as the number of trees and providing more computational power could have improved the RF model's performance. Regardless, the benefits of the epistatic-modeling methods was not as clear in the forward validation scenario.

Despite this, we next investigated all pairwise marker-marker interactions (Figure 3.3). We detected numerous significant interactions in both datasets, which is consistent with the expectation from the five-fold cross validation results and with previous studies that have suggested epistatic influence in solid stem inheritance (Hayat et al. 1995; Bainsla et al. 2020). Additionally, the primary locus influencing stem solidness in bread wheat, *Solid-stem locus1* (*SS1l*) (Cook et al. 2004) has been shown to have epistatic interactions with other solid stem quantitative trait loci (QTL) (Nilsen et al. 2017; Dhariwal et al. 2022). The *SS1l* gene is thought to correspond to copy number variation of the transcription factor *TdDof* which likely has a

regulatory function (Guo et al. 2009; Nilsen et al. 2020; Bathini et al.) and would further support the notion of epistatic gene interactions.

While both datasets detected numerous epistatic gene interactions, the broad patterns of these interactions varied substantially (Figure 3.3). Of particular interest, in the F6 removed dataset, there were numerous interactions at the very end of the long arm of chromosome 3B, right at the location of the *SStI* gene (Figure 3.3; Varella et al. 2015; Oiestad et al. 2017; Nilsen et al. 2017). Using the full dataset however, these interactions were not detected, and instead, interactions were detected in the middle of the long arm of 3B (Figure 3.3). The F6 removed dataset also had significant interactions between the long arm of chromosome 3B and the long arm of chromosome 4A where a QTL for stem cutting percentage (but not stem solidness) has previously been reported (Figure 3.3; Sherman et al. 2010; Varella et al. 2015). Notable interactions within the full dataset were between two markers on the short arm of chromosome 5A and two markers on the short arm of chromosome 4A, though sawfly-related QTL have not yet been mapped to either region (Figure 3.3).

Finally, we conducted an MLM GWAS in order to identify the most important marker-sawfly resistance associations in either dataset. Interestingly, in the full dataset, we detected a strong association on the long arm of chromosome 4A (Figure 3.4; Table 3.3), which likely corresponded to the 4A QTL associated with stem cutting percentage previously described (Sherman et al. 2010; Varella et al. 2015). Of the significant associations detected in the F6 removed data, Dhariwal et al. (2022) also detected minor QTL for stem solidness at similar positions on chromosomes 1DS and 5DL, but the associations on chromosomes 4B and 7D appear to be novel.

Conclusion

Our goal was to utilize historic agronomic data collected by the MSU Spring Wheat Breeding Program in order to perform genomic selection (GS) for wheat stem sawfly resistance. We found that while sawfly resistance was highly variable year to year and lines present largely unbalanced, broad-sense heritability with years was often high. Further, we were able to obtain strong prediction accuracies in five-fold cross validations and accuracies above the conventional 0.3 threshold (Heffner et al. 2010) in a challenging forward validation scenario. We also found that GS models that accounted for epistatic gene interactions outperformed those that did not. Investigating this further, we found evidence for numerous marker-marker epistatic interactions in our dataset. Finally, we were able to identify important GWAS loci that had been previously described and potentially three novel loci as well. Overall, we demonstrated that GS for sawfly resistance is a viable tool for the MSU Spring Wheat Program, and the implementation of this tool could lead to superior sawfly resistance in new varieties and thereby help manage this major wheat pest.

References

- Bainsla NK, Yadav R, Singh GP, Sharma RK (2020) Additive Genetic Behavior of Stem Solidness in Wheat (*Triticum aestivum* L.). *Sci Rep* 10:7336.
<https://doi.org/10.1038/s41598-020-64470-x>
- Bates D, Mächler M, Bolker B, Walker S (2015) Fitting Linear Mixed-Effects Models Using **lme4**. *J Stat Soft* 67:1. <https://doi.org/10.18637/jss.v067.i01>
- Bathini A, Mendu L, Singh NP, et al A “solid” solution for wheat stem sawfly (Hymenoptera: Cephidae) resistance: Genetics, breeding and development of solid stem wheat.
<https://doi.org/10.1002/csc2.20998>

- Beres BL, Cárcamo HA, Byers JR (2007) Effect of Wheat Stem Sawfly Damage on Yield and Quality of Selected Canadian Spring Wheat. *J Econ Entomol* 100:79–87. <https://doi.org/10.1093/jee/100.1.79>
- Beres BL, Dosdall LM, Weaver DK, et al (2011) Biology and integrated management of wheat stem sawfly and the need for continuing research. *The Canadian Entomologist* 143:105–125. <https://doi.org/10.4039/n10-056>
- Beres BL, Hill BD, Cárcamo HA, et al (2017) An artificial neural network model to predict wheat stem sawfly cutting in solid-stemmed wheat cultivars. *Can J Plant Sci* 97:329–336. <https://doi.org/10.1139/cjps-2016-0364>
- Beres BL, McKenzie RH, Cárcamo HA, et al (2012) Influence of Seeding Rate, Nitrogen Management, and Micronutrient Blend Applications on Pith Expression in Solid-Stemmed Spring Wheat. *Crop Science* 52:1316–1329. <https://doi.org/10.2135/cropsci2011.05.0239>
- Bernardo R, Yu J (2007) Prospects for Genomewide Selection for Quantitative Traits in Maize. *Crop Science* 47:1082–1090. <https://doi.org/10.2135/cropsci2006.11.0690>
- Bonnett D, Li Y, Crossa J, et al (2022) Response to Early Generation Genomic Selection for Yield in Wheat. *Front Plant Sci* 12:. <https://doi.org/10.3389/fpls.2021.718611>
- Brault C, Conley EJ, Read AC, et al (2025) Improving genomic prediction for plant disease using environmental covariates. *Plant Methods* 21:114. <https://doi.org/10.1186/s13007-025-01418-0>
- Breiman L (2001) Random Forests. *Machine Learning* 45:5–32. <https://doi.org/10.1023/A:1010933404324>
- Browning BL, Tian X, Zhou Y, Browning SR (2021) Fast two-stage phasing of large-scale sequence data. *The American Journal of Human Genetics* 108:1880–1890. <https://doi.org/10.1016/j.ajhg.2021.08.005>
- Browning BL, Zhou Y, Browning SR (2018) A One-Penny Imputed Genome from Next-Generation Reference Panels. *The American Journal of Human Genetics* 103:338–348. <https://doi.org/10.1016/j.ajhg.2018.07.015>
- Charmet G, Tran L-G, Auzanneau J, et al (2020) BWGS: A R package for genomic selection and its application to a wheat breeding programme. *PLOS ONE* 15:e0222733. <https://doi.org/10.1371/journal.pone.0222733>
- Cook JP, Wichman DM, Martin JM, et al (2004) Identification of Microsatellite Markers Associated with a Stem Solidness Locus in Wheat. *Crop Science* 44:1397–1402. <https://doi.org/10.2135/cropsci2004.1397>

- Crossa J, Pérez-Rodríguez P, Cuevas J, et al (2017) Genomic Selection in Plant Breeding: Methods, Models, and Perspectives. *Trends in Plant Science* 22:961–975.
<https://doi.org/10.1016/j.tplants.2017.08.011>
- Cullis BR, Smith AB, Coombes NE (2006) On the design of early generation variety trials with correlated data. *JABES* 11:381–393. <https://doi.org/10.1198/108571106X154443>
- Daetwyler HD, Pong-Wong R, Villanueva B, Woolliams JA (2010) The Impact of Genetic Architecture on Genome-Wide Evaluation Methods. *Genetics* 185:1021–1031.
<https://doi.org/10.1534/genetics.110.116855>
- Danecek P, Bonfield JK, Liddle J, et al (2021) Twelve years of SAMtools and BCFtools. *GigaScience* 10:. <https://doi.org/10.1093/gigascience/giab008>
- de los Campos G, Hickey JM, Pong-Wong R, et al (2013) Whole-Genome Regression and Prediction Methods Applied to Plant and Animal Breeding. *Genetics* 193:327–345.
<https://doi.org/10.1534/genetics.112.143313>
- Dhariwal R, Hiebert CW, Randhawa HS (2022) QTL analysis identified two major all-internodes solidness loci from a completely solid-stemmed spring wheat line. *Front Plant Sci* 13:.
<https://doi.org/10.3389/fpls.2022.1035620>
- Ermatinger LS, Powell SL, Peterson RKD, Weaver DK (2025) Mapping Wheat Stem Sawfly (*Cephus cinctus* Norton) Infestations in Spring and Winter Wheat Fields via Multiway Modelling of Multitemporal Sentinel 2 Images. *Remote Sensing* 17:1950.
<https://doi.org/10.3390/rs17111950>
- Gianola D, Fernando RL, Stella A (2006) Genomic-Assisted Prediction of Genetic Value With Semiparametric Procedures. *Genetics* 173:1761–1776.
<https://doi.org/10.1534/genetics.105.049510>
- Gill HS, Brar N, Halder J, et al (2023) Multi-trait genomic selection improves the prediction accuracy of end-use quality traits in hard winter wheat. *The Plant Genome* 16:e20331.
<https://doi.org/10.1002/tpg2.20331>
- Giovanny C-P (2016) Genome assisted prediction of quantitative traits using the R package sommer. *PLoS ONE* 11:1–15
- González JR, Armengol L, Solé X, et al (2007) SNPAssoc: an R package to perform whole genome association studies. *Bioinformatics* 23:654–655.
<https://doi.org/10.1093/bioinformatics/btm025>
- Graham SE (2025) Integrating Genomic and Phenomic Breeding Tools for Wheat Stem Sawfly Resistance in Hard Winter Wheat. Ph.D., The University of Nebraska - Lincoln

- Guo Y, Qin G, Gu H, Qu L-J (2009) Dof5.6/HCA2, a Dof transcription factor gene, regulates interfascicular cambium formation and vascular tissue development in Arabidopsis. *Plant Cell* 21:3518–3534. <https://doi.org/10.1105/tpc.108.064139>
- Gutierrez-Gonzalez JJ, Mascher M, Poland J, Muehlbauer GJ (2019) Dense genotyping-by-sequencing linkage maps of two Synthetic W7984×Opata reference populations provide insights into wheat structural diversity. *Sci Rep* 9:1793. <https://doi.org/10.1038/s41598-018-38111-3>
- Hager MS, Cook JP, Bothner B, Weaver DK (2025) Untargeted metabolomic and transcriptomic analysis in spring and durum wheat reveals potential mechanisms associated with the early stem solidness phenotype and resistance to wheat stem sawfly. *Front Plant Sci* 16:. <https://doi.org/10.3389/fpls.2025.1497732>
- Hayat MA, Martin JM, Lanning SP, et al (1995) Variation for stem solidness and its association with agronomic traits in spring wheat. *Can J Plant Sci* 75:775–780. <https://doi.org/10.4141/cjps95-131>
- Heffner EL, Lorenz AJ, Jannink J-L, Sorrells ME (2010) Plant Breeding with Genomic Selection: Gain per Unit Time and Cost. *Crop Science* 50:1681–1690. <https://doi.org/10.2135/cropsci2009.11.0662>
- Holmes ND (1984) THE EFFECT OF LIGHT ON THE RESISTANCE OF HARD RED SPRING WHEATS TO THE WHEAT STEM SAWFLY, CEPHUS CINCTUS (HYMENOPTERA: CEPHIDAE). *The Canadian Entomologist* 116:677–684. <https://doi.org/10.4039/Ent116677-5>
- Howard R, Carriquiry AL, Beavis WD (2014) Parametric and Nonparametric Statistical Methods for Genomic Selection of Traits with Additive and Epistatic Genetic Architectures. *G3 Genes|Genomes|Genetics* 4:1027–1046. <https://doi.org/10.1534/g3.114.010298>
- Jiang Y, Reif JC (2015) Modeling Epistasis in Genomic Selection. *Genetics* 201:759–768. <https://doi.org/10.1534/genetics.115.177907>
- Knaus BJ, Grünwald NJ (2016) VcfR: an R package to manipulate and visualize VCF format data. *BioRxiv*
- Lanning SP, Fox P, Elser J, et al (2006) Microsatellite Markers Associated with a Secondary Stem Solidness Locus in Wheat. *Crop Science* 46:1701–1703. <https://doi.org/10.2135/cropsci2005.10-0379>
- McKenzie H (1965) INHERITANCE OF SAWFLY REACTION AND STEM SOLIDNESS IN SPRING WHEAT CROSSES: STEM SOLIDNESS. *Can J Plant Sci* 45:591–600. <https://doi.org/10.4141/cjps65-111>

- McKinney BA, Reif DM, Ritchie MD, Moore JH (2006) Machine Learning for Detecting Gene-Gene Interactions. *Appl Bioinformatics* 5:77–88. <https://doi.org/10.2165/00822942-200605020-00002>
- McNeal FH (1955) Inheritance of Stem Solidness and Spikelet Number in a Thatcher X Rescue Wheat Cross
- Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. *Genetics* 157:1819–1829. <https://doi.org/10.1093/genetics/157.4.1819>
- Moreno V, Gonzalez JR, Pelegri D (2024) SNPAssoc: SNPs-Based Whole Genome Association Studies
- Morota G, Gianola D (2014) Kernel-based whole-genome prediction of complex traits: a review. *Front Genet* 5:. <https://doi.org/10.3389/fgene.2014.00363>
- Nansen C, Macedo TB, Weaver DK, Peterson RKD (2005) Spatiotemporal distributions of wheat stem sawfly eggs and larvae in dryland wheat fields. *The Canadian Entomologist* 137:428–440. <https://doi.org/10.4039/n04-094>
- Nilsen KT, N'Diaye A, MacLachlan PR, et al (2017) High density mapping and haplotype analysis of the major stem-solidness locus SSt1 in durum and common wheat. *PLOS ONE* 12:e0175285. <https://doi.org/10.1371/journal.pone.0175285>
- Nilsen KT, Walkowiak S, Xiang D, et al (2020) Copy number variation of TdDof controls solid-stemmed architecture in wheat. *Proceedings of the National Academy of Sciences* 117:28708–28718. <https://doi.org/10.1073/pnas.2009418117>
- Ogutu JO, Piepho H-P, Schulz-Streeck T (2011) A comparison of random forests, boosting and support vector machines for genomic selection. *BMC Proc* 5:S11. <https://doi.org/10.1186/1753-6561-5-S3-S11>
- Oiestad AJ, Martin JM, Cook J, et al (2017) Identification of Candidate Genes Responsible for Stem Pith Production Using Expression Analysis in Solid-Stemmed Wheat. *The Plant Genome* 10:plantgenome2017.02.0008. <https://doi.org/10.3835/plantgenome2017.02.0008>
- Olfert O, Weiss RM, Catton H, et al (2019) Bioclimatic assessment of abiotic factors affecting relative abundance and distribution of wheat stem sawfly (Hymenoptera: Cephidae) in western Canada. *The Canadian Entomologist* 151:16–33. <https://doi.org/10.4039/tce.2018.46>
- Park T, Casella G (2008) The Bayesian Lasso. *Journal of the American Statistical Association* 103:681–686. <https://doi.org/10.1198/016214508000000337>

- Pruim R, Kaplan DT, Horton NJ (2017) The mosaic Package: Helping Students to “Think with Data” Using R. *The R Journal* 9:77–102
- R Core Team (2024) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria
- Schmidt P, Hartung J, Rath J, Piepho H-P (2019) Estimating Broad-Sense Heritability with Unbalanced Data from Agricultural Cultivar Trials. *Crop Science* 59:525–536. <https://doi.org/10.2135/cropsci2018.06.0376>
- Sherman JD, Weaver DK, Hofland ML, et al (2010) Identification of Novel QTL for Sawfly Resistance in Wheat. *Crop Science* 50:73–86. <https://doi.org/10.2135/cropsci2009.03.0145>
- Talbert LE, Sherman JD, Hofland ML, et al (2014) Resistance to *Cephus cinctus* Norton, the wheat stem sawfly, in a recombinant inbred line population of wheat derived from two resistance sources. *Plant Breeding* 133:427–432. <https://doi.org/10.1111/pbr.12184>
- Tibshirani R (1996) Regression Shrinkage and Selection Via the Lasso. *Royal Statistical Society Journal Series B: Methodological* 58:267–288. <https://doi.org/10.1111/j.2517-6161.1996.tb02080.x>
- Tran LG, Charmet G (2021) BWGS: BreedWheat Genomic Selection Pipeline
- VanRaden PM (2008) Efficient Methods to Compute Genomic Predictions. *Journal of Dairy Science* 91:4414–4423. <https://doi.org/10.3168/jds.2007-0980>
- Varella AC, Weaver DK, Sherman JD, et al (2015) Association Analysis of Stem Solidness and Wheat Stem Sawfly Resistance in a Panel of North American Spring Wheat Germplasm. *Crop Science* 55:2046–2055. <https://doi.org/10.2135/cropsci2014.12.0852>
- Varella AC, Zhang H, Weaver DK, et al (2019) A Novel QTL in Durum Wheat for Resistance to the Wheat Stem Sawfly Associated with Early Expression of Stem Solidness. *G3 Genes|Genomes|Genetics* 9:1999–2006. <https://doi.org/10.1534/g3.119.400240>
- Wallace LE, McNeal FH (1966) Stem Sawflies of Economic Importance in Grain Crops in the United States. Agricultural Research Service, U.S. Department of Agriculture
- Wang M, Xu S (2019) Statistical power in genome-wide association studies and quantitative trait locus mapping. *Heredity* 123:287–306. <https://doi.org/10.1038/s41437-019-0205-3>
- Weaver D (2023) Wheat stem sawfly (*Cephus cinctus* Norton). In: Eigenbrode S, Rashed A (eds) *Burleigh Dodds Series in Agricultural Science*. Burleigh Dodds Science Publishing, pp 93–134

- Wong ML, Bruckner PL, Berg JE, et al (2023) Evaluation of wheat stem sawfly-resistant solid stem Qss.msub-3BL alleles in hard red winter wheat. *Crop Science* 63:556–567. <https://doi.org/10.1002/csc2.20866>
- Yin L, Zhang H, Tang Z, et al (2021) rMVP: A Memory-Efficient, Visualization-Enhanced, and Parallel-Accelerated Tool for Genome-Wide Association Study. *genom proteom bioinform* 19:619–628. <https://doi.org/10.1016/j.gpb.2020.10.007>
- Yin L, Zhang H, Tang Z, et al (2026) rMVP: Memory-Efficient, Visualize-Enhanced, Parallel-Accelerated GWAS Tool
- Zhang A, Wang H, Beyene Y, et al (2017) Effect of Trait Heritability, Training Population Size and Marker Density on Genomic Prediction Accuracy Estimation in 22 bi-parental Tropical Maize Populations. *Front Plant Sci* 8:. <https://doi.org/10.3389/fpls.2017.01916>
- Zhu T, Wang L, Rimbert H, et al (2021) Optical maps refine the bread wheat *Triticum aestivum* cv. Chinese Spring genome assembly. *The Plant Journal* 107:303–314. <https://doi.org/10.1111/tpj.15289>

CHAPTER 5

CONCLUSION

The goal of this thesis work was to apply next generation sequencing (NGS) techniques to different aspects of cultivar development in the Montana State University Spring Wheat Breeding program. In Chapter 1, fine mapping procedures were applied to a quantitative trait locus (QTL) controlling productive tiller number (PTN) in Montana germplasm. Productive tiller number is a primary component of grain yield (Lai et al. 2025) and contributes to yield plasticity (Peltonen-Sainio et al. 2011; Nasseer et al. 2016) which is critical for bread wheat (*Triticum aestivum* L.) production in Montana. Our goal was to discover the underlying causal variant(s) within the QTL for a more precise genetic control over PTN. We found that the lighting conditions in the greenhouses resulting from different times of year had a huge confounding effect on our experiments. Switching to the new, high intensity LED grow lights resulted in a massive increase in tiller numbers across all genetic groups. Further, we were able to detect differences between recombinant haplotypes using the new lights; however, we could not detect differences in tiller number between our control haplotypes, preventing us from confidently reducing the QTL region.

We next tried searching for wheat orthologs of known tiller genes in other plant species, especially rice (*Oryza sativa*). We found a candidate ortholog of the major rice tillering gene *MONOCULMI* (*MOCI*) located near our QTL of interest. We leveraged RNA sequencing data to evaluate gene expression and perform variant calling between a pair of near isogenic lines (NILs) that varied for tiller number. We found that the potential *MOCI* ortholog contained two exonic single nucleotide polymorphisms (SNPs) between the NILs, and that it was expressed in both

groups, though not differentially. From all this, we concluded that this potential *MOCI* ortholog was a strong candidate gene for our QTL and warrants further investigation.

In Chapter 2, genomic selection (GS) models were developed for twelve important end-use quality traits in the breeding program. End-use quality describes traits which are related to the milling, dough, or baking characteristics of a grain sample. The hard red wheat market class, the primary market class grown in Montana, is known for certain quality traits such as protein content and gluten strength (Subedi et al. 2023). Due to the difficulty of data collection (Battenfield et al. 2016) and the polygenic nature of these traits (Gill et al. 2025b), genomic selection was investigated as a tool for earlier generation quality improvement. We found that all of the tested GS models following the conventional two-step approach performed quite well across two validation procedures for the majority of the quality traits. However, we did not find any benefits to including extra fixed effects in these models to account for the effects of known major genes. We did however find that single-step reaction norm models which incorporated more complex genotype-by-environment interactions also tended to perform well and could be used depending on selection goals. Overall, we were able to develop genomic selection models that could be used in the breeding program for earlier generation improvement of many end-use quality traits.

In Chapter 3, genomic selection models were developed to select for wheat stem sawfly resistance. Wheat stem sawfly (*Cephus cinctus* Norton) (hereafter sawfly) is the primary insect pest of wheat in Montana, and has been estimated to cause up to US \$400 million of wheat crop loss annually across North America (Beres et al. 2017). Resistance to sawfly is a complex, polygenic trait (Bathini et al. 2023) and sawfly exert inconsistent pressures year to year (Cook et

al. 2004), making the trait difficult to measure and select on. Therefore, GS models could allow for genome-wide, consistent, and earlier generation improvement for sawfly resistance. We found that GS models which accounted for epistasis tended to outperform models which did not across two different validation methods. We found early evidence for a series of complex epistatic interactions related to sawfly resistance by conducting a naïve test of all marker-to-marker interactions. We discovered several markers associated with sawfly resistance via genome wide association studies, several of which were potentially novel including one marker on chromosome 4B and three markers on chromosome 7D. Through all this, we found that genomic selection models which accounted for epistasis could be usefully incorporated into the breeding program for improving sawfly resistance.

Across all three chapters, a new candidate gene for productive tiller number control was proposed, multiple approaches for using genomic selection models were tested and found successful for end-use quality prediction, and genomic selection was successfully developed for sawfly resistance prediction and the genetic basis for sawfly resistance was more thoroughly explored. The combination of these three projects means that next generation sequencing technologies have been thoroughly incorporated into the MSU Spring Wheat Breeding Program. Using these technologies will allow the program to identify major genes and precisely incorporate beneficial alleles for primary traits like grain yield reliably into germplasm. Further, the genomic selection methods established in this thesis will allow for consistent earlier generation selection for a comprehensive set of traits, including quality and sawfly resistance. Together, these technologies will allow for a more efficient and precise breeding pipeline which will ensure that the best possible varieties are being developed for Montana growers.

CUMULATIVE REFERENCES CITED

- AACC (2000) Approved methods of the American association of cereal chemists. American Association of Cereal Chemists
- Alahmad S, Rambla C, Voss-Fels KP, Hickey LT (2022) Accelerating Breeding Cycles. In: Reynolds MP, Braun H-J (eds) Wheat Improvement: Food Security in a Changing Climate. Springer International Publishing, Cham, pp 557–571
- Alemu A, Åstrand J, Montesinos-López OA, et al (2024) Genomic selection in plant breeding: Key factors shaping two decades of progress. *Molecular Plant* 17:552–578. <https://doi.org/10.1016/j.molp.2024.03.007>
- Aoun M, Carter A, Thompson YA, et al (2021) Environment characterization and genomic prediction for end-use quality traits in soft white winter wheat. *The Plant Genome* 14:e20128. <https://doi.org/10.1002/tpg2.20128>
- Arruda MP, Lipka AE, Brown PJ, et al (2016) Comparing genomic selection and marker-assisted selection for Fusarium head blight resistance in wheat (*Triticum aestivum* L.). *Mol Breeding* 36:84. <https://doi.org/10.1007/s11032-016-0508-5>
- Aury J-M, Engelen S, Istace B, et al (2022) Long-read and chromosome-scale assembly of the hexaploid wheat genome achieves high resolution for research and breeding. *Gigascience* 11:giac034. <https://doi.org/10.1093/gigascience/giac034>
- Awan MJA, Amin I, Rasheed A, et al (2024) Knockout mutation in TaD27 enhances number of productive tillers in hexaploid wheat. *Front Genome Ed* 6:. <https://doi.org/10.3389/fgeed.2024.1455761>
- Bainsla NK, Yadav R, Singh GP, Sharma RK (2020) Additive Genetic Behavior of Stem Solidness in Wheat (*Triticum aestivum* L.). *Sci Rep* 10:7336. <https://doi.org/10.1038/s41598-020-64470-x>
- Bates D, Mächler M, Bolker B, Walker S (2015) Fitting Linear Mixed-Effects Models Using **lme4**. *J Stat Soft* 67:. <https://doi.org/10.18637/jss.v067.i01>
- Bathini A, Mendu L, Singh NP, et al (2023) A “solid” solution for wheat stem sawfly (Hymenoptera: Cephidae) resistance: Genetics, breeding and development of solid stem wheat. *Crop Science* 63:1934–1948. <https://doi.org/10.1002/csc2.20998>
- Bathini A, Mendu L, Singh NP, et al A “solid” solution for wheat stem sawfly (Hymenoptera: Cephidae) resistance: Genetics, breeding and development of solid stem wheat. <https://doi.org/10.1002/csc2.20998>

- Battenfield SD, Guzmán C, Gaynor RC, et al (2016) Genomic Selection for Processing and End-Use Quality Traits in the CIMMYT Spring Bread Wheat Breeding Program. *The Plant Genome* 9:plantgenome2016.01.0005. <https://doi.org/10.3835/plantgenome2016.01.0005>
- Battenfield SD, Sheridan JL, Silva LDCE, et al (2018) Breeding-assisted genomics: Applying meta-GWAS for milling and baking quality in CIMMYT wheat breeding program. *PLOS ONE* 13:e0204757. <https://doi.org/10.1371/journal.pone.0204757>
- Bennett T, Sieberer T, Willett B, et al (2006) The Arabidopsis MAX Pathway Controls Shoot Branching by Regulating Auxin Transport. *Current Biology* 16:553–563. <https://doi.org/10.1016/j.cub.2006.01.058>
- Bentley AR, Scutari M, Gosman N, et al (2014) Applying association mapping and genomic selection to the dissection of key traits in elite European wheat. *Theor Appl Genet* 127:2619–2633. <https://doi.org/10.1007/s00122-014-2403-y>
- Beres BL, Cárcamo HA, Byers JR (2007) Effect of Wheat Stem Sawfly Damage on Yield and Quality of Selected Canadian Spring Wheat. *J Econ Entomol* 100:79–87. <https://doi.org/10.1093/jee/100.1.79>
- Beres BL, Dosdall LM, Weaver DK, et al (2011) Biology and integrated management of wheat stem sawfly and the need for continuing research. *The Canadian Entomologist* 143:105–125. <https://doi.org/10.4039/n10-056>
- Beres BL, Hill BD, Cárcamo HA, et al (2017) An artificial neural network model to predict wheat stem sawfly cutting in solid-stemmed wheat cultivars. *Can J Plant Sci* 97:329–336. <https://doi.org/10.1139/cjps-2016-0364>
- Beres BL, McKenzie RH, Cárcamo HA, et al (2012) Influence of Seeding Rate, Nitrogen Management, and Micronutrient Blend Applications on Pith Expression in Solid-Stemmed Spring Wheat. *Crop Science* 52:1316–1329. <https://doi.org/10.2135/cropsci2011.05.0239>
- Bernardo R (2020) *Breeding for Quantitative Traits in Plants*, Third. Stemma Press
- Bernardo R (2016) Bandwagons I, too, have known. *Theor Appl Genet* 129:2323–2332. <https://doi.org/10.1007/s00122-016-2772-5>
- Bernardo R (2014) Genomewide Selection when Major Genes Are Known. *Crop Science* 54:68–75. <https://doi.org/10.2135/cropsci2013.05.0315>
- Bernardo R, Yu J (2007a) Prospects for Genomewide Selection for Quantitative Traits in Maize. *Crop Science* 47:1082–1090. <https://doi.org/10.2135/cropsci2006.11.0690>
- Bernardo R, Yu J (2007b) Prospects for Genomewide Selection for Quantitative Traits in Maize. *Crop Science* 47:1082–1090. <https://doi.org/10.2135/cropsci2006.11.0690>

- Blake NK, Stougaard RN, Weaver DK, et al (2011) Identification of a quantitative trait locus for resistance to *Sitodiplosis mosellana* (Géhin), the orange wheat blossom midge, in spring wheat. *Plant Breeding* 130:25–30. <https://doi.org/10.1111/j.1439-0523.2010.01809.x>
- Bonnett D, Li Y, Crossa J, et al (2022) Response to Early Generation Genomic Selection for Yield in Wheat. *Front Plant Sci* 12:. <https://doi.org/10.3389/fpls.2021.718611>
- BOS H, NEUTEBOOM J (1998) Morphological Analysis of Leaf and Tiller Number Dynamics of Wheat (*Triticum aestivum*L.): Responses to Temperature and Light Intensity. *Annals of Botany* 81:131–139. <https://doi.org/10.1006/anbo.1997.0531>
- Brault C, Conley EJ, Read AC, et al (2025) Improving genomic prediction for plant disease using environmental covariates. *Plant Methods* 21:114. <https://doi.org/10.1186/s13007-025-01418-0>
- Breiman L (2001) Random Forests. *Machine Learning* 45:5–32. <https://doi.org/10.1023/A:1010933404324>
- Browning BL, Tian X, Zhou Y, Browning SR (2021) Fast two-stage phasing of large-scale sequence data. *The American Journal of Human Genetics* 108:1880–1890. <https://doi.org/10.1016/j.ajhg.2021.08.005>
- Browning BL, Zhou Y, Browning SR (2018a) A One-Penny Imputed Genome from Next-Generation Reference Panels. *The American Journal of Human Genetics* 103:338–348. <https://doi.org/10.1016/j.ajhg.2018.07.015>
- Browning BL, Zhou Y, Browning SR (2018b) A One-Penny Imputed Genome from Next-Generation Reference Panels. *The American Journal of Human Genetics* 103:338–348. <https://doi.org/10.1016/j.ajhg.2018.07.015>
- Burgueño J, de los Campos G, Weigel K, Crossa J (2012) Genomic Prediction of Breeding Values when Modeling Genotype $\times$ Environment Interaction using Pedigree and Dense Molecular Markers. *Crop Science* 52:707–719. <https://doi.org/10.2135/cropsci2011.06.0299>
- Bushnell B (2014) BBMap: A Fast, Accurate, Splice-Aware Aligner. Walnut Creek, CA
- Cao L, Li T, Geng S, et al (2023) TaSPL14-7A is a conserved regulator controlling plant architecture and yield traits in common wheat (*Triticum aestivum* L.). *Front Plant Sci* 14:. <https://doi.org/10.3389/fpls.2023.1178624>
- Cardon G, Höhmann S, Klein J, et al (1999) Molecular characterisation of the *Arabidopsis* SBP-box genes. *Gene* 237:91–104. [https://doi.org/10.1016/S0378-1119\(99\)00308-X](https://doi.org/10.1016/S0378-1119(99)00308-X)

- Cardon GH, Höhmann S, Nettlesheim K, et al (1997) Functional analysis of the *Arabidopsis thaliana* SBP-box gene *SPL3*: a novel gene involved in the floral transition. *The Plant Journal* 12:367–377. <https://doi.org/10.1046/j.1365-313X.1997.12020367.x>
- Casal JJ, Sanchez RA, Deregis VA (1986) The effect of plant density on tillering: The involvement of R/FR ratio and the proportion of radiation intercepted per plant. *Environmental and Experimental Botany* 26:365–371. [https://doi.org/10.1016/0098-8472\(86\)90024-9](https://doi.org/10.1016/0098-8472(86)90024-9)
- Charmet G, Tran L-G, Auzanneau J, et al (2020) BWGS: A R package for genomic selection and its application to a wheat breeding programme. *PLOS ONE* 15:e0222733. <https://doi.org/10.1371/journal.pone.0222733>
- Chen X, Zhang Z, Liu D, et al (2010) SQUAMOSA Promoter-Binding Protein-Like Transcription Factors: Star Players for Plant Growth and Development. *Journal of Integrative Plant Biology* 52:946–951. <https://doi.org/10.1111/j.1744-7909.2010.00987.x>
- Cook JP, Wichman DM, Martin JM, et al (2004) Identification of Microsatellite Markers Associated with a Stem Solidness Locus in Wheat. *Crop Science* 44:1397–1402. <https://doi.org/10.2135/cropsci2004.1397>
- Crespo-Herrera LA, Crossa J, Huerta-Espino J, et al (2018) Genetic Gains for Grain Yield in CIMMYT's Semi-Arid Wheat Yield Trials Grown in Suboptimal Environments. *Crop Science* 58:1890–1898. <https://doi.org/10.2135/cropsci2018.01.0017>
- Crossa J, Pérez-Rodríguez P, Cuevas J, et al (2017a) Genomic Selection in Plant Breeding: Methods, Models, and Perspectives. *Trends in Plant Science* 22:961–975. <https://doi.org/10.1016/j.tplants.2017.08.011>
- Crossa J, Pérez-Rodríguez P, Cuevas J, et al (2017b) Genomic Selection in Plant Breeding: Methods, Models, and Perspectives. *Trends in Plant Science* 22:961–975. <https://doi.org/10.1016/j.tplants.2017.08.011>
- Cullis BR, Smith AB, Coombes NE (2006) On the design of early generation variety trials with correlated data. *JABES* 11:381–393. <https://doi.org/10.1198/108571106X154443>
- Daetwyler HD, Pong-Wong R, Villanueva B, Woolliams JA (2010) The Impact of Genetic Architecture on Genome-Wide Evaluation Methods. *Genetics* 185:1021–1031. <https://doi.org/10.1534/genetics.110.116855>
- Danecek P, Auton A, Abecasis G, et al (2011) The variant call format and VCFtools. *Bioinformatics* 27:2156–2158. <https://doi.org/10.1093/bioinformatics/btr330>
- Danecek P, Bonfield JK, Liddle J, et al (2021a) Twelve years of SAMtools and BCFtools. *GigaScience* 10:giab008. <https://doi.org/10.1093/gigascience/giab008>

- Danecek P, Bonfield JK, Liddle J, et al (2021b) Twelve years of SAMtools and BCFtools. GigaScience 10:. <https://doi.org/10.1093/gigascience/giab008>
- de los Campos G, Hickey JM, Pong-Wong R, et al (2013) Whole-Genome Regression and Prediction Methods Applied to Plant and Animal Breeding. *Genetics* 193:327–345. <https://doi.org/10.1534/genetics.112.143313>
- Dhariwal R, Hiebert CW, Randhawa HS (2022) QTL analysis identified two major all-internodes solidness loci from a completely solid-stemmed spring wheat line. *Front Plant Sci* 13:. <https://doi.org/10.3389/fpls.2022.1035620>
- Dijk EL van, Auger H, Jaszczyszyn Y, Thermes C (2014) Ten years of next-generation sequencing technology. *Trends in Genetics* 30:418–426. <https://doi.org/10.1016/j.tig.2014.07.001>
- Dixon J, Braun HJ, Kosina P, Crouch JH (2009) Wheat Facts and Futures 2009. CIMMYT
- Dixon LE, Greenwood JR, Bencivenga S, et al (2018) TEOSINTE BRANCHED1 Regulates Inflorescence Architecture and Development in Bread Wheat (*Triticum aestivum*). *Plant Cell* 30:563–581. <https://doi.org/10.1105/tpc.17.00961>
- Dobin A, Davis CA, Schlesinger F, et al (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29:15–21. <https://doi.org/10.1093/bioinformatics/bts635>
- Dong C, Zhang L, Zhang Q, et al (2023) Tiller Number1 encodes an ankyrin repeat protein that controls tillering in bread wheat. *Nat Commun* 14:836. <https://doi.org/10.1038/s41467-023-36271-z>
- Eagles HA, Hollamby GJ, Gororo NN, Eastwood RF (2002) Estimation and utilisation of glutenin gene effects from the analysis of unbalanced data from wheat breeding programs. *Aust J Agric Res* 53:367–377. <https://doi.org/10.1071/ar01074>
- Endelman JB (2011) Ridge Regression and Other Kernels for Genomic Selection with R Package rrBLUP. *The Plant Genome* 4:. <https://doi.org/10.3835/plantgenome2011.08.0024>
- Ermatinger LS, Powell SL, Peterson RKD, Weaver DK (2025) Mapping Wheat Stem Sawfly (*Cephus cinctus* Norton) Infestations in Spring and Winter Wheat Fields via Multiway Modelling of Multitemporal Sentinel 2 Images. *Remote Sensing* 17:1950. <https://doi.org/10.3390/rs17111950>
- EVERS JB, VOS J, ANDRIEU B, STRUIK PC (2006) Cessation of Tillering in Spring Wheat in Relation to Light Interception and Red : Far-red Ratio. *Ann Bot* 97:649–658. <https://doi.org/10.1093/aob/mcl020>

- Fang Z, Ji Y, Hu J, et al (2020) Strigolactones and Brassinosteroids Antagonistically Regulate the Stability of the D53–OsBZR1 Complex to Determine FC1 Expression in Rice Tillering. *Molecular Plant* 13:586–597. <https://doi.org/10.1016/j.molp.2019.12.005>
- Fox J, Weisberg S (2019) *An R Companion to Applied Regression*, Third. Sage, Thousand Oaks CA
- Fradgley NS, Bentley AR, Gardner KA, et al (2023) Maintenance of UK bread baking quality: Trends in wheat quality traits over 50 years of breeding and potential for future application of genomic-assisted selection. *The Plant Genome* 16:e20326. <https://doi.org/10.1002/tpg2.20326>
- Gao Y, Dhillon GS, Joshi P, et al (2025) Enhancing genomic predictive ability of yield and yield-related traits in spring wheat by integrating major plant adaptation genes as a fixed effect. *Theor Appl Genet* 138:290. <https://doi.org/10.1007/s00122-025-05075-8>
- Garrick DJ, Taylor JF, Fernando RL (2009) Deregressing estimated breeding values and weighting information for genomic regression analyses. *Genet Sel Evol* 41:55. <https://doi.org/10.1186/1297-9686-41-55>
- Gasparis S, Miłoszewski MM (2023) Genetic Basis of Grain Size and Weight in Rice, Wheat, and Barley. *International Journal of Molecular Sciences* 24:16921. <https://doi.org/10.3390/ijms242316921>
- Gianola D, Fernando RL, Stella A (2006) Genomic-Assisted Prediction of Genetic Value With Semiparametric Procedures. *Genetics* 173:1761–1776. <https://doi.org/10.1534/genetics.105.049510>
- Gill HS, Blecha S, Brault C, et al (2025a) Genetic gains from 60 years of spring wheat breeding in the Northern Plains of the United States. *Crop Science* 65:e70106. <https://doi.org/10.1002/csc2.70106>
- Gill HS, Brar N, Halder J, et al (2023) Multi-trait genomic selection improves the prediction accuracy of end-use quality traits in hard winter wheat. *The Plant Genome* 16:e20331. <https://doi.org/10.1002/tpg2.20331>
- Gill HS, Conley E, Brault C, et al (2025b) Association mapping and genomic prediction for processing and end-use quality traits in wheat (*Triticum aestivum* L.). *The Plant Genome* 18:e20529. <https://doi.org/10.1002/tpg2.20529>
- Giovanny C-P (2016) Genome assisted prediction of quantitative traits using the R package sommer. *PLoS ONE* 11:1–15
- González JR, Armengol L, Solé X, et al (2007) SNPAssoc: an R package to perform whole genome association studies. *Bioinformatics* 23:654–655. <https://doi.org/10.1093/bioinformatics/btm025>

- Graham SE (2025) Integrating Genomic and Phenomic Breeding Tools for Wheat Stem Sawfly Resistance in Hard Winter Wheat. Ph.D., The University of Nebraska - Lincoln
- Greb T, Clarenz O, Schäfer E, et al (2003) Molecular analysis of the LATERAL SUPPRESSOR gene in Arabidopsis reveals a conserved control mechanism for axillary meristem formation. *Genes Dev* 17:1175–1187. <https://doi.org/10.1101/gad.260703>
- Guo Y, Qin G, Gu H, Qu L-J (2009) Dof5.6/HCA2, a Dof transcription factor gene, regulates interfascicular cambium formation and vascular tissue development in Arabidopsis. *Plant Cell* 21:3518–3534. <https://doi.org/10.1105/tpc.108.064139>
- Gupta PK, Rustgi S, Kumar N (2006) Genetic and molecular basis of grain size and grain number and its relevance to grain productivity in higher plants. *Genome* 49:565–571. <https://doi.org/10.1139/g06-063>
- Gutierrez-Gonzalez JJ, Mascher M, Poland J, Muehlbauer GJ (2019) Dense genotyping-by-sequencing linkage maps of two Synthetic W7984×Opata reference populations provide insights into wheat structural diversity. *Sci Rep* 9:1793. <https://doi.org/10.1038/s41598-018-38111-3>
- Hager MS, Cook JP, Bothner B, Weaver DK (2025) Untargeted metabolomic and transcriptomic analysis in spring and durum wheat reveals potential mechanisms associated with the early stem solidness phenotype and resistance to wheat stem sawfly. *Front Plant Sci* 16:. <https://doi.org/10.3389/fpls.2025.1497732>
- Hale CO, Tillett BJ, Martin JM, et al (2024) The Grain Number Increase 1 alleles GNI-A1-105Y and -105K increase grain number in spring wheat. *Crop Science* 64:2212–2222. <https://doi.org/10.1002/csc2.21267>
- Hale CO, Volkman MM, Martin JM, et al (2025) A Teosinte Branched-B1 null mutation increases durum wheat tillering, increasing grain yield in certain environments. *Crop Science* 65:e70136. <https://doi.org/10.1002/csc2.70136>
- Hayat MA, Martin JM, Lanning SP, et al (1995) Variation for stem solidness and its association with agronomic traits in spring wheat. *Can J Plant Sci* 75:775–780. <https://doi.org/10.4141/cjps95-131>
- He G, Liu J, Dong H, Sun J (2019) The Blue-Light Receptor CRY1 Interacts with BZR1 and BIN2 to Modulate the Phosphorylation and Nuclear Function of BZR1 in Repressing BR Signaling in Arabidopsis. *Molecular Plant* 12:689–703. <https://doi.org/10.1016/j.molp.2019.02.001>
- He J, Meng S, Zhao T, et al (2017) An innovative procedure of genome-wide association analysis fits studies on germplasm population and plant breeding. *Theor Appl Genet* 130:2327–2343. <https://doi.org/10.1007/s00122-017-2962-9>

- Heffner EL, Lorenz AJ, Jannink J-L, Sorrells ME (2010) Plant Breeding with Genomic Selection: Gain per Unit Time and Cost. *Crop Science* 50:1681–1690. <https://doi.org/10.2135/cropsci2009.11.0662>
- Heslot N, Jannink J-L, Sorrells ME (2013) Using Genomic Prediction to Characterize Environments and Optimize Prediction Accuracy in Applied Breeding Data. *Crop Science* 53:921–933. <https://doi.org/10.2135/cropsci2012.07.0420>
- Holland JB, Piepho H-P (2024) Don't BLUP Twice. *G3 Genes|Genomes|Genetics* 14:jkae250. <https://doi.org/10.1093/g3journal/jkae250>
- Holmes ND (1984) THE EFFECT OF LIGHT ON THE RESISTANCE OF HARD RED SPRING WHEATS TO THE WHEAT STEM SAWFLY, *CEPHUS CINCTUS* (HYMENOPTERA: CEPHIDAE). *The Canadian Entomologist* 116:677–684. <https://doi.org/10.4039/Ent116677-5>
- Hothorn T, Bretz F, Westfall P (2008) Simultaneous Inference in General Parametric Models. *Biometrical Journal* 50:346–363
- Hou H, Li J, Gao M, et al (2013) Genomic Organization, Phylogenetic Comparison and Differential Expression of the SBP-Box Family Genes in Grape. *PLOS ONE* 8:e59358. <https://doi.org/10.1371/journal.pone.0059358>
- Hou S, Yu L, Guo W, et al (2026) Fine mapping and verification of a tiller inhibition gene in synthetic hexaploid wheat derivative. *Mol Breeding* 46:17. <https://doi.org/10.1007/s11032-026-01640-1>
- Howard R, Carriquiry AL, Beavis WD (2014) Parametric and Nonparametric Statistical Methods for Genomic Selection of Traits with Additive and Epistatic Genetic Architectures. *G3 Genes|Genomes|Genetics* 4:1027–1046. <https://doi.org/10.1534/g3.114.010298>
- Huang M, Liu X, Zhou Y, et al (2019) BLINK: a package for the next level of genome-wide association studies with both individuals and markers in the millions. *GigaScience* 8:giy154. <https://doi.org/10.1093/gigascience/giy154>
- Hyles J, Vautrin S, Pettolino F, et al (2017) Repeat-length variation in a wheat cellulose synthase-like gene is associated with altered tiller number and stem cell wall composition. *J Exp Bot* 68:1519–1529. <https://doi.org/10.1093/jxb/erx051>
- Ibba MI, Crossa J, Montesinos-López OA, et al (2020) Genome-based prediction of multiple wheat quality traits in multiple years. *The Plant Genome* 13:e20034. <https://doi.org/10.1002/tpg2.20034>
- Ignatiadis N, Klaus B, Zaugg JB, Huber W (2016) Data-driven hypothesis weighting increases detection power in genome-scale multiple testing. *Nat Methods* 13:577–580. <https://doi.org/10.1038/nmeth.3885>

- Ikeda A, Ueguchi-Tanaka M, Sonoda Y, et al (2001) slender Rice, a Constitutive Gibberellin Response Mutant, Is Caused by a Null Mutation of the SLR1 Gene, an Ortholog of the Height-Regulating Gene GAI/RGA/RHT/D8. *Plant Cell* 13:999–1010. <https://doi.org/10.1105/tpc.13.5.999>
- Isik F, Holland J, Maltecca C (2017) Genomic Selection. In: Isik F, Holland J, Maltecca C (eds) *Genetic Data Analysis for Plant and Animal Breeding*. Springer International Publishing, Cham, pp 355–384
- Jannink J-L, Lorenz AJ, Iwata H (2010) Genomic selection in plant breeding: from theory to practice. *Brief Funct Genomics* 9:166–177. <https://doi.org/10.1093/bfpg/elq001>
- Jarquín D, Crossa J, Lacaze X, et al (2014) A reaction norm model for genomic selection using high-dimensional genomic and environmental data. *Theor Appl Genet* 127:595–607. <https://doi.org/10.1007/s00122-013-2243-1>
- Jarquín D, Lemes da Silva C, Gaynor RC, et al (2017) Increasing Genomic-Enabled Prediction Accuracy by Modeling Genotype $\times$ Environment Interactions in Kansas Wheat. *The Plant Genome* 10:plantgenome2016.12.0130. <https://doi.org/10.3835/plantgenome2016.12.0130>
- Jian C, Pan Y, Liu S, et al (2024) The TaGW2-TaSPL14 module regulates the trade-off between tiller number and grain weight in wheat. *Journal of Integrative Plant Biology* 66:1953–1965. <https://doi.org/10.1111/jipb.13723>
- Jiang DJ, Hao XH, Zhang J, et al (2021) Reducing expression of TaOTUB1s decreases tiller number in wheat. *Plant Signaling & Behavior* 16:2018217. <https://doi.org/10.1080/15592324.2021.2018217>
- Jiang Y, Reif JC (2015) Modeling Epistasis in Genomic Selection. *Genetics* 201:759–768. <https://doi.org/10.1534/genetics.115.177907>
- Jiao Y, Wang Y, Xue D, et al (2010) Regulation of OsSPL14 by OsmiR156 defines ideal plant architecture in rice. *Nat Genet* 42:541–544. <https://doi.org/10.1038/ng.591>
- Jones BH (2020) Genetic dissection of grain yield and yield component traits in hexaploid spring wheat
- Jones BH, Blake NK, Heo H-Y, et al (2021) Allelic response of yield component traits to resource availability in spring wheat. *Theor Appl Genet* 134:603–620. <https://doi.org/10.1007/s00122-020-03717-7>
- Kawahara Y, de la Bastide M, Hamilton JP, et al (2013) Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. *Rice* 6:4. <https://doi.org/10.1186/1939-8433-6-4>

- Khan Y, Xiong Z, Zhang H, et al (2022) Expression and roles of GRAS gene family in plant growth, signal transduction, biotic and abiotic stress resistance and symbiosis formation—a review. *Plant Biology* 24:404–416. <https://doi.org/10.1111/plb.13364>
- Knaus BJ, Grünwald NJ (2016) VcfR: an R package to manipulate and visualize VCF format data. *BioRxiv*
- Komatsu M, Maekawa M, Shimamoto K, Kyoizuka J (2001) The *LAX1* and *FRIZZY PANICLE 2* Genes Determine the Inflorescence Architecture of Rice by Controlling Rachis-Branch and Spikelet Development. *Developmental Biology* 231:364–373. <https://doi.org/10.1006/dbio.2000.9988>
- Krasileva KV, Vasquez-Gross HA, Howell T, et al (2017) Uncovering hidden variation in polyploid wheat. *Proceedings of the National Academy of Sciences* 114:E913–E921. <https://doi.org/10.1073/pnas.1619268114>
- Kuraparthi V, Sood S, Dhaliwal HS, et al (2007) Identification and mapping of a tiller inhibition gene (*tin3*) in wheat. *Theor Appl Genet* 114:285–294. <https://doi.org/10.1007/s00122-006-0431-y>
- Kuroda H, Takahashi N, Shimada H, et al (2002) Classification and Expression Analysis of Arabidopsis F-Box-Containing Protein Genes. *Plant Cell Physiol* 43:1073–1085. <https://doi.org/10.1093/pcp/pcf151>
- Lai X, Zhu Z, Zhang Y, et al (2025) The selection and application of tiller number QTLs in modern wheat breeding. *Theor Appl Genet* 138:124. <https://doi.org/10.1007/s00122-025-04908-w>
- Laidig F, Piepho H-P, Rentel D, et al (2017) Breeding progress, environmental variation and correlation of winter wheat yield and quality traits in German official variety trials and on-farm during 1983–2014. *Theor Appl Genet* 130:223–245. <https://doi.org/10.1007/s00122-016-2810-3>
- Lande R, Thompson R (1990) Efficiency of Marker-Assisted Selection in the Improvement of Quantitative Traits. *Genetics* 124:743–756. <https://doi.org/10.1093/genetics/124.3.743>
- Lander ES, Linton LM, Birren B, et al (2001) Initial sequencing and analysis of the human genome. *Nature* 409:860–921. <https://doi.org/10.1038/35057062>
- Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9:357–359. <https://doi.org/10.1038/nmeth.1923>
- Lanning S p., Carlson G r., Nash D, et al (2004) Registration of ‘Choteau’ Wheat. *Crop Science* 44:2264–2265. <https://doi.org/10.2135/cropsci2004.2264>

- Lanning SP, Carlson GR, Nash D, et al (2006a) Registration of “Vida” Wheat. *Crop Science* 46:2315–2316
- Lanning SP, Fox P, Elser J, et al (2006b) Microsatellite Markers Associated with a Secondary Stem Solidness Locus in Wheat. *Crop Science* 46:1701–1703. <https://doi.org/10.2135/cropsci2005.10-0379>
- Lanning SP, Talbert LE, McGuire CF, et al (1994) Registration of ‘McNeal’ Wheat. *Crop Science* 34:cropsci1994.0011183X003400040060x. <https://doi.org/10.2135/cropsci1994.0011183X003400040060x>
- Lenssen AW, Johnson GD, Carlson GR (2007) Cropping sequence and tillage system influences annual crop production and water use in semiarid Montana, USA. *Field Crops Research* 100:32–43. <https://doi.org/10.1016/j.fcr.2006.05.004>
- Lenth RV (2025) emmeans: Estimated Marginal Means, aka Least-Squares Means
- Lenth RV (2024) emmeans: Estimated Marginal Means, aka Least-Squares Means
- Li X, Li Y, Yu X, et al (2020) Genomics-Enabled Analysis of Puroindoline b2 Genes Identifies New Alleles in Wheat and Related Triticeae Species. *International Journal of Molecular Sciences* 21:. <https://doi.org/10.3390/ijms21041304>
- Li X, Qian Q, Fu Z, et al (2003) Control of tillering in rice. *Nature* 422:618–621. <https://doi.org/10.1038/nature01518>
- Liao Z, Yu H, Duan J, et al (2019) SLR1 inhibits MOC1 degradation to coordinate tiller number and plant height in rice. *Nat Commun* 10:2738. <https://doi.org/10.1038/s41467-019-10667-2>
- Liu J, Cheng X, Liu P, Sun J (2017) miR156-Targeted SBP-Box Transcription Factors Interact with DWARF53 to Regulate TEOSINTE BRANCHED1 and BARREN STALK1 Expression in Bread Wheat. *Plant Physiol* 174:1931–1948. <https://doi.org/10.1104/pp.17.00445>
- Liu J, Wang T, Lan Y, et al (2025) Fine-mapping and candidate gene identification for *QPtn.sau-4B* showing potential in increasing productive tiller number and yield in wheat. *The Crop Journal* 13:480–489. <https://doi.org/10.1016/j.cj.2025.01.014>
- Lloveras J, Manent J, Viudas J, et al (2004) Seeding Rate Influence on Yield and Yield Components of Irrigated Winter Wheat in a Mediterranean Climate. *Agronomy Journal* 96:1258–1265. <https://doi.org/10.2134/agronj2004.1258>
- Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* 15:550. <https://doi.org/10.1186/s13059-014-0550-8>

- Lu Z, Shao G, Xiong J, et al (2015) *MONOCULM 3*, an Ortholog of *WUSCHEL* in Rice, Is Required for Tiller Bud Formation. *Journal of Genetics and Genomics* 42:71–78. <https://doi.org/10.1016/j.jgg.2014.12.005>
- Ma Z, Zhao D, Zhang C, et al (2007) Molecular genetic analysis of five spike-related traits in wheat using RIL and immortalized F2 populations. *Mol Genet Genomics* 277:31–42. <https://doi.org/10.1007/s00438-006-0166-0>
- Manning K, Tör M, Poole M, et al (2006) A naturally occurring epigenetic mutation in a gene encoding an SBP-box transcription factor inhibits tomato fruit ripening. *Nat Genet* 38:948–952. <https://doi.org/10.1038/ng1841>
- Mao Z, He S, Xu F, et al (2020) Photoexcited CRY1 and phyB interact directly with ARF6 and ARF8 to regulate their DNA-binding activity and auxin-induced hypocotyl elongation in *Arabidopsis*. *New Phytologist* 225:848–865. <https://doi.org/10.1111/nph.16194>
- Maqbool R, Nagarajan R, Mutti JS, Gill KS (2025) Wheat Tiller -1 (WT-1), a GRAS domain encoding gene, controls both tillering and spikelet number per spike in wheat. *Plant Mol Biol* 115:128. <https://doi.org/10.1007/s11103-025-01629-y>
- Margulies M, Egholm M, Altman WE, et al (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437:376–380. <https://doi.org/10.1038/nature03959>
- Marulanda JJ, Mi X, Utz HF, et al (2021) Optimum breeding strategies using genomic and phenotypic selection for the simultaneous improvement of two traits. *Theor Appl Genet* 134:4025–4042. <https://doi.org/10.1007/s00122-021-03945-5>
- McKenna A, Hanna M, Banks E, et al (2010) The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 20:1297–1303. <https://doi.org/10.1101/gr.107524.110>
- McKenzie H (1965) INHERITANCE OF SAWFLY REACTION AND STEM SOLIDNESS IN SPRING WHEAT CROSSES: STEM SOLIDNESS. *Can J Plant Sci* 45:591–600. <https://doi.org/10.4141/cjps65-111>
- McKinney BA, Reif DM, Ritchie MD, Moore JH (2006) Machine Learning for Detecting Gene-Gene Interactions. *Appl Bioinformatics* 5:77–88. <https://doi.org/10.2165/00822942-200605020-00002>
- McNeal FH (1955) Inheritance of Stem Solidness and Spikelet Number in a Thatcher X Rescue Wheat Cross
- Merrick LF, Lozada DN, Chen X, Carter AH (2022) Classification and Regression Models for Genomic Selection of Skewed Phenotypes: A Case for Disease Resistance in Winter Wheat (*Triticum aestivum* L.). *Front Genet* 13:. <https://doi.org/10.3389/fgene.2022.835781>

- Messing J, Bharti AK, Karlowski WM, et al (2004) Sequence composition and genome organization of maize. *Proc Natl Acad Sci U S A* 101:14349–14354. <https://doi.org/10.1073/pnas.0406163101>
- Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. *Genetics* 157:1819–1829. <https://doi.org/10.1093/genetics/157.4.1819>
- Michel S, Kummer C, Gallee M, et al (2018) Improving the baking quality of bread wheat by genomic selection in early generations. *Theor Appl Genet* 131:477–493. <https://doi.org/10.1007/s00122-017-2998-x>
- Michel S, Löschenberger F, Ametz C, et al (2019) Simultaneous selection for grain yield and protein content in genomics-assisted wheat breeding. *Theor Appl Genet* 132:1745–1760. <https://doi.org/10.1007/s00122-019-03312-5>
- Miura K, Ikeda M, Matsubara A, et al (2010) OsSPL14 promotes panicle branching and higher grain productivity in rice. *Nat Genet* 42:545–549. <https://doi.org/10.1038/ng.592>
- Montesinos-López A, Montesinos-López OA, Montesinos-López JC, et al (2021) A guide for kernel generalized regression methods for genomic-enabled prediction. *Heredity* 126:577–596. <https://doi.org/10.1038/s41437-021-00412-1>
- Morales N, Ogbonna AC, Ellerbrock BJ, et al (2022) Breedbase: a digital ecosystem for modern plant breeding. *G3 Genes|Genomes|Genetics* 12:jkac078. <https://doi.org/10.1093/g3journal/jkac078>
- Moreno V, Gonzalez JR, Pelegri D (2024) SNPassoc: SNPs-Based Whole Genome Association Studies
- Morota G, Gianola D (2014) Kernel-based whole-genome prediction of complex traits: a review. *Front Genet* 5:. <https://doi.org/10.3389/fgene.2014.00363>
- Nagasawa N, Miyoshi M, Kitano H, et al (1996) Mutations associated with floral organ number in rice. *Planta* 198:627–633. <https://doi.org/10.1007/BF00262651>
- Nansen C, Macedo TB, Weaver DK, Peterson RKD (2005) Spatiotemporal distributions of wheat stem sawfly eggs and larvae in dryland wheat fields. *The Canadian Entomologist* 137:428–440. <https://doi.org/10.4039/n04-094>
- Naruoka Y, Talbert LE, Lanning SP, et al (2011) Identification of quantitative trait loci for productive tiller number and its relationship to agronomic traits in spring wheat. *Theor Appl Genet* 123:1043. <https://doi.org/10.1007/s00122-011-1646-0>

- Nasseer A m., Martin J m., Heo H y., et al (2016) Impact of a Quantitative Trait Locus for Tiller Number on Plasticity of Agronomic Traits in Spring Wheat. *Crop Science* 56:595–602. <https://doi.org/10.2135/cropsci2015.05.0325>
- Nilsen KT, N'Diaye A, MacLachlan PR, et al (2017) High density mapping and haplotype analysis of the major stem-solidness locus SSt1 in durum and common wheat. *PLOS ONE* 12:e0175285. <https://doi.org/10.1371/journal.pone.0175285>
- Nilsen KT, Walkowiak S, Xiang D, et al (2020) Copy number variation of TdDof controls solid-stemmed architecture in wheat. *Proceedings of the National Academy of Sciences* 117:28708–28718. <https://doi.org/10.1073/pnas.2009418117>
- Ogawa M, Kusano T, Katsumi M, Sano H (2000) Rice gibberellin-insensitive gene homolog, *OsGAI*, encodes a nuclear-localized protein capable of gene activation at transcriptional level. *Gene* 245:21–29. [https://doi.org/10.1016/S0378-1119\(00\)00018-4](https://doi.org/10.1016/S0378-1119(00)00018-4)
- Ogutu JO, Piepho H-P, Schulz-Streeck T (2011) A comparison of random forests, boosting and support vector machines for genomic selection. *BMC Proc* 5:S11. <https://doi.org/10.1186/1753-6561-5-S3-S11>
- Oiestad AJ, Martin JM, Cook J, et al (2017) Identification of Candidate Genes Responsible for Stem Pith Production Using Expression Analysis in Solid-Stemmed Wheat. *The Plant Genome* 10:plantgenome2017.02.0008. <https://doi.org/10.3835/plantgenome2017.02.0008>
- Olfert O, Weiss RM, Catton H, et al (2019) Bioclimatic assessment of abiotic factors affecting relative abundance and distribution of wheat stem sawfly (Hymenoptera: Cephidae) in western Canada. *The Canadian Entomologist* 151:16–33. <https://doi.org/10.4039/tce.2018.46>
- Park T, Casella G (2008) The Bayesian Lasso. *Journal of the American Statistical Association* 103:681–686. <https://doi.org/10.1198/016214508000000337>
- Payne PI (1987) Genetics of Wheat Storage Proteins and the Effect of Allelic Variation on Bread-Making Quality. *Annual Review of Plant Biology* 38:141–153. <https://doi.org/10.1146/annurev.pp.38.060187.001041>
- Peltonen-Sainio P, Jauhiainen L, Sadras VO (2011) Phenotypic plasticity of yield and agronomic traits in cereals and rapeseed at high latitudes. *Field Crops Research* 124:261–269. <https://doi.org/10.1016/j.fcr.2011.06.016>
- Peña RJ, Trethowan R, Pfeiffer WH, Ginkel MV (2002) Quality (End-Use) Improvement in Wheat: Compositional, Genetic, and Environmental Factors. *Journal of Crop Production* 5:1–37. https://doi.org/10.1300/J144v05n01_02

- Peng ZS, Yen C, Yang JL (1998) Genetic control of oligo-culms in common wheat [*Triticum aestivum*]. Wheat Information Service (Japan)
- Pérez P, de los Campos G (2014) Genome-Wide Regression and Prediction with the BGLR Statistical Package. *Genetics* 198:483–495. <https://doi.org/10.1534/genetics.114.164442>
- Pérez-Rodríguez P, Crossa J, Rutkoski J, et al (2017) Single-Step Genomic and Pedigree Genotype $\times$ Environment Interaction Models for Predicting Wheat Lines in International Environments. *The Plant Genome* 10:plantgenome2016.09.0089. <https://doi.org/10.3835/plantgenome2016.09.0089>
- Persa R, Iwata H, Jarquin D (2020) Use of family structure information in interaction with environments for leveraging genomic prediction models. *The Crop Journal* 8:843–854. <https://doi.org/10.1016/j.cj.2020.06.004>
- Poland JA, Brown PJ, Sorrells ME, Jannink J-L (2012) Development of High-Density Genetic Maps for Barley and Wheat Using a Novel Two-Enzyme Genotyping-by-Sequencing Approach. *PLOS ONE* 7:e32253. <https://doi.org/10.1371/journal.pone.0032253>
- Prasad M, Kumar N, Kulwal P, et al (2003) QTL analysis for grain protein content using SSR markers and validation studies using NILs in bread wheat. *Theor Appl Genet* 106:659–667. <https://doi.org/10.1007/s00122-002-1114-y>
- Pruim R, Kaplan DT, Horton NJ (2017) The mosaic Package: Helping Students to “Think with Data” Using R. *The R Journal* 9:77–102
- Pysh LD, Wysocka-Diller JW, Camilleri C, et al (1999) The GRAS gene family in Arabidopsis: sequence characterization and basic expression analysis of the SCARECROW-LIKE genes. *The Plant Journal* 18:111–119. <https://doi.org/10.1046/j.1365-313X.1999.00431.x>
- Qiao L, Zheng X, Zhao J, et al (2025) Genetic dissection of flag leaf morphology traits and fine mapping of a novel QTL (Qflw.sxau-6BL) in bread wheat (*Triticum aestivum* L.). *Theor Appl Genet* 138:21. <https://doi.org/10.1007/s00122-024-04802-x>
- Quintero A, Molero G, Reynolds MP, Calderini DF (2018) Trade-off between grain weight and grain number in wheat depends on G $\times$ E interaction: A case study of an elite CIMMYT panel (CIMCOG). *European Journal of Agronomy* 92:17–29. <https://doi.org/10.1016/j.eja.2017.09.007>
- R Core Team (2024) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria
- Rai A, Han SS (2023) Recent developments on the contribution of glutenin and puroindoline proteins to improve wheat grain quality. *Cereal Chemistry* 100:56–71. <https://doi.org/10.1002/cche.10607>

- Reis-Filho JS (2009) Next-generation sequencing. *Breast Cancer Res* 11:S12. <https://doi.org/10.1186/bcr2431>
- Roberts S, Brooks K, Nogueira L, Walters CG (2022) The role of quality characteristics in pricing hard red winter wheat. *Food Policy* 108:102246. <https://doi.org/10.1016/j.foodpol.2022.102246>
- Rutkoski JE, Poland JA, Singh RP, et al (2014) Genomic Selection for Quantitative Adult Plant Stem Rust Resistance in Wheat. *The Plant Genome* 7:plantgenome2014.02.0006. <https://doi.org/10.3835/plantgenome2014.02.0006>
- Sadras VO, Rebetzke GJ (2013) Plasticity of wheat grain yield is associated with plasticity of ear number. *cpsc* 64:234–243. <https://doi.org/10.1071/CP13117>
- Sandhu KS, Aoun M, Morris CF, et al (2021) Genomic Selection for End-Use Quality and Processing Traits in Soft White Winter Wheat Breeding Program with Machine and Deep Learning Models. *Biology* 10:. <https://doi.org/10.3390/biology10070689>
- Sandhu KS, Patil SS, Aoun M, Carter AH (2022) Multi-Trait Multi-Environment Genomic Prediction for End-Use Quality Traits in Winter Wheat. *Frontiers in Genetics* 13:
- Schmidt P, Hartung J, Rath J, Piepho H-P (2019) Estimating Broad-Sense Heritability with Unbalanced Data from Agricultural Cultivar Trials. *Crop Science* 59:525–536. <https://doi.org/10.2135/cropsci2018.06.0376>
- Schnable PS, Ware D, Fulton RS, et al (2009) The B73 Maize Genome: Complexity, Diversity, and Dynamics. *Science* 326:1112–1115. <https://doi.org/10.1126/science.1178534>
- Schoen A, Yadav I, Wu S, et al (2023) Identification and high-resolution mapping of a novel tiller number gene (*tin6*) by combining forward genetics screen and MutMap approach in bread wheat. *Funct Integr Genomics* 23:157. <https://doi.org/10.1007/s10142-023-01084-2>
- Schumacher K, Schmitt T, Rossberg M, et al (1999) The Lateral suppressor (*Ls*) gene of tomato encodes a new member of the VHIID protein family. *Proceedings of the National Academy of Sciences* 96:290–295. <https://doi.org/10.1073/pnas.96.1.290>
- Shahi D, Guo J, Pradhan S, et al (2024) Genome-Wide Association Study and Genomic Prediction of Soft Wheat End-Use Quality Traits Under Post-Anthesis Heat-Stressed Conditions. *Biology* 13:. <https://doi.org/10.3390/biology13120962>
- Shang Q, Wang Y, Tang H, et al (2021) Genetic, hormonal, and environmental control of tillering in wheat. *The Crop Journal* 9:986–991. <https://doi.org/10.1016/j.cj.2021.03.002>
- Shao G, Lu Z, Xiong J, et al (2019) Tiller Bud Formation Regulators MOC1 and MOC3 Cooperatively Promote Tiller Bud Outgrowth by Activating *FON1* Expression in Rice. *Molecular Plant* 12:1090–1102. <https://doi.org/10.1016/j.molp.2019.04.008>

- Shendure J, Ji H (2008) Next-generation DNA sequencing. *Nat Biotechnol* 26:1135–1145. <https://doi.org/10.1038/nbt1486>
- Shendure J, Mitra RD, Varma C, Church GM (2004) Advanced sequencing technologies: methods and goals. *Nat Rev Genet* 5:335–344. <https://doi.org/10.1038/nrg1325>
- Shendure J, Porreca GJ, Reppas NB, et al (2005) Accurate Multiplex Polony Sequencing of an Evolved Bacterial Genome. *Science* 309:1728–1732. <https://doi.org/10.1126/science.1117389>
- Sherman JD, Weaver DK, Hofland ML, et al (2010) Identification of Novel QTL for Sawfly Resistance in Wheat. *Crop Science* 50:73–86. <https://doi.org/10.2135/cropsci2009.03.0145>
- Si Y, Lu Q, Tian S, et al (2022) Fine mapping of the tiller inhibition gene TIN5 in *Triticum urartu*. *Theor Appl Genet* 135:2665–2673. <https://doi.org/10.1007/s00122-022-04140-w>
- Slafer GA, Savin R, Sadras VO (2014) Coarse and fine regulation of wheat yield components in response to genotype and environment. *Field Crops Research* 157:71–83. <https://doi.org/10.1016/j.fcr.2013.12.004>
- Souvorov A, Kapustin Y, Kiryutin B, et al (2010) Gnomon: NCBI Eukaryotic Gene Prediction Tool. National Center for Biotechnology Information, Bethesda, MD
- Souza EJ, Sneller C, Guttieri MJ, et al (2012) Basis for Selecting Soft Wheat for End-Use Quality. *Crop Science* 52:21–31. <https://doi.org/10.2135/cropsci2011.02.0090>
- Spielmeyer W, Richards RA (2004) Comparative mapping of wheat chromosome 1AS which contains the tiller inhibition gene (tin) with rice chromosome 5S. *Theor Appl Genet* 109:1303–1310. <https://doi.org/10.1007/s00122-004-1745-2>
- Spindel JE, Begum H, Akdemir D, et al (2016) Genome-wide prediction models that incorporate de novo GWAS are a powerful new tool for tropical rice improvement. *Heredity* 116:395–408. <https://doi.org/10.1038/hdy.2015.113>
- Steiner B, Michel S, Maccaferri M, et al (2019) Exploring and exploiting the genetic variation of Fusarium head blight resistance for genomic-assisted breeding in the elite durum wheat gene pool. *Theor Appl Genet* 132:969–988. <https://doi.org/10.1007/s00122-018-3253-9>
- Stirnberg P, van de Sande K, Leyser HMO (2002) MAX1 and MAX2 control shoot lateral branching in *Arabidopsis*. *Development* 129:1131–1141. <https://doi.org/10.1242/dev.129.5.1131>
- Subedi M, Ghimire B, Bagwell JW, et al (2023) Wheat end-use quality: State of art, genetics, genomics-assisted improvement, future challenges, and opportunities. *Front Genet* 13:. <https://doi.org/10.3389/fgene.2022.1032601>

- Sun L, Lai M, Ghouri F, et al (2024) Modern Plant Breeding Techniques in Crop Improvement and Genetic Diversity: From Molecular Markers and Gene Editing to Artificial Intelligence—A Critical Review. *Plants* 13:2676. <https://doi.org/10.3390/plants13192676>
- Tabuchi H, Zhang Y, Hattori S, et al (2011) LAX PANICLE2 of Rice Encodes a Novel Nuclear Protein and Regulates the Formation of Axillary Meristems. *Plant Cell* 23:3276–3287. <https://doi.org/10.1105/tpc.111.088765>
- Tadesse W, Sanchez-Garcia M, Assefa SG, et al (2019) Genetic Gains in Wheat Breeding and Its Role in Feeding the World. *Crop Breeding, Genetics and Genomics* 1:. <https://doi.org/10.20900/cbgg20190005>
- Talbert LE, Sherman JD, Hofland ML, et al (2014) Resistance to *Cephus cinctus* Norton, the wheat stem sawfly, in a recombinant inbred line population of wheat derived from two resistance sources. *Plant Breeding* 133:427–432. <https://doi.org/10.1111/pbr.12184>
- The International Wheat Genome Sequencing Consortium (IWGSC), Appels R, Eversole K, et al (2018) Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* 361:eaar7191. <https://doi.org/10.1126/science.aar7191>
- Tibshirani R (1996) Regression Shrinkage and Selection Via the Lasso. *Royal Statistical Society Journal Series B: Methodological* 58:267–288. <https://doi.org/10.1111/j.2517-6161.1996.tb02080.x>
- Tokatlidis I s. (2014) Addressing the yield by density interaction is a prerequisite to bridge the yield gap of rain-fed wheat. *Annals of Applied Biology* 165:27–42. <https://doi.org/10.1111/aab.12121>
- Tran LG, Charmet G (2021) BWGS: BreedWheat Genomic Selection Pipeline
- Tripathi RK, Goel R, Kumari S, Dahuja A (2017) Genomic organization, phylogenetic comparison, and expression profiles of the SPL family genes and their regulation in soybean. *Dev Genes Evol* 227:101–119. <https://doi.org/10.1007/s00427-017-0574-7>
- Ugarte CC, Trupkin SA, Ghiglione H, et al (2010) Low red/far-red ratios delay spike and stem growth in wheat. *J Exp Bot* 61:3151–3162. <https://doi.org/10.1093/jxb/erq140>
- Upton RN, Correr FH, Lile J, et al (2023) Design, execution, and interpretation of plant RNA-seq analyses. *Front Plant Sci* 14:. <https://doi.org/10.3389/fpls.2023.1135455>
- USDA NASS (2026) USDA/NASS 2025 State Agriculture Overview for Montana. https://www.nass.usda.gov/Quick_Stats/Ag_Overview/stateOverview.php?state=MONTANA. Accessed 19 Apr 2026
- USDA NASS (2024) Small Grains 2024 Summary 09/30/2024. United States Department of Agriculture

- Van der Auwera GA, O'Connor BD (2020) Genomics in the Cloud: Using Docker, GATK, and WDL in Terra, 1st edn.
- VanRaden PM (2008) Efficient Methods to Compute Genomic Predictions. *Journal of Dairy Science* 91:4414–4423. <https://doi.org/10.3168/jds.2007-0980>
- Varella AC, Weaver DK, Sherman JD, et al (2015) Association Analysis of Stem Solidness and Wheat Stem Sawfly Resistance in a Panel of North American Spring Wheat Germplasm. *Crop Science* 55:2046–2055. <https://doi.org/10.2135/cropsci2014.12.0852>
- Varella AC, Zhang H, Weaver DK, et al (2019) A Novel QTL in Durum Wheat for Resistance to the Wheat Stem Sawfly Associated with Early Expression of Stem Solidness. *G3 Genes|Genomes|Genetics* 9:1999–2006. <https://doi.org/10.1534/g3.119.400240>
- Wallace JG, Rodgers-Melnick E, Buckler ES (2018) On the Road to Breeding 4.0: Unraveling the Good, the Bad, and the Boring of Crop Quantitative Genomics. *Annual Review of Genetics* 52:421–444. <https://doi.org/10.1146/annurev-genet-120116-024846>
- Wallace LE, McNeal FH (1966) Stem Sawflies of Economic Importance in Grain Crops in the United States. Agricultural Research Service, U.S. Department of Agriculture
- Wang J, Zhang Z (2021) GAPIT Version 3: Boosting Power and Accuracy for Genomic Association and Prediction. *Genomics, Proteomics & Bioinformatics* 19:629–640. <https://doi.org/10.1016/j.gpb.2021.08.005>
- Wang M, Xu S (2019) Statistical power in genome-wide association studies and quantitative trait locus mapping. *Heredity* 123:287–306. <https://doi.org/10.1038/s41437-019-0205-3>
- Wang S, Wong D, Forrest K, et al (2014) Characterization of polyploid wheat genomic diversity using a high-density 90 000 single nucleotide polymorphism array. *Plant Biotechnology Journal* 12:787–796. <https://doi.org/10.1111/pbi.12183>
- Wang W, Guo W, Le L, et al (2023) Integration of high-throughput phenotyping, GWAS, and predictive models reveals the genetic architecture of plant height in maize. *Molecular Plant* 16:354–373. <https://doi.org/10.1016/j.molp.2022.11.016>
- Wang W, Lu X, Li L, et al (2018) Photoexcited CRYPTOCHROME1 Interacts with Dephosphorylated BES1 to Regulate Brassinosteroid Signaling and Photomorphogenesis in Arabidopsis. *Plant Cell* 30:1989–2005. <https://doi.org/10.1105/tpc.17.00994>
- Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* 10:57–63. <https://doi.org/10.1038/nrg2484>
- Wang Z, Wu F, Chen X, et al (2022) Fine mapping of the tiller inhibition gene TIN4 contributing to ideal plant architecture in common wheat. *Theor Appl Genet* 135:527–535. <https://doi.org/10.1007/s00122-021-03981-1>

- Weaver D (2023) Wheat stem sawfly (*Cephus cinctus* Norton). In: Eigenbrode S, Rashed A (eds) Burleigh Dodds Series in Agricultural Science. Burleigh Dodds Science Publishing, pp 93–134
- Wong ML, Bruckner PL, Berg JE, et al (2023) Evaluation of wheat stem sawfly-resistant solid stem Qss.msub-3BL alleles in hard red winter wheat. *Crop Science* 63:556–567. <https://doi.org/10.1002/csc2.20866>
- Xie Q, Mayes S, Sparkes DL (2016) Optimizing tiller production and survival for grain yield improvement in a bread wheat $\times$ spelt mapping population. *Ann Bot* 117:51–66. <https://doi.org/10.1093/aob/mcv147>
- Xu C, Wang Y, Yu Y, et al (2012) Degradation of MONOCULM 1 by APC/CTAD1 regulates rice tillering. *Nat Commun* 3:750. <https://doi.org/10.1038/ncomms1743>
- Xu F, He S, Zhang J, et al (2018) Photoactivated CRY1 and phyB Interact Directly with AUX/IAA Proteins to Inhibit Auxin Signaling in Arabidopsis. *Molecular Plant* 11:523–541. <https://doi.org/10.1016/j.molp.2017.12.003>
- Yan B, Yang Z, He G, et al (2021) The blue light receptor CRY1 interacts with GID1 and DELLA proteins to repress gibberellin signaling and plant growth. *Plant Communications* 2:100245. <https://doi.org/10.1016/j.xplc.2021.100245>
- Yan Y, Ding C, Zhang G, et al (2023) Genetic and environmental control of rice tillering. *The Crop Journal* 11:1287–1302. <https://doi.org/10.1016/j.cj.2023.05.009>
- Yang Z, Kan W, Wang Z, et al (2025) Genome-wide identification and expression analysis of phytochrome gene family in Aikang58 wheat (*Triticum aestivum* L.). *Front Plant Sci* 15:. <https://doi.org/10.3389/fpls.2024.1520457>
- Yao E, Blake VC, Sharma PK, et al (2025) GrainGenes: genetics, genomes, and pangenomes. *GENETICS* iyaf270. <https://doi.org/10.1093/genetics/iyaf270>
- Yin L, Zhang H, Tang Z, et al (2021) rMVP: A Memory-Efficient, Visualization-Enhanced, and Parallel-Accelerated Tool for Genome-Wide Association Study. *genom proteom bioinform* 19:619–628. <https://doi.org/10.1016/j.gpb.2020.10.007>
- Yin L, Zhang H, Tang Z, et al (2026) rMVP: Memory-Efficient, Visualize-Enhanced, Parallel-Accelerated GWAS Tool
- Yoshida S (1972) Physiological Aspects of Grain Yield. *Annual Review of Plant Biology* 23:437–464. <https://doi.org/10.1146/annurev.pp.23.060172.002253>
- Zadoks JC, Chang TT, Konzak CF (1974) A decimal code for the growth stages of cereals. *Weed Research* 14:415–421. <https://doi.org/10.1111/j.1365-3180.1974.tb01084.x>

- Zhai R, Feng Y, Wang H, et al (2013) Transcriptome analysis of rice root heterosis by RNA-Seq. *BMC Genomics* 14:19. <https://doi.org/10.1186/1471-2164-14-19>
- Zhang A, Wang H, Beyene Y, et al (2017) Effect of Trait Heritability, Training Population Size and Marker Density on Genomic Prediction Accuracy Estimation in 22 bi-parental Tropical Maize Populations. *Front Plant Sci* 8:. <https://doi.org/10.3389/fpls.2017.01916>
- Zhang B, Liu X, Xu W, et al (2015) Novel function of a putative MOC1 ortholog associated with spikelet number per spike in common wheat. *Sci Rep* 5:12211. <https://doi.org/10.1038/srep12211>
- Zhang J, Wu J, Liu W, et al (2013) Genetic mapping of a fertile tiller inhibition gene, *ftin*, in wheat. *Mol Breeding* 31:441–449. <https://doi.org/10.1007/s11032-012-9801-0>
- Zhang N, Liu Y, Gui S, Wang Y (2025) Regulation of tillering and panicle branching in rice and wheat. *Journal of Genetics and Genomics* 52:869–886. <https://doi.org/10.1016/j.jgg.2024.12.005>
- Zhang-Biehn S, Fritz AK, Zhang G, et al (2021) Accelerating wheat breeding for end-use quality through association mapping and multivariate genomic prediction. *The Plant Genome* 14:e20164. <https://doi.org/10.1002/tpg2.20164>
- Zhao B, Wu TT, Ma SS, et al (2020) TaD27-B gene controls the tiller number in hexaploid wheat. *Plant Biotechnology Journal* 18:513–525. <https://doi.org/10.1111/pbi.13220>
- Zhao C, Liu X, Liu H, et al (2022) Fine mapping of QFlw-5B, a major QTL for flag leaf width in common wheat (*Triticum aestivum* L.). *Theor Appl Genet* 135:2531–2541. <https://doi.org/10.1007/s00122-022-04135-7>
- Zhao Y, Mette MF, Gowda M, et al (2014) Bridging the gap between marker-assisted and genomic selection of heading time and plant height in hybrid wheat. *Heredity* 112:638–645. <https://doi.org/10.1038/hdy.2014.1>
- Zhu T, Liu Y, Ma L, et al (2020) Genome-wide identification, phylogeny and expression analysis of the SPL gene family in wheat. *BMC Plant Biol* 20:420. <https://doi.org/10.1186/s12870-020-02576-0>
- Zhu T, Wang L, Rimbert H, et al (2021) Optical maps refine the bread wheat *Triticum aestivum* cv. Chinese Spring genome assembly. *The Plant Journal* 107:303–314. <https://doi.org/10.1111/tpj.15289>
- Zhuang H, Guo Z, Wang J, Chen T (2024a) Genome-wide identification and comprehensive analysis of the phytochrome-interacting factor (PIF) gene family in wheat. *PLOS ONE* 19:e0296269. <https://doi.org/10.1371/journal.pone.0296269>

Zhuang L, Liu H, Hou J, et al (2024b) Genetic improvement of important agronomic traits in Chinese wheat breeding over the past 70 years. BMC Plant Biol 24:1151.
<https://doi.org/10.1186/s12870-024-05841-8>

(2026) FAOSTAT. <https://www.fao.org/faostat/>. Accessed 19 Apr 2026

(2010) *Oryza sativa* Japonica Group IPA1 (IPA1) gene, complete cds