

DISSERTATION

CHARACTERIZING RESISTANCE TO WHEAT STEM SAWFLY (*CEPHUS CINCTUS*
NORTON) IN HARD WINTER WHEAT (*TRITICUM AESTIVUM* L.)

Submitted By

Mik Hammers

Department of Soil and Crop Sciences

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Summer 2026

Doctoral Committee:

Advisor: R. Esten Mason

Geoff Morris
Punya Nachappa
Katherine Frels

Copyright by Mik Hammers 2026

All Rights Reserved

ABSTRACT

CHARACTERIZING RESISTANCE TO WHEAT STEM SAWFLY (*CEPHUS CINCTUS* NORTON) IN HARD WINTER WHEAT (*TRITICUM AESTIVUM* L.)

Wheat stem sawfly (WSS) is a native grass feeding pest of winter wheat (*Triticum aestivum* L) which is difficult to control since most of its life cycle occurs within the stem of wheat plants. The only well-characterized genetic resistance to WSS is the solid stem locus (*Sst1*) on chromosome 3B, which exhibits environmental variability. It is critical to further characterize the solid stem trait and identify novel forms of genetic resistance outside of *Sst1* to improve the overall resistance of wheat to WSS.

In the second chapter, genome-wide association studies (GWAS) were performed on lines in the Colorado State University wheat breeding program grown between 2014 and 2025 field seasons for three traits of interest: heading date (HD), WSS damage in the form of stem cutting (CUT), and stem solidity (SOLID). Significant marker-trait associations (MTA) from these GWAS were used to identify beneficial allelic combinations (AC) for WSS resistance. The stem cutting ACs which had the lowest damage estimates were those in which lines possessed the resistant haplotype at every locus assessed (CUT = 2.35, error = 0.21, N = 56). Lines that had all resistant alleles in the stem solidity ACs showed the same superior estimate (SOLID = 14.7, error = 0.78, N = 22). MTA for GWAS and AC models did not overlap, indicating the presence of WSS resistance in the population that is not associated with stem solidity. Understanding these different AC groups and their effects on stem solidity and cutting can improve efficiency in breeding for WSS resistance.

Despite the large effect of *SSII*, additional loci associated with these WSS resistant phenotypes have previously been identified. This oligogenic genetic control lends itself to genomic prediction. In chapter three, best linear unbiased prediction (GBLUP), GBLUP+*SSII*, Bayesian (Bayes A and BayesLASSO), and random forest models were assessed for predictive ability using cross- and forward-validation. Additionally, environments were culled based on heritability to evaluate the effect of training population heritability on predictive ability. Overall, random forest models performed well in cross- and forward-validation for cutting ($\bar{r}_{CV} = 0.67$, $\bar{r}_{FV} = 0.45$) and GBLUP performed well for solidity ($\bar{r}_{CV} = 0.68$, $\bar{r}_{FV} = 0.64$). Increasing heritability of the training population did not improved cross-validation predictive ability for cutting ($\bar{r}_{BayesA} = 0.75$, $\bar{r}_{random\ forest} = 0.77$) or solidity ($\bar{r}_{GBLUP} = 0.85$, $\bar{r}_{BayesA} = 0.86$). For forward-validation, increasing heritability did not improve predictive ability for cutting ($\bar{r}_{BayesA} = 0.39$, $\bar{r}_{random\ forest} = 0.43$) and decreased predictive ability for solidity ($\bar{r}_{GBLUP} = 0.54$, $\bar{r}_{BayesA} = 0.55$). These results highlight the need to identify models best suited to the genetic architecture of individual traits of interest.

In chapters four and five, two pair of near isogenic lines (NILs) were grown over two field seasons to understand the metabolome, proteome, and phytohormone profile differences in response to WSS infestation in hollow stem and solid stem wheat. It has been reported that Clearfield®(CL) wheat varieties exhibit lower WSS infestation. To further understand potential molecular differences between CL non-CL varieties, chapter four utilized the herbicide tolerant line ‘Byrd CL Plus’ and non-herbicide tolerant line ‘Byrd’. Stem samples were collected for three replicates from each NIL during pre-, peak, and post- infestation time points. Five differentially expressed proteins were identified across all three time points. At peak infestation, two protein accessions A0A3B6JCK1 (Log₂FC = -3.48, p-value_{FDR} < 0.01) and A0A3B6NUD3 (Log₂FC = -

2.30, $p\text{-value}_{\text{FDR}} < 0.01$) were found to be associated with the flowering gene *FLOWERING LOCUS C (FLC)* and glycolysis, respectively. One unannotated metabolite was upregulated during peak infestation ($\text{Log}_2\text{FC} = 3.20$, $p\text{-value}_{\text{FDR}} = 0.02$). Additionally, exploratory analysis of significant metabolites ($\text{Log}_2\text{FC} \geq |6|$, $p\text{-value}_{\text{unadjusted}} < 0.05$) revealed strong correlations with metabolites and phytohormones, abscisic acid and jasmonate iso-leucine, which were associated with plant defense pathways. While these results potential low levels differences in plant responses and defenses, it also showed the complexity of the of plant responses to infestation and underlying defense mechanisms.

To further investigate molecular differences between hollow stem and solid stem wheat varieties, chapter 5 focuses on a NIL pair containing the hollow stem line ‘CO20SFD047R.4’ and solid stem line ‘CO20SFD047R.1’ were utilized in the same study as chapter 4. Post-infestation, two protein accessions A0A3B6NPX7 ($\text{Log}_2\text{FC} = -3.32$, $p\text{-value}_{\text{FDR}} = 0.01$) and A0A3B6SBN0 ($\text{Log}_2\text{FC} = -2.45$, $p\text{-value}_{\text{FDR}} = 0.02$) were found to be associated with lipid transport and galactose transport and cell wall component synthesis, respectively. Additionally, the defense signaling phytohormone jasmonate isoleucine (JA-Ile) was significantly differentially expressed at peak infestation ($\text{Log}_2\text{FC} = 1.59$, $p\text{-value}_{\text{FDR}} = 0.02$) and post-infestation ($\text{Log}_2\text{FC} = -1.46$, $p\text{-value}_{\text{FDR}} = 0.04$). These results highlight the complex downstream effects associated with *SstI* resistance to WSS.

Solid stem expression is variable and the tradeoff with yield component traits is not well understood. In chapter six, eight sets of near isogenic lines (NIL) were grown across six environments to investigate the impact of the *SstI* locus on 11 yield component traits and solidity. There were little to no consistent effects of the *SstI* locus on biomass traits, plant height, and heading date. Analysis supported the identification of three NIL groups, NIL1, NIL6, and NIL8

that were consistently solid across all environments with solidity estimates of 14.5, 14.5, and 12.9, respectively. However, grain yield in solid stem lines was 205 kg ha⁻¹ greater than the hollow stem line for NIL1 and 747 kg ha⁻¹, and 260 kg ha⁻¹ lower than the hollow stem allele for NIL6 and NIL8. These results highlighted NIL1 as being an ideal genetic background for the solid stem trait due to its reliable solidity and minimal yield-related penalties. Furthermore, these results indicate no consistent trade off in agronomic traits associated with the *SS1* locus across different genetic backgrounds. Further understanding the relationship between sawfly resistance and agronomic trade-offs could be beneficial for developing more elite WSS resistant cultivars.

ACKNOWLEDGEMENTS

As the past four years come to a close, I cannot help but think about all of the people who have helped me succeed and made this experience possible. I would like to start by thanking everyone closest to me. To my friends, thank you for being there through all the highs and lows, helping me laugh along the way, and giving me some of the most amazing memories. To my family, thank you for always believing in me and encouraging me to try new things. I could not have done this without your unwavering support.

I would also like to thank my advisor, Dr. Esten Mason, and all of the members of the Colorado State University Wheat Breeding and Genetics crew: Emily Hudson-Arns, Forrest Wold-McGimsey, John Stromberger, Scott Seifert, Dr. Zach Winn, Meena Santra, Hong Wang, Collin Marshall, and Emily Billow. It has been a privilege to work with such an amazing group of people. I hope to spend my career surrounded by coworkers who are as kind as you all have been. I will always remember the great times we shared, even on the longest field days. Thank you all for being there for me throughout my six years in Fort Collins.

TABLE OF CONTENTS

ABSTRACT.....	ii
ACKNOWLEDGEMENTS	vi
CHAPTER I – LITURATURE REVIEW	1
Wheat and Food Security	1
Wheat Stem Sawfly.....	1
Management Strategies for Wheat Stem Sawfly	2
Genetic Resistance to Wheat Stem Sawfly	3
Environmental Effects and Trade-Offs	5
Additional Wheat Stem Sawfly Genomic Resistance.....	6
Tools and Strategies for Breeding.....	6
Genomic Selection	6
Genome Wide Association Studies.....	8
Multi-omics.....	9
CHAPTER II - IDENTIFYING MARKER-TRAIT ASSOCIATIONS FOR WHEAT STEM SAWFLY RESISTANCE.....	11
INTRODUCTION	11
MATERIALS AND METHODS.....	16
Germplasm.....	16
Data Collection	16
Genotyping.....	17
Statistical Analysis.....	18
Best Linear Unbiased Estimates	18
Heritability	19

CUT, SOLID, and HD GWAS	20
GWAS with Marker Covariates.....	21
Allelic Combination Analysis.....	22
AC Marker Analysis	23
RESULTS	24
Phenotypic Evaluations.....	24
Marker Distribution and Population Structure.....	26
Cutting Score, Solid Stem, and Heading Date GWAS	26
Cutting Score GWAS With Marker Covariates	27
Solid Stem GWAS With Marker Covariates	29
Allelic Combination Analysis.....	34
AC Marker Analysis	36
DISCUSSION.....	39
CONCLUSIONS.....	44
CHAPTER III – ASSESSING PREDICTIVE ABILITY OF GENOMIC PREDICTION MODELS FOR WHEAT STEM SAWFLY RESISTANCE	45
INTRODUCTION	45
MATERIALS AND METHODS.....	49
Germplasm.....	49
Data Collection	49
Genotyping.....	50
Statistical Analysis.....	51
Phenotypic Data Analysis	51
Heritability	52
Genomic Prediction	53

GBLUP	53
Bayes A and Bayes LASSO.....	53
Random Forest	55
Cross-Validation Predictive Ability.....	57
Forward-Validation.....	57
Heritability Culling	57
RESULTS	58
Phenotypic Analysis.....	58
Genomic Prediction using All Environments	61
DISCUSSION	68
CONCLUSIONS.....	72
CHAPTER IV – UTILIZING MULTI-OMICS TO UNDERSTAND THE LINK BETWEEN CLEARFIELD AND WHEAT STEM SAWFLY RESISTANCE.....	73
INTRODUCTION	73
MATERIALS AND METHODS.....	77
Germplasm.....	77
Sample Collection.....	77
Phytohormone and Metabolomic Sample Preparation	79
Proteomics Sample Preparation	82
Metabolomics Sample Processing	84
Statistical Analysis.....	86
Differential Expression Analysis	86
Statistical Analysis - Proteomics	87
Phenotypic Analysis.....	87
RESULTS	88

Variation in the Proteome	88
Variation in the Metabolome	92
Variation in the Phytohormone Profile	96
DISCUSSION	99
CONCLUSIONS.....	104
CHAPTER V – UTILIZING MULTI-OMICS TO UNDERSTAND MOLECULAR MECHANISMS OF SOLID STEM RESISTANCE IN WINTER WHEAT (<i>Triticum aestivum</i> L.)	106
INTRODUCTION	106
MATERIALS AND METHODS.....	110
Germplasm.....	110
Sample Collection	111
Sample and Data Preparation and Statistical Analysis	112
RESULTS	112
Variation in the Proteome	112
Variation in the Metabolome	116
Variation in the Phytohormone Profile	120
DISCUSSION.....	125
CONCLUSIONS.....	129
CHAPTER VI – EFFECTS OF SS _{t1} , SOLIDITY, AND ENVIRONMENT ON YIELD- COMPONENT TRAITS IN HARD WINTER WHEAT	131
INTRODUCTION	131
MATERIALS AND METHODS.....	134
Germplasm.....	134
Sample Collection	135
Genotyping.....	137

Data Analysis	137
RESULTS	139
Population and Data Assessment	139
General assessments of traits across environments	139
Phylogenetic Analysis.....	143
Effect of Allelic Variation at <i>SSrI</i> Across Environment	144
Effect significance across traits	144
Biomass Traits	146
Physiological Traits	146
Components of Yield	148
Yield and Test Weight	149
Solidity	150
DISCUSSION.....	152
CONCLUSIONS.....	155
CHAPTER VII – CONCLUSIONS.....	157
REFERENCES	160
APPENDIX.....	195
CHAPTER II SUPPLEMENTAL.....	196
CHAPTER III SUPPLEMENTAL	220
CHAPTER IV SUPPLEMENTAL	224
CHAPTER V SUPPLEMENTAL	242
CHAPTER VI SUPPLEMENTAL	269

CHAPTER I – LITURATURE REVIEW

Wheat and Food Security

Wheat is one of the three most important cereal crops globally, with annual production exceeding 700 million tons (FAO, 2023). Its global importance is largely due to its broad adaptability, with cultivation spanning from 67°N to 45°S (Shewry, 2009). Approximately 95% of wheat produced is bread wheat (*Triticum aestivum* L.), which is primarily used for baked goods, while the remaining 5% is durum wheat (*Triticum durum*), used in pasta and semolina products. Additional specialty wheats, such as emmer (*Triticum dicoccum*) and spelt (*Triticum spelta*), are grown on a smaller scale (Allan, 1987).

In the United States, five major wheat classes are cultivated: hard red winter, hard red spring, soft red winter, white wheat, and durum wheat. The distribution of these classes is shaped by environmental adaptation and market demands. For example, spring wheat is typically grown in colder, drier regions (Allan, 1987). Beyond its agronomic importance, wheat is a key nutritional resource, providing protein, B vitamins, dietary fiber, and various phytochemicals (Shewry & Hey, 2015).

Wheat Stem Sawfly

The wheat stem sawfly (*Cephus cinctus* Norton; WSS) is a native insect pest that significantly impacts common and durum wheat and other cereal crops, including barley (*Hordeum vulgare*) and triticale ($\times$ *Triticosecale* Wittmack) (Beres et al., 2012). Historically a pest of spring wheat in the Northern United States since the 19th century, WSS expanded its impact to winter

wheat systems in the 1980s by adapting its emergence timing to align with winter wheat phenology (Peirce et al., 2022).

Most of the WSS life cycle occurs within the stem, making them difficult to manage. Female WSS oviposit eggs within the stem, where larvae develop by feeding on parenchymal tissue and other larvae until a single larva remains. This larva migrates to the base of the stem, forms a hibernaculum (“stub”), and overwinters before emerging the following spring (Ainslie, 1920; Beres et al., 2007). As plants senesce, larval feeding and stub formation weakens stems, leading to lodging (“cutting”), which complicates harvest and contributes to yield loss. Combined losses from lodging and feeding damage are estimated at approximately \$350 million annually in North America (Beres et al., 2011).

Management Strategies for Wheat Stem Sawfly

Since WSS larvae develop within stems, conventional insecticide applications are largely ineffective, necessitating alternative management strategies. Cultural practices such as trap cropping exploit WSS host preference by planting susceptible cultivars alongside resistant ones to divert infestation (Beres et al., 2009). Similarly, interseeding and border-row strategies aim to reduce overall field damage by distributing preferred and non-preferred hosts (Weaver et al., 2009).

Mechanical practices, including swathing and windrowing, can reduce stem cutting but are labor-intensive and less effective than other approaches (Weaver et al., 2005). Biological control through parasitoids, particularly *Bracon cephi* and *B. lissogaster*, has shown promise (Rand et al., 2017). Although tillage can reduce overwintering populations by burying stubble (Portman et al., 2018), it is generally discouraged in dryland systems due to negative impacts on soil moisture

retention, soil structure, and beneficial parasitoids (Al-Kaisi et al., 2002, 2004). Among all strategies, host plant resistance remains the most sustainable and widely adopted approach. Resistance mechanisms are broadly categorized as antixenosis (non-preference) and antibiosis (reduced host suitability) (Vänninen, 2005).

Genetic Resistance to Wheat Stem Sawfly

Currently, the most widely utilized genetic resistance to WSS in released cultivars comes from the solid stem trait associated with the *SStI* locus on the long arm of chromosome 3B (Cook et al., 2004a) which was originally derived from the Portuguese landrace ‘S-615’ (Larson & Macdonald, 1966). A homeologue for the 3B locus has been identified as *Qssmsub-3DL* on the long arm of chromosome 3D (Lanning et al., 2006). In non-solid stem wheat, the center of each stem is comprised of ground tissue made up of parenchyma cells during development (Ainslie, 1920), which normally break down during maturation. These undifferentiated parenchyma cells make up the pith, and leave a central cavity called the internodal lacuna in the wheat stem (Kebrom et al., 2012; Kirby, n.d.). In solid and semisolid wheat cultivars, these parenchyma cells do not break down entirely to form the hollow stem, which increases mortality of WSS eggs through mechanical interference (Beres et al., 2012; Holmes & Peterson, 1960, 1961) increases mortality of eggs and neonates shortly after hatching due to increased stem moisture (Holmes & Peterson, 1964), and hinders the ability of WSS larvae to move about the stem and feed (Evert et al., 2006; Kemp, 1934).

The decreased damage to the stems experienced by semi-solid stem or solid stem plants reduces lodging caused by WSS damage. The causal polymorphism conferring stem solidity at the *SStI* locus was identified as a *TdDof* transcription factor and is the result of increased expression

due to copy number variation (Nilsen et al., 2017). It was found that at least three copies of the *TdDof* transcription factor confer solid stems in cultivars, and that a unique 25 base pair insertion sequence was present between each copy of the transcription factors but was entirely missing from hollow stem cultivars. Increased expression of *SStI* caused 76% of variation for stem solidness of cultivars (Cook et al., 2004a).

However, variation in resistance among lines carrying *SStI* suggests additional genetic complexity. For example, Talbert et al. (2014) demonstrated that cultivars with similar solid stem expression can differ in WSS resistance, indicating the influence of alternative haplotypes and additional loci such as a quantitative trait locus (QTL) on chromosome 4A. Subsequent work by Cook et al. (2017) identified 26 haplotypes for the *SStI* locus using nine kompetitive allele specific PCR (KASP) assays. Three of these alleles included the solid stem allele most used in breeding programs derived from S-615, the Conan allele, and the hollow stem allele. All other alleles identified were novel to the study and indicate that additional haplotypes from more diverse wheat germplasm could be used as alternative forms of resistance.

Beyond *SStI*, additional loci contributing to stem solidity have been identified. Minor QTL have been reported on chromosomes 2D and 5A in common wheat (Cook et al., 2017), chromosome 3AL in durum wheat (*Triticum turgidum* L.) (Varella, Zhang, et al., 2019) and chromosome 1B and 5D in hexaploid wheat landraces and spring wheat lines (Sherman et al., 2010; Varella et al., 2015; Varella, Weaver, et al., 2019a). Furthermore, haplotype variation at the 3B locus has been characterized, including the Reeder allele (*Qss.msub-3BLa*), which confers a hollow stem; the Choteau allele (*Qss.msub-3BLb*), which maintains stem solidity throughout development; and the Conan allele (*Qss.msub-3BLc*), which confers early stem solidity followed

by a loss of pith integrity as the plant approaches maturity due to assimilate repartitioning (Cook et al., 2019a; Varella et al., 2015; Varella, Weaver, et al., 2019b).

Environmental Effects and Trade-Offs

One limitation of the solid stem trait is its sensitivity to environmental variables (Beres et al., 2012), including temperature, light intensity, moisture, and precipitation. Lower temperatures and light intensity (Holmes, 1984), as well as increased moisture and precipitation, have been shown to decrease stem solidness (Platt & Farstad, 1946). Additionally, the agronomic trade-offs associated with the solid stem trait are not fully understood. It has been hypothesized, and demonstrated in some studies, that solid stem wheat cultivars may yield less than hollow stem cultivars due to the repartitioning of assimilates to stem tissues (Sherman et al., 2015; Weiss & Morrill, 1992). This observation aligns with source–sink and resistance trade-off theories in crop improvement, which suggest that gains in one trait may come at the expense of others (Dwivedi et al., 2021).

However, reports on the relationship between stem solidity and grain yield are inconsistent. Several studies have found no significant genetic or phenotypic correlation between grain yield and stem solidity (Cook et al., 2004a, 2019a; Hayat et al., 1995; Sherman et al., 2015), whereas McNeal & Berg (1979) reported a negative correlation. These uncertainties highlight the need for further characterization of solid stem resistance, as well as the identification of alternative resistance mechanisms, to improve field performance and economic outcomes for growers.

Additional Wheat Stem Sawfly Genomic Resistance

Additional studies have identified resistance loci independent of the solid stem trait and the *SStI* locus across diverse germplasm, including hollow stem lines, spring wheat, and landraces. Quantitative trait loci analyses have identified antixenosis loci such as *Q_{wss.msub-2D}* on chromosome 2D and *Q_{wss.msub-4A.1}* on chromosome 4A, which reduce oviposition and plant attractiveness, respectively (Varella et al., 2017). Additionally, Sherman et al. (2010) identified QTL on chromosomes 1B, 2D, 3B, 4A, and 5B associated with cutting and infestation in a spring wheat recombinant inbred line population. Similarly, Varella et al. (2015) reported a cutting locus on chromosome 4A, along with larval mortality loci on chromosomes 2A, 3A, and 5B. Further work by Varella et al. (2019) identified loci associated with infestation, larval mortality, and cutting on chromosomes 1B, 4A, 4B, and 6A in recombinant inbred populations derived from landraces.

Joukhadar et al. (2013) evaluated a population of 134 wheat genotypes, including landraces, synthetic hexaploid wheat, and elite germplasm, and identified five genomic regions on chromosomes 1D, 3BL, 5BL, 6BS, and 7AS associated with the number of cut stems. Marker–trait associations on 3BL and 5BL were likely linked to the *SStI* locus and the major vernalization locus *VrnB1*, respectively, while the remaining three associations were novel.

Tools and Strategies for Breeding

Genomic Selection

Genomic selection (GS) uses genomic estimated breeding values (GEBVs), derived from genome-wide molecular markers and historical phenotypic data, to select lines based on their

additive genetic merit for traits of interest (Ganal et al., 2019; Larkin et al., 2019). These GEBVs enable selection in genotyped but unphenotyped populations for traits such as disease resistance or grain yield (Meuwissen et al., 2001). By incorporating markers across the genome, GS captures both small- and large-effect alleles, improving prediction accuracy and accelerating genetic gain by enabling earlier selection within breeding cycles (Newell & Jannink, 2014).

Genomic prediction (GP) has been successfully applied to quantitative traits in wheat, with varying levels of success. Multi-environment and multi-trait GS models have been developed for wheat trait selection (Ward et al., 2019), offering potential for improving flowering time and yield component traits in breeding programs.

Non-frequentist approaches, like Bayesian models, have gained popularity in GP. Unlike frequentist models, which assume normally distributed marker effects centered near zero, Bayesian models incorporate prior distributions and estimate posterior distributions of marker effects. For example, BayesA assumes a common prior for all markers but allows for shrinkage of small-effect markers (de Almeida Filho et al., 2019; Meuwissen et al., 2001), while Bayesian LASSO (BayesLASSO) assumes a Laplace distribution, making it more flexible for traits influenced by large-effect loci (Kärkkäinen & Sillanpää, 2012). These models have been successfully applied to improve predictive ability across traits and species (Meher et al., 2023; Onogi, 2023; Song et al., 2023).

Machine learning (ML) models provide an alternative framework for genomic prediction by identifying patterns within complex datasets (Singh et al., 2016). Although ML methods vary depending on the task (Pridmore et al., 2012), they generally aim to capture underlying relationships within training data. Random forest models are robust, non-parametric approaches

capable of handling complex interaction structures in high-dimensional datasets (González-Recio & Forni, 2011). Their tree-based structure and bootstrap sampling allow them to model non-linear relationships and interactions among markers (Montesinos López et al., 2022), making them well suited for genomic prediction of complex traits, including those influenced by polygenic and epistatic effects 6/17/2026 5:53:00 PM.

Genome Wide Association Studies

Genome-wide association studies (GWAS) are widely used to identify marker–trait associations (MTA) by detecting relationships between genotypes and phenotypes within a population (Visscher et al., 2017). GWAS leverages genome-wide single nucleotide polymorphisms (SNPs) to identify genomic regions associated with traits of interest. These associations rely on linkage disequilibrium between SNPs and phenotypic variation (Hamblin et al., 2011; Visscher et al., 2017).

Compared to biparental QTL mapping, GWAS offers the advantage of utilizing diverse germplasm, increasing allelic diversity and mapping resolution (Rathan et al., 2022). GWAS has been successfully applied to a range of wheat traits, including yield (Bhatta et al., 2018; F. Li et al., 2019; Sukumaran et al., 2015), morphology (Pang et al., 2020), quality (Hao et al., 2022; Lou et al., 2021), and agronomic performance (Jamil et al., 2019; Y. Liu et al., 2017). However, only one published GWAS study has specifically examined WSS resistance (Joukhadar et al., 2013), highlighting a clear need for further research across diverse germplasm pools.

Multi-omics

Multi-omics approaches including metabolomics, proteomics, transcriptomics, and phytohormone analyses are increasingly used to elucidate molecular mechanisms in plants including plant defenses ([Kärkkäinen & Sillanpää, 2012](#)), phenology ([J. Liu et al., 2021](#)), stress responses ([X. Chen et al., 2022](#); [Ramadan et al., 2026](#)), development ([Fenn & Giovannoni, 2021](#)), and yield related traits ([Daba et al., 2020](#)).

These approaches have also been applied to understanding WSS resistance in both solid and non-solid stem wheat. For solid stem wheat, [Lavergne et al. \(2020\)](#) evaluated four wheat genotypes—‘Hatcher’ (PI 638512) ([Haley et al., 2005](#)), ‘Denali’ (PI 664256) ([Haley, Johnson, Peairs, Stromberger, Hudson, Seifert, Kottke, Valdez, Rudolph, Martin, et al., 2012](#)), ‘Conan’ (PI 607549), and ‘Reeder’ (PI 613586)—using integrated proteomic and metabolomic analyses under WSS infestation. Significant differences in both metabolomic and proteomic profiles were observed between infested and control treatments across all cultivars. Several pathways potentially contributing to resistance were identified, including detoxification, proteinase inhibition, systemic signaling, and calcium (Ca²⁺) flux.

Similarly, [Hager et al. \(2025\)](#) investigated these mechanisms using the hollow-stem cultivar ‘Reeder’, the semi-solid cultivar ‘Conan’, and derived near-isogenic line (NIL) populations. Differences in metabolite abundance were observed among genotypes with hollow stems, early solidness, and fully solid stems across multiple developmental stages. Across these groups, pathways related to phenylpropanoid biosynthesis, carbohydrate metabolism, and programmed cell death were consistently associated with WSS resistance.

Additional studies have examined non-solid stem resistance mechanisms. Lavergne et al. (2020) also reported significant metabolomic and proteomic differences in diverse WSS-infested cultivars, identifying pathways related to detoxification, proteinase inhibition, systemic signaling, and Ca^{2+} flux. Hager et al. (2024) further explored this using two NIL populations derived from ‘Reeder’ and ‘Conan’. In hollow stem NILs, genes involved in lignin biosynthesis, such as caffeic acid O-methyltransferase (COMT; *TraesCS3B02G612000*), and cell wall modification, such as pectin methylesterase (PME; *TraesCS3B02G610100*), were upregulated during early growth stages in lines carrying the 3BLa allele.

CHAPTER II - IDENTIFYING MARKER-TRAIT ASSOCIATIONS FOR WHEAT STEM SAWFLY RESISTANCE

INTRODUCTION

Wheat stem sawfly (*Cephus cinctus* Norton) (WSS) is a native grass-feeding insect pest of many agronomically relevant grasses including common wheat (*Triticum aestivum* L.), durum wheat (*Triticum durum*), barley (*Hordeum vulgare*), and triticale ($\times$ *Triticosecale* Wittmack) (Beres et al., 2012). Wheat stem sawfly has been noted as a pest of spring wheat in the Northern United States (US) since the 19th century. However, due to changes in climatic patterns, WSS began emerging earlier in the season in the latter 20th century and have adapted to regularly infest hard winter wheat of the Great Plains region (Peirce et al., 2022). Currently, the annual estimated financial loss due to WSS in the Great Plains and Canada is approximately \$350 million United States dollars (Beres et al., 2011).

Most of the WSS life cycle occurs within the stem of the host, making cultural or chemical control difficult (Beres et al., 2011). In late spring, WSS oviposit their eggs within the stem (Ainslie, 1920). Once hatched, the larvae feed on the parenchymal tissues of the plant while also cannibalizing other larvae until one larva remains. The remaining sawfly larva then eats its way towards the base of the stem where it creates a hibernaculum, or “stub” to pupate and overwinter. During the formation of the hibernaculum, the larva will chew the surrounding stem and fill the opening with frass (Ainslie, 1920; Beres et al., 2007). Once infested wheat has senesced, the weakened stems are predisposed to lodging which disrupts conventional agricultural practices and reduces yield.

Currently, the most widely utilized genetic resistance to WSS in released wheat cultivars is derived from the *SStI* locus on the long arm of chromosome 3B which confers the “solid stem” phenotype (Cook et al., 2004a; Nilsen et al., 2017, 2020). Likewise, there is a less frequently utilized homologue to *SStI*, *Qssmsub-3DL*, located on the long arm of chromosome 3D (Lanning et al., 2006). In non-solid stem wheat, the center of each stem is comprised of ground tissue consisting of parenchyma cells which normally break down during maturation (Ainslie, 1920). These undifferentiated parenchyma cells constitute the pith. When the pith inside the wheat stem breaks down it leaves a central cavity called the internodal lacuna (Kebrom et al., 2012; Kirby, n.d.). In solid and semi-solid stemmed wheat cultivars, these parenchyma cells do not break down entirely which increases the mortality of WSS eggs through mechanical interference (Beres et al., 2012; Holmes & Peterson, 1960, 1961), increases the mortality of eggs and neonates through increased stem moisture (Holmes & Peterson, 1964), and hinders the ability of WSS larvae to move about the stem and feed (Evert et al., 2006; Kemp, 1934). The decreased damage to semi-solid stem or solid stem plants reduces lodging caused by WSS damage, thus mitigating some yield losses (Beres et al., 2007).

The molecular mechanism conferring the *SStI* solid stem phenotype is attributed to increased expression of the gene *TdDof* (*TRITD3Bv1G280530*) (Nilsen et al., 2020). Nilsen et al (2020) demonstrated that at least 3 copies of the *TdDof* transcription factor confer the solid stem phenotype, and that a unique 25 base pair insertion sequence was present between each gene copy. Increased expression of *SStI* explained 76% of the variance for stem solidness in a doubled haploid winter wheat population derived from a solid stem by hollow stem cross (Nilsen et al., 2020). The solid stem phenotype conferred by *SStI* is also influenced by environmental variables such as temperature, light intensity, moisture, and precipitation (Beres et al., 2012). Lower temperatures

and light intensity (Holmes, 1984) as well as increased moisture and precipitation were found to decrease stem solidness (Platt & Farstad, 1946). There are mixed reports about the effects of the solid stem trait on grain yield (Cook et al., 2019a; Sherman et al., 2015)

It is crucial to find additional sources of WSS resistance, especially those not related to solid stem. Important sources of non-solid stem resistance could be related to host non-preference or other agronomic traits such as straw strength. It is also important to improve our understanding and consistency of solid stem expression. Increasing sources of genetic resistance to WSS may enhance the overall stability and durability of WSS resistance (MacIntosh, 2019). In prior studies, Talbert et al. (2014) used two wheat recombinant inbred lines (RILs), ‘Scholar’ and ‘Conan’, that show varying resistance to WSS despite both having the solid stem *SStI* locus solid allele. It was found that while both varieties expressed solid stems, Conan had lower infestation and cutting scores than Scholar, despite having more hollow stems than Scholar. The authors found that these two lines had different alleles of the *SStI* locus, which could affect WSS resistance; this difference in WSS resistance was also compounded by a separate quantitative trait locus (QTL) on chromosome 4A. The QTL on 4A (wmc161) showed little effect on solid stem expression, but appeared to provide additive resistance with 20% reduced cutting in infested stems in the ‘Conan’ haplotype (Talbert et al., 2014).

Cook et al. (2017) identified 26 haplotypes for the *SStI* locus using nine kompetitive allele specific PCR (polymerase chain reaction) (KASP) assays. Three of these differing haplotypes included the solid stem *SStI* allele most used in breeding programs derived from S-615 (the Conan allele) and the hollow stem allele. All other haplotypes identified were novel to the study and this suggests that novel recombination events and haplotypes from diverse wheat germplasm could be used as alternative forms of resistance.

Genome-wide association studies (GWAS) are a commonly used method for analyzing the genetic architecture of traits (Korte & Farlow, 2013). At the time of this publication, there is only one peer-reviewed article that outlines a GWAS for WSS resistance (Joukhadar et al., 2013). Joukhadar et al. (2013) used a population of 134 wheat genotypes that included landraces, synthetic hexaploid wheat, and some elite germplasm and identified five genomic regions on chromosomes 1D, 3BL, 5BL, 6BS, and 7AS associated with number of cut stems. The marker-trait associations (MTA) identified on 3BL and 5BL were identified as most-likely linked to the *SStI* locus and the major effect vernalization locus *VrnBI*, respectively. However, the three remaining MTA were novel to the study. While there are multiple QTL studies for WSS resistance (Cook et al., 2019a; Sherman et al., 2010; Varella et al., 2015; Varella, Weaver, et al., 2019a; Varella, Zhang, et al., 2019), comparing these existing results to a GWAS on a new population of elite and agronomically viable lines could potentially assist in identifying consistent QTL for WSS resistance across different populations or uncover previously unidentified MTAs.

The Colorado State University (CSU) Wheat Breeding and Genetics program has collected trait information on nurseries screened for WSS resistance including both cutting scores (CUT) and stem solidness ratings (SOLID) since the 2014 field season. While the solid stem phenotype conferred by *SStI* on chromosome 3BL is present in the program, work has been ongoing to further characterize genomic regions related to solid stem and non-solid stem resistance within the program. It has been hypothesized that genetic variation for solidity in the CSU germplasm is oligogenic and that non-solid stem resistance exists in the CSU germplasm because of the observed variation in levels of cutting and solidity amongst cultivars with and without *SStI* (Supplementary Table 2.1). In the following study, 2,503 individuals from the CSU hard winter wheat germplasm were utilized in a GWAS to elucidate the genetic architecture of WSS resistance in CSU material

and to investigate allelic combinations (ACs) of MTAs to understand their effect on WSS resistance traits.

MATERIALS AND METHODS

Germplasm

The population used in this study included 2,503 lines from the CSU wheat breeding and genetics program that had unbalanced phenotypic data for stem solidness or WSS cutting damage across many years. All data was collected between 2014 and 2025 and totaled 19,436 observations in 45 unique site-year environments (Table 2.1).

Table 2.1. The number of unique genotypes collected within each year and the number of environments that are represented within the year.

Year	Number of Lines	Environments
2014	237	2
2016	345	2
2017	689	2
2018	643	3
2019	550	3
2020	941	4
2021	788	5
2022	1,135	3
2023	1,298	7
2024	2,555	8
2025	2,321	6

Data Collection

Observations were collected in Sidney, NE, Hemingford, NE, Grant, NE, Mead, NE, Orchard, CO, New Raymer, CO, Roggen, CO, Akron, CO, Genoa, CO, and Fort Collins, CO (Supplementary Table 2.2). Heading date (HD) was defined as the number of days from January 1st (Julian day) that half the wheat heads in a plot had fully emerged from the flag leaf. Cutting (CUT; i.e., lodging due to WSS damage) was scored on an ordinal visual scale ranging from one to nine where one was no visible cutting and nine was completely cut. Solidity (SOLID) was rated

by collecting 20 stems randomly from each plot. A cross section was taken from each internode of the stem and rated on a visual ordinal scale of solidness from one to five where one was completely hollow and five was completely solid. An average is then taken across all internodal segments present on each stem from five to 25 where five is completely hollow throughout the whole length of the stem and 25 is completely solid throughout the whole length of the stem. A maximum of four internodal ratings were taken from each stem. Ratings for stems with less than four internodal regions were adjusted based on number of internodal ratings to meet the equivalent 5 to 25 scale. The rating for each plot is then taken by averaging the stem solidness across all 20 stems. Over all years, there were 7,048 data points for HD, 9,475 for CUT, and 2,913 for SOLID.

Genotyping

Lines originating from CSU were whole genome sequenced via genotyping-by-sequencing (GBS). Ten seeds were planted for each line and roughly one cm tissue samples were collected from each plant for deoxyribonucleic acid (DNA) extraction. The DNA was extracted and quantified from the samples using ThermoFisher's MagMax plant DNA and PicoGreen kits (ThermoFisher Scientific, Waltham, Massachusetts, USA), respectively. Extracted DNA was normalized to a concentration of 20 ng μL^{-1} . Sequencing libraries were prepped following the protocol from Poland et al. (2012) and sequenced on a NovaSeq 6000 sequencer (Illumina, San Diego, California, USA) at 384-plex density per lane. Reads were then aligned to the RefSeq Chinese Spring wheat reference sequence v2.1 (Appels et al., 2018) using the burrow-wheel aligner (H. Li & Durbin, 2009). The reads obtained from the alignment were processed using the TASSEL 2.0 standalone pipeline (Glaubitz et al., 2014).

All single nucleotide polymorphisms (SNPs) were filtered using the following parameters: SNPs with greater than 20% missing data were removed; SNPs with greater than 10%

heterozygosity were removed; SNPs with a read depth of zero or greater than 100 were removed; SNPs with a site allele frequency less than 5% or greater than 95% were removed, and unaligned or indel SNPs were removed; Only biallelic SNPs were retained. After filtering, 22,207 high-quality SNPs remained. Missing data was imputed using Beagle V 5.4 (Browning et al., 2018). A synthetic wheat biparental cross between ‘W7984’ and ‘Opata’ was used to estimate the recombination distance-based map for imputation due to its established use for high-density mapping (Gutierrez-Gonzalez et al., 2019).

Statistical Analysis

Best Linear Unbiased Estimates

Statistical analysis was performed in R version 4.5.1 (Yao et al., 2022) using ASReml-R version 4.1.0.160 (Butler et al., 2023). Best linear unbiased estimates (BLUEs) were obtained for all traits for each year-location environment using the following mixed linear univariate model:

$$y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij}$$

Where y_{ij} is the observed phenotype; μ is the overall mean; τ_i is the fixed effect of the i^{th} genotype; β_j is the random row by column effect of the trial design; and ε_{ij} is the residual error term, where $\varepsilon_{ij} \sim N(0, \mathbf{I}\sigma_\varepsilon^2)$, where $\mathbf{I}$ is an identity matrix and σ_ε^2 is the residual error variance. To find the best fit model, five BLUE models were run for each environment. The first model consisted of a simple linear regression without random effects. The second model incorporated row and column random effects as identically and independently distributed (IID) variables. The third model utilized an IID row effect and an autocorrelated column effect modeled with an autoregressive (AR1) distribution where the residual error of the AR1 effect is $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2 \Sigma_r(\rho_r) \otimes \sigma_\varepsilon^2 \Sigma_c(\rho_c))$, where $\Sigma_r(\rho_r)$ and $\Sigma_c(\rho_c)$ are the correlation matrices for the row

model (order r , autocorrelation parameter ρ_r) and the column model (order c , autocorrelation parameter ρ_c). The fourth model featured an AR1 structure for rows and an IID effect with common variance for columns. Finally, the fifth model employed a row-by-column interaction with autocorrelated structures for both dimensions. The model with the lowest Bayesian info criterion (BIC) was selected for predicting BLUEs within that environment.

For each single environment model which converged and had the lowest BIC among assessed models, BLUEs were calculated for use in a second stage model to calculate adjusted means across environments. The following model was used to calculate BLUEs across environment:

$$y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij}$$

where y_{ij} is the observed phenotype; μ is the overall mean; τ_i is the fixed effect of the i^{th} genotype; β_j is the random environment effect; and ε_{ij} is the residual error term, where $\varepsilon_{ij} \sim N(0, I\sigma^2_\varepsilon)$, where I is an identity matrix and σ^2_ε is the residual error variance. After multi-environment BLUEs were calculated, there were 2,503 unique individuals remaining with 1,783 BLUEs for HD, 2,399 for CUT, and 640 for SOLID.

Heritability

Heritability of each trait across environments was estimated using a genomic best linear unbiased prediction (GBLUP) model:

$$y = X\beta + Zu + \varepsilon$$

Where β is a vector of fixed effects, X is the design matrix of fixed effects, u is a vector of genotypic values with an assumed normal distribution $u \sim N(0, G\sigma_u^2)$, where G is the genomic

VanRaden additive relationship matrix (VanRaden, 2008) and σ_u^2 is the variance of individual genotypes, $\mathbf{Z}$ is the design matrix relating genotype effects to phenotypic values ($\mathbf{y}$), and $\boldsymbol{\varepsilon}$ is the vector of random normal deviates with variance σ_ε^2 (Clark and van der Werf, 2013). the residual error with assumed distribution $\varepsilon_i \sim N(0, \mathbf{I}\sigma_\varepsilon^2)$ where $\mathbf{I}$ is the identity matrix and σ_u^2 is the variance of residual errors. The previously calculated within-environment BLUE of each trait was included in the model as the fixed effects instead of raw phenotypic information. (Endelman, 2011). Genomic heritability (h_{GBLUP}^2) was calculated as follows:

$$h_{GBLUP}^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_\varepsilon^2}$$

where $\sigma_{g(y)}^2$ is the genetic variance estimated with $\mathbf{G}$ of a trait and $\sigma_{\varepsilon(y)}^2$ is the residual error associated with that trait.

CUT, SOLID, and HD GWAS

All genome-wide association studies were performed using the statgenGWAS R package (Van Rossum & Kruijer, 2020). Genome-wide association studies were performed for both cutting and stem solidness using the following mixed model (Kang et al., 2010):

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon}$$

Where $\mathbf{y}$ is the vector of n phenotypes, $\mathbf{X}$ is the $n \times q$ design matrix for the fixed effects of the model including the intercept, covariates, and SNP of interest, $\boldsymbol{\beta}$ is the $q \times 1$ vector of coefficients of the fixed effects including the SNP effect of interest, β_{SNP} , $\mathbf{Z}$ is the $n \times t$ matrix assigning individuals to t genotypes, $\mathbf{u}$ is the vector of random effects with $\text{var}(\mathbf{u}) = \sigma_g^2 \mathbf{K}$, where $\mathbf{K}$ is a marker-based relatedness matrix, and $\boldsymbol{\varepsilon}$ represents the $n \times 1$ vector of residual variance $\text{Var}[\boldsymbol{\varepsilon}] = \mathbf{I}_n \sigma_\varepsilon^2$. The phenotypic variance-covariance matrix of the model is $\Sigma = \sigma_g^2 \mathbf{Z}\mathbf{K}\mathbf{Z}^t +$

$\sigma_e^2 \mathbf{I}_n$. Because not all lines in the population had BLUEs for each trait, the population was subset for the CUT and SOLID GWAS. Based on population structure, the top four and top two principal components (PCs) were used for CUT and SOLID models, respectively, for all CUT and SOLID models in this study. The number of PCs to include was based on scree plots for population structure for lines with CUT data and lines with SOLID data, respectively. Markers were considered significant at the Bonferroni threshold ($-\log_{10}(\text{p-value}) \geq 5.65$) or significant at an exploratory threshold that was slightly lower ($5.65 \geq -\log_{10}(\text{p-value}) \geq 4$).

Markers significant at the Bonferroni threshold from the HD, CUT, and SOLID markers were linkage disequilibrium (LD) pruned with the *gaston* R package (Perdry & Dandine-Roulland, 2015) with the maximum LD threshold set to 0.9 and max distance set to the positional difference of the furthest two identified significant markers for the respective trait GWAS for more strict pruning.

GWAS with Marker Covariates

Pruned markers from initial GWAS were used as fixed effects in CUT and SOLID models to identify loci for CUT and SOLID that were not associated with the *SStI* locus or HD. Additionally, GWAS models were run utilizing populations that separated hollow stem or solid stem individuals based on if they had the hollow stem or solid stem allele at the most significant *SStI* locus marker (S3B_844887153). Based on these parameters, six additional GWASs were performed: CUT with 5 pruned CUT markers as fixed effects (C1), CUT with the 30 pruned SOLID markers and 6 pruned HD markers as fixed effects (C2), CUT using only individuals with the hollow stem allele at the *SStI* locus (C3), SOLID with the 30 pruned SOLID markers as fixed effects (S1), SOLID with the 30 pruned SOLID and 6 pruned HD markers as fixed effects (S2), and SOLID using only individuals with the solid stem allele at the *SStI* locus (S3).

Loci significant at the Bonferroni threshold or within the exploratory threshold from all models were compiled. If multiple markers from the same region were significant in the model, the SNP with the highest $-\log_{10}(p)$ was selected as the peak marker for that region. Any significant markers that were present on the same chromosome as significant HD markers were assessed to ensure they were associated with WSS resistance and not HD. To do this, the LD was calculated between the significant CUT or SOLID marker and the nearby HD marker to confirm HD markers were not included for further analyses. Markers identified on chromosome 2D had their LD calculated between the marker closest to *Ppd-D1*(S2D_33652726) based on the gene specified in the GrainGenes database (*TraesCS2D02G079600*)(Yao et al., 2022).

Allelic Combination Analysis

Allelic combinations (ACs) were identified for lines that varied in allelic state for significant MTA. These ACs were used to assess the effect of different combined haplotypes on SOLID and CUT. To select SNPs for generating ACs, a stepwise linear regression was used to assess markers and marker effects that had significant effects on the trait of interest using ASReml-R version 4.1.0.160 (Butler et al., 2023):

$$y = X\tau + Zu + \varepsilon$$

Where y is the vector of observations, X is the design matrix of fixed effects, τ is the vector of fixed effects that included two principle components for SOLID models and four principle components for CUT models and the independent markers and combinations of all significant GWAS markers, Z is the design matrix for random effects, u is a vector of genotypic values with an assumed normal distribution $u \sim N(0, G\sigma_u^2)$, where G is the genomic VanRaden additive

relationship matrix (VanRaden, 2008) and u is the variance of individual genotypes, and ϵ is a vector of random residual effects with a distribution $\sim N(0, I\sigma^2)$.

This model was used for all significant CUT markers using CUT as the response variable and then again using all significant SOLID markers using SOLID as the response variable. The markers were recoded to have resistant (R), heterozygous (H), or susceptible (S) alleles where the resistant allele for markers significant from CUT models was determined by which allele lowered CUT and the resistant allele for markers significant in SOLID models was determined by which allele increased SOLID. For each trait, an initial model was run that included the additive and epistatic effects for all input markers and all possible combinations of marker interactions from those markers. The model used backward elimination, meaning for each step of the model, each marker and marker interaction was assessed to determine if the marker interaction effects that were significant beyond the Bonferroni threshold $\left(\frac{0.05}{n\ effects}\right)$. Significant markers and marker interactions were advanced to the next step of the model until only markers and marker interactions significant at the Bonferroni threshold remained.

AC Marker Analysis

The allelic combinations were assessed in a linear mixed model (LMM) using ASReml-R version 4.1.0.160 (Butler et al., 2023) using the same structure as the markers effects data, but with PCs and ACs being used as fixed effects instead of PCs and markers. These models were used to predict the response variable (CUT or SOLID) for each AC to assess which resistant/susceptible marker combinations effect resistance to WSS.

RESULTS

Phenotypic Evaluations

Trait BLUEs ranged from -0.4 to 8.7 for CUT with a mean of 4.1, 1.8 to 22 for SOLID with a mean of 11, and 143 to 156 days for HD with a mean of 150 (Table 2.2). There were multiple significant correlations amongst BLUEs with the most notable being CUT and SOLID ($r = -0.54$, $p < 0.001$), CUT and HD ($r = -0.16$, $p < 0.001$), SOLID and HD ($r = 0.17$, $p < 0.001$) (Figure 1) Heritability for CUT, SOLID, and HD was 0.32, 0.50, and 0.62, respectively.

Table 2. Summary statistics for the BLUEs of heading date (HD)(Julian days), cutting score (CUT), and stem solidness (SOLID) including number of individuals (N), mean, standard deviation (Std. Dev.), minimum (Min), maximum (Max), heritability(h^2_{GBLUP}), and the standard error for heritability(h^2 SE).

Trait	N	Mean	Std. Dev.	Min	Max	h^2_{GBLUP}	h^2 SE
CUT	2,399	4.1	1.8	-0.4	8.7	0.32	0.02
SOLID	640	11	3.6	1.8	22	0.5	0.03
HD	1,783	150	2	143	156	0.62	0.02

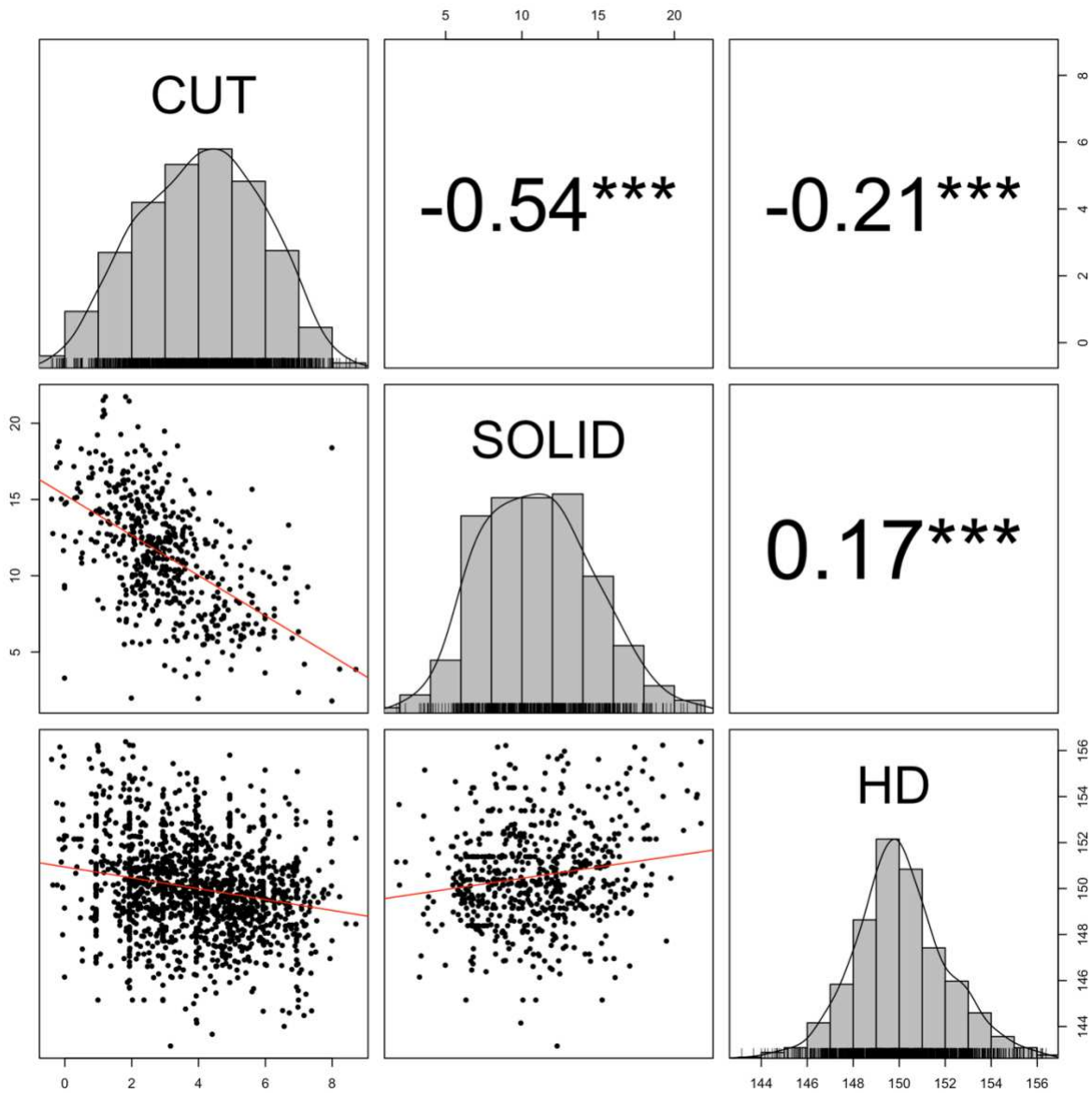


Figure 2.1. Pairs plot of univariate, multi-environmental best linear unbiased estimates of agronomic trait heading date (HD) in Julian days, and wheat stem sawfly resistance traits cutting score (CUT), and solidity (SOLID). The lower left triangle of the grid displays the scatterplot of the two traits of interest with a red linear line of best fit. The upper right triangle of the grid displays Pearson's correlation coefficient (r) of corresponding traits where significance is defined by asterisks (*=0.05, **=0.01, and ***=0.0001). The diagonal plots in the center of the figure are histograms and density plots for each agronomic and wheat stem sawfly trait.

Marker Distribution and Population Structure

A total of 22,207 high-quality SNP markers distributed across the genome were used in this study. There were 8,254 total markers in the A sub-genome, 10,607 in the B sub-genome, and 3,346 in the D sub-genome (Supplementary Table 2.3). Population structure was inferred by principal component analysis (PCA) (Figure 2.2). The first, second, and third PCs accounted for 8.20%, 5.34%, and 3.30% of the total variation, respectively. Because not all lines in the population had BLUEs for each trait, the population was subset for the CUT and SOLID GWAS. Percent total variance accounted for per PC was subjected to scree plot analysis for each subset of the population to determine the number of PCs to use as covariates in each model. All CUT models used 4 PCs as covariates while all SOLID models used 2 PCs as covariates.

PCAs for Molecular Markers

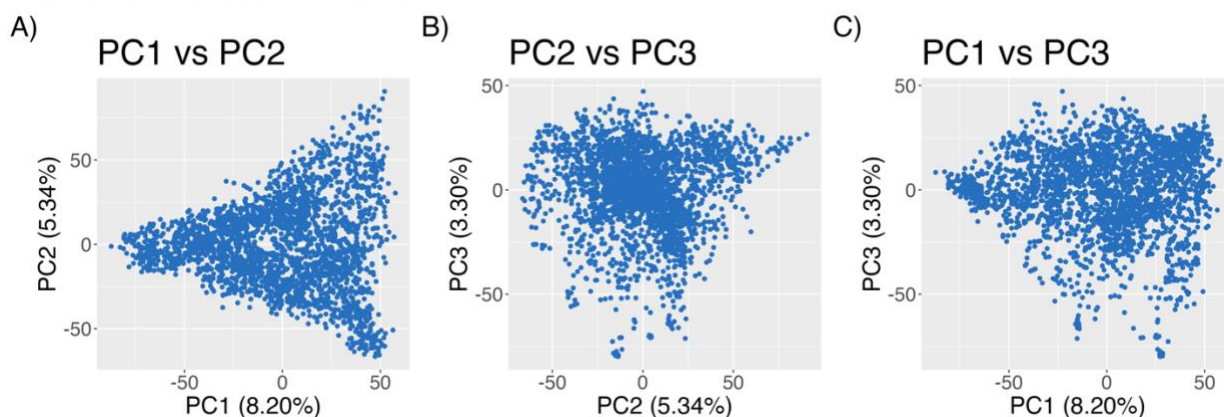


Figure 2.2. Principal component analysis plots for A) principal component 1 versus principal component 2, B) principal component 2 versus principal component 3, and C) principal component 1 versus principal component 3 to infer population structure of all 2,503 genotypes.

Cutting Score, Solid Stem, and Heading Date GWAS

Genome-wide association studies were performed for CUT, SOLID, and HD (Figure 2.3, Supplemental Figure 2.1). Markers falling above the Bonferroni significance threshold were LD pruned and used as fixed effects in additional CUT and SOLID models. A total of 77 marker-trait

associations (MTA) were detected above the Bonferroni threshold across all three models; 20 for CUT, 51 for SOLID, and 6 for HD.

All significant markers above the Bonferroni threshold for CUT and SOLID were associated with the solid stem trait on chromosome 3B. The peak SNP above the Bonferroni for CUT was S3B_844992127 which had an effect size of -0.54 on CUT. Mean CUT score for the minor allele was 3.05 and mean CUT for the major allele was 5.02. The peak marker for the locus above the Bonferroni for SOLID was SNP S3B_844887153. The marker had an effect size of 1.74 and a mean SOLID score of 11.79 for the minor allele and 6.77 for the major allele. There were two different loci above the Bonferroni for HD on chromosomes 7D and 7A. The marker on chromosome 7D was S7D_70444241 which had a mean of 150 Julian days for the minor allele and a mean of 151 Julian days for the major allele and had an effect size of -0.51 days. The marker on chromosome 7A was S7A_80297844 which had an effect size of -0.49. After LD pruning, 5 MTA for CUT, 30 for SOLID, and 6 for HD remained for use in further GWAS models.

Cutting Score GWAS With Marker Covariates

For all fixed effect and specified population (all solid stem or hollow stem individuals), the marker with the highest $-\log_{10}(\text{p-value})$ at a given peak is the marker discussed in the following sections (Table 2.3, Supplementary Table 2.4). For the C1 model (CUT with pruned CUT markers as fixed effects) ($n = 2,399$), there were no markers above the Bonferroni threshold ($-\log_{10}(\text{p-value}) \geq 5.65$) and 19 markers within the exploratory threshold ($5.65 \geq -\log_{10}(\text{p-value}) \geq 4$). Exploratory markers were identified on chromosomes 2B, 2D, 3A, 3B, 5A, and 5D (Figure 4). The markers on chromosome 3B were located in the same region as the *SStI* locus and were not considered for further analysis. The highest $-\log_{10}(\text{p-value})$ marker on chromosome 2B (S2B_5645271) had a $-\log_{10}(\text{p-value})$ of 5.29 and an effect size of 0.25 on cutting score with mean CUT of 4.35 for the

minor allele and 4.04 for the major allele. The highest $-\log_{10}(\text{p-value})$ marker on chromosome 3A (S3A_610265762) had an effect size of -0.19 on cutting score with mean CUT of 3.62 for the minor allele and 4.34 for the major allele. The highest $-\log_{10}(\text{p-value})$ marker on chromosome 5D (S5D_62660804) had an effect size of 0.31 with mean CUT of 5.19 for the minor allele and 3.99 for the major allele. Chromosome 5D also had a peak in a different region (S5D_164599835, $-\log_{10}(\text{p-value}) = 4.32$) which had an effect size of 0.22 and a mean CUT of 4.37 for the minor allele and 4.03 for the major allele. Finally, the highest $-\log_{10}(\text{p-value})$ marker on chromosome 5A (S5A_705490182) had a $-\log_{10}(\text{p-value})$ of 4.23 and an effect size of -0.24 on cutting score with mean CUT of 3.36 for the minor allele and 4.31 for the major allele.

For the C2 model (CUT with pruned SOLID and HD markers as fixed effects, $n = 2,399$), there were no markers above the Bonferroni threshold (Figure 2.4). There were 17 markers within the exploratory threshold on chromosomes 2B, 2D, 3A, 4A, 5A, and 5D. The loci on chromosomes 2B, 2D, 3A, 5A, and 5D were the same markers identified in C1. The markers on 4A (S4A_331523741) had an effect size of 0.18 on cutting score and a $-\log_{10}(\text{p-value})$ of 4.06. The mean CUT was 3.57 for the minor allele and 4.11 for the major allele.

The C3 model (CUT using only hollow stem individuals, $n = 1,212$) had no loci above the Bonferroni threshold and three markers within the exploratory threshold (Figure 2.4). One of the markers was still associated with the *SStI* locus, presumably due to high LD ($r^2 = 0.88$) between the peak *SStI* locus and the identified marker. The other two markers were associated with a peak on chromosome 2B. The more prominent of these two markers (S2B_65833439) had a $-\log_{10}(\text{p-value})$ of 4.32 and an effect size of -0.29 on cutting score with a mean CUT of 4.69 for the minor allele and 5.14 for the major allele.

Solid Stem GWAS With Marker Covariates

The S1 model (SOLID with pruned SOLID markers as fixed effects, $n = 640$) had no significant markers above the Bonferroni threshold (Figure 2.5). There were four markers within the exploratory threshold on chromosomes 3B, 4B, and 7D. The major allele for both the 4B and 2D markers had positive effects on SOLID. Marker S4B_480704002 (effect = 1.08, $-\log_{10}(\text{p-value}) = 4.55$) had mean SOLID of 13.1 for the minor allele and 10.4 for the major allele and marker S2D_654117623 (effect = 0.62, $-\log_{10}(\text{p-value}) = 4.28$) had mean SOLID of 11.6 for the minor allele and 10.7 for the major allele.

The S2 model (SOLID with pruned SOLID and HD markers as fixed effects) ($n = 640$) had no markers above the Bonferroni threshold and four markers within the exploratory threshold, which were the same markers that were identified in the S1 model (Figure 2.5). The S3 model (SOLID using only individuals with the solid stem allele) ($n = 506$) had no markers above the Bonferroni threshold and five markers within the exploratory threshold on chromosomes 2A, 3B, 4B, and 6A (Figure 2.5). All three markers on 2A (S2A_760522014, effect = 1.11, $-\log_{10}(\text{p-value}) = 4.19$), 4B (S4B_480704002, effect = 1.18, $-\log_{10}(\text{p-value}) = 4.17$), and 6A (S6A_524789877, effect = 1.38, $-\log_{10}(\text{p-value}) = 4.38$) have a beneficial alternate allele in this population with S2A_760522014 having a mean SOLID of 14.2 for the minor allele and 11.5 for the major allele, S4B_480704002 having a mean SOLID of 13.6 for the minor allele and 11.3 for the major allele, and S6A_524789877 having a mean SOLID of 14.6 for the minor allele and 11.6 for the major allele.

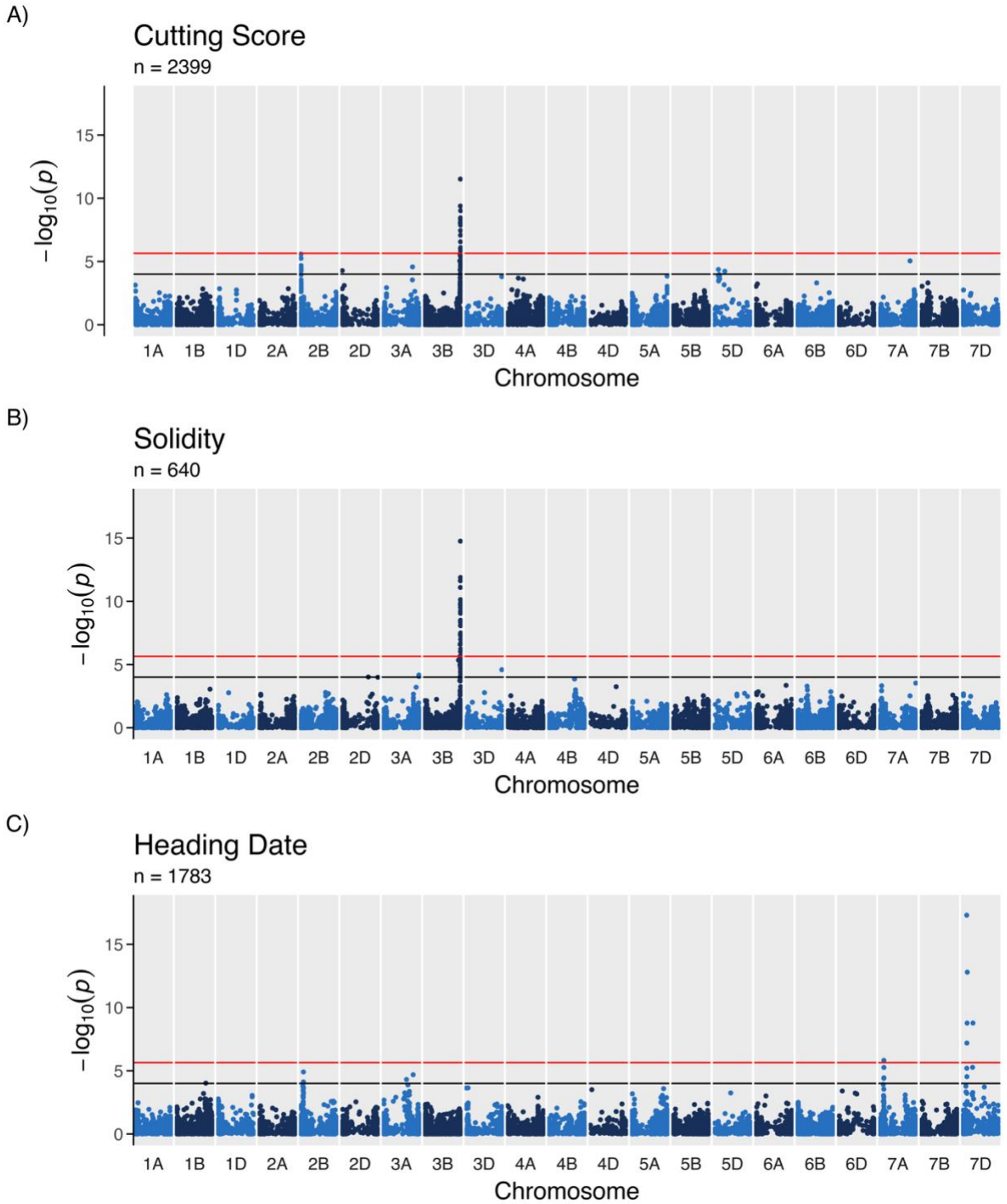


Figure 2.3. Manhattan plots for genome-wide association studies for (A) cutting score (CUT) (n = 2399), (B) solidity (SOLID) (n = 640), and (C) heading date (HD) (n = 1783). The black significance line indicates the threshold for a $-\log_{10}(p\text{-value})$ of 4 while the red significance line indicates the Bonferroni correction threshold with a $-\log_{10}(p\text{-value})$ of 5.65.

Table 2.3. Significant results for six different fixed effect GWAS and the GWAS including specific populations. Model includes the six GWAS models: CUT with 5 pruned CUT markers as fixed effects (C1), CUT with the 30 pruned SOLID markers and 6 pruned HD markers as fixed effects (C2), CUT using only individuals with the hollow stem allele at the *SStI* locus (C3), SOLID with the 30 pruned SOLID markers as fixed effects (S1), SOLID with the 30 pruned SOLID and 6 pruned HD markers as fixed effects (S2), and SOLID using only individuals with the solid stem allele at the *SStI* locus (S3), SNP^a gives the name of the marker, Alleles indicates the two letter alleles at that SNP locus, SNP MAF^b, Effect indicates the effect of the SNP marker on the trait of interest in reference to the reference allele, Effect SE is the standard error of the effect estimate.

Model	SNP	Alleles	MAF	P-Value	Effect	Effect SE	-log ₁₀ (p-value)
C1	S2B_5645271	T/C	0.24	5.13E-06	0.25	0.06	5.29
	S3A_610265762	G/A	0.35	1.11E-05	-0.19	0.04	4.95
	S5D_62660804	G/A	0.08	1.77E-05	0.31	0.07	4.75
	S5D_164599835	T/C	0.18	3.47E-05	0.23	0.05	4.46
	S2D_2896056	C/G	0.21	4.03E-05	-0.19	0.05	4.39
	S5A_705490182	C/T	0.2	5.90E-05	-0.24	0.06	4.23
C2	S2B_5645271	T/C	0.24	5.55E-06	0.25	0.06	5.26
	S3A_610265762	G/A	0.35	1.08E-05	-0.19	0.04	4.97
	S5D_62660804	G/A	0.08	1.64E-05	0.32	0.07	4.78
	S5D_164599835	T/C	0.18	2.45E-05	0.23	0.05	4.61
	S2D_2896056	C/G	0.21	2.79E-05	-0.2	0.05	4.55
	S5A_705490182	C/T	0.2	5.96E-05	-0.24	0.06	4.22
C3	S2B_65833439	A/G	0.27	4.82E-05	-0.29	0.07	4.32
S1	S4B_480704002	C/A	0.19	2.82E-05	1.08	0.26	4.55
	S2D_654117623	T/C	0.3	5.29E-05	0.62	0.16	4.28
S2	S4B_480704002	C/A	0.19	3.65E-05	1.06	0.26	4.44
	S2D_654117623	T/C	0.3	3.74E-05	0.63	0.15	4.43
S3	S6A_524789877	A/G	0.08	4.12E-05	1.38	0.34	4.38
	S2A_760522014	C/G	0.1	6.45E-05	1.11	0.28	4.19
	S4B_480704002	C/A	0.22	6.72E-05	1.18	0.3	4.17

^a Single nucleotide polymorphism, where the chromosome and position are separated by an underscore.

^b Minor allele frequency

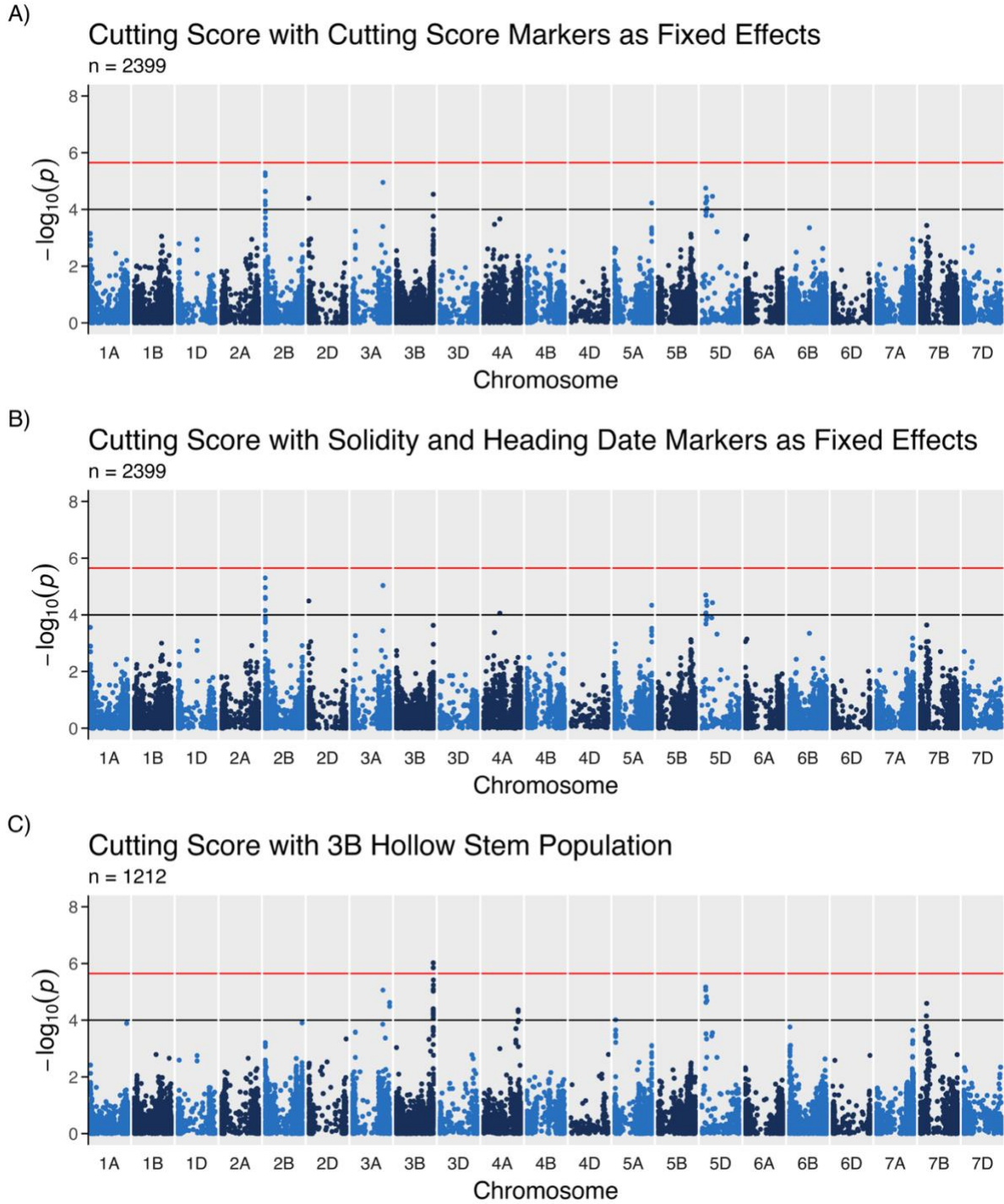


Figure 2.4. Manhattan plots for genome-wide association studies for (A) cutting score (CUT) with significant results from the original CUT GWAS ($n = 2399$), (B) CUT with significant markers from the CUT and HD GWAS ($n = 2399$), and (C) CUT using only individuals that have the hollow stem allele at the *SS1I* locus ($n = 534$). The black significance line indicates the threshold for a $-\log_{10}(\text{p-value})$ of 4 while the red significance line indicates the Bonferroni correction threshold with a $-\log_{10}(\text{p-value})$ of 5.65.

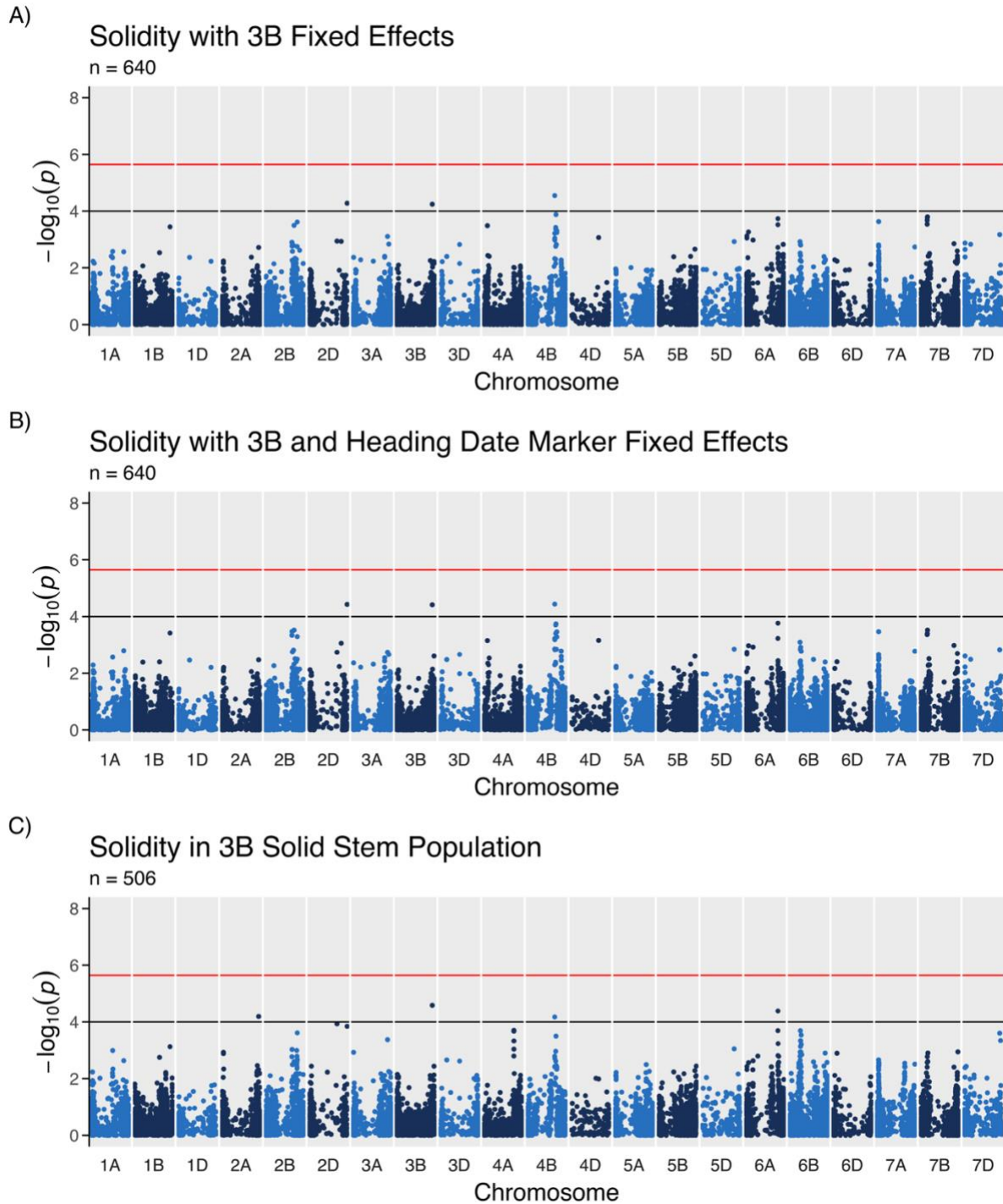


Figure 2.5. Manhattan plots for genome-wide association studies for (A) stem solidity (SOLID) with significant results from the original SOLID GWAS (n = 640), (B) SOLID with significant results from the original SOLID and heading date (HD) GWAS (n = 640), and (C) a SOLID GWAS only using a population that has the solid stem allele at the SSt1 locus (n = 506). The black significance line indicates the threshold for a $-\log_{10}(\text{p-value})$ of 4 while the red significance line indicates the Bonferroni correction threshold with a $-\log_{10}(\text{p-value})$ of 5.65.

Allelic Combination Analysis

Of all loci above the Bonferroni or within the exploratory threshold for either CUT or SOLID identified in the fixed effect GWASs, only four were on the same chromosome as a significant HD marker. These markers were on chromosome 2B, 3A, and 7A and had LD r^2 values ranging from 0.001 to 0.01, indicating the identified loci for CUT and SOLID were not associated with HD (Supplementary Table 5). The highest LD r^2 was 0.01 between the HD marker S2B_60764872 and the CUT marker S2B_5645271, indicating that markers for CUT were not in LD with markers for HD.

Seven identified CUT markers were used to develop CUT ACs and included markers from 2B, 2D, 3A, 3B, 4B, 5A, and 5D. After stepwise linear regression identified the markers that had a significant effect on CUT at the Bonferroni threshold ($0.005 \geq p\text{-value}$), there were no significant epistatic effects (Supplementary Table 2.6). Five markers remained in the AC group (S3B_844887153, S2B_564527, S3A_610265762, S5A_705490182, S5D_62660804) (Supplementary Table 2.7) with the 3B marker representing *SStI*. From these five markers, 29 unique CUT ACs were developed and were used to predict of group means and standard errors for CUT and SOLID (Supplementary Table 2.8). The five ACs with the lowest CUT predictions all had the solid stem allele at the *SStI* locus with the AC possessing all five resistance alleles having the lowest CUT prediction (CUT = 2.35, error = 0.21, n = 56). However, the sixth lowest CUT prediction had the hollow stem allele at the *SStI* locus and resistant alleles at all other loci (CUT = 3.70, error = 0.28, n = 28). Predictions were also made for SOLID using the CUT ACs, though some CUT ACs lacked data for SOLID. While there was no SOLID data for the two lines that lacked the solid stem allele at the *SStI* locus, the SOLID estimates for the five other ACs all ranged between 10.6 and 13.1 (Figure 2.6).

Three identified SOLID markers were used to develop SOLID ACs and included markers from 2D, 3B, 4B with the 3B marker representing *SStI*. After stepwise linear regression identified the markers and marker combinations that had a significant effect on solidity at the Bonferroni threshold ($0.007 \geq p\text{-value}$) at the final step of the model, there were no significant epistatic effects (Supplementary Table 2.9), but all three markers remained in the AC group (S3B_844887153, S2D_654117623, S4B_480704002). From these three markers, we were able to calculate predictions and standard errors for SOLID for eight unique ACs. The top four ACs for SOLID all had the solid stem allele at the *SStI* locus. The AC with the resistant allele at all three markers had the highest SOLID estimate (SOLID = 14.7, effect = 0.78, N = 22) (Supplementary Table 2.10). Predictions were also made for CUT using the SOLID AC groups. The CUT predictions were inversely related to the SOLID rating, but the CUT predictions only ranged from 3.16 to 4.62 (Figure 2.7).

Trait Predictions from Cutting Score Allelic Combinations

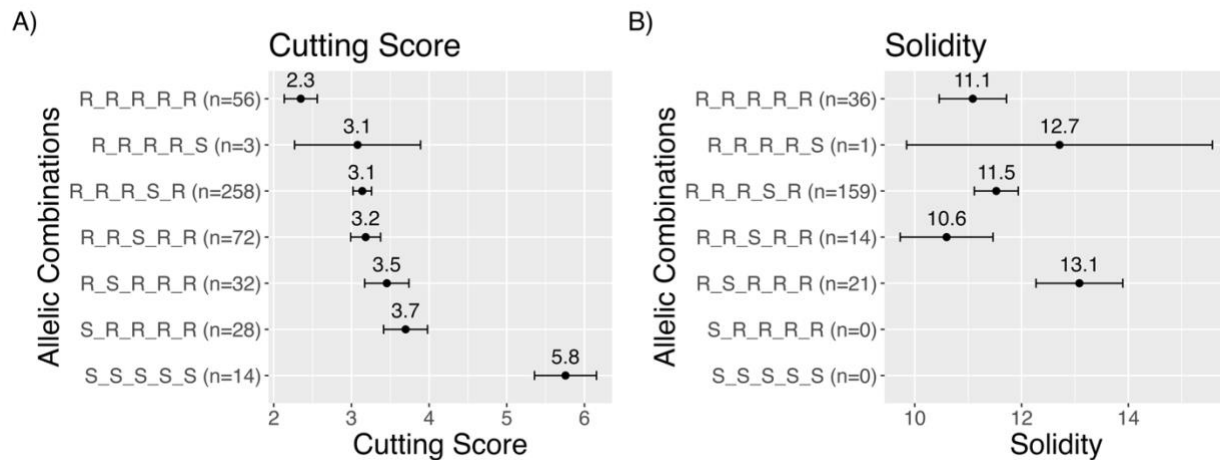


Figure 2.6. Distribution of trait predictions for allelic combinations (ACs) for cutting score (CUT). The y-axis shows the top six ACs for CUT in order from top to bottom and the susceptible AC on the bottom. The markers on the y-axis are in order with S3B_844887153 on the left followed by S2B_564527, S3A_610265762, S5A_705490182, and S5D_62660804. The x-axis shows the predicted (A) CUT and (B) SOLID of each AC with the standard errors of the prediction.

Trait Predictions from Solidity Allelic Combinations

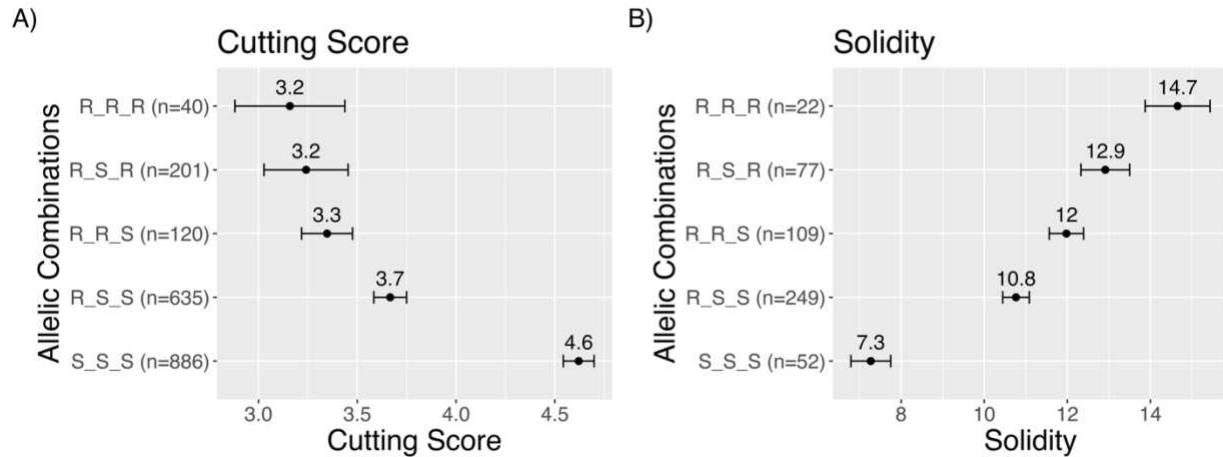


Figure 2.7. Distribution of trait predictions for allelic combinations (ACs) for solidity (SOLID). The y-axis shows the top four ACs for SOLID in order from top to bottom and the susceptible AC on the bottom. The markers on the y-axis are in order with S3B_844887153 on the left followed by S2D_654117623 and S4B_480704002. The x-axis shows the predicted (A) CUT and (B) SOLID of each AC with the standard errors of the prediction.

AC Marker Analysis

To further investigate the markers associated with each AC, individual marker effects were extracted for both CUT and SOLID from the stepwise linear regression models used to determine ACs (Table 2.4). For the five markers associated with CUT ACs, two of the resistant alleles are the major allele in the population (S2B_5645271 and S5D_62660804). The CUT marker with the largest effect on CUT was associated with the SSt1 locus (S3B_844887153) and had a CUT effect of -0.94 with a standard error of 0.12. All four other markers outside of the SSt1 marker had high SOLID standard errors relative to their estimated SOLID effect. However, the SSt1 resistant allele had an estimated effect of 3.63 on SOLID with a standard error of 0.41, which was expected. Additionally, not all resistant alleles for CUT had SOLID estimates that indicated increased SOLID in those alleles. For the markers associated with the SOLID ACs, all three resistant alleles were the minor allele in the population, and all three resistant alleles also predicted lower CUT. However, similar to the CUT markers, the estimates for CUT using the SOLID AC markers had

high standard errors relative to the CUT estimates. The only exception was the *SStI* marker (S3B_844887153), which had similar estimates for CUT and SOLID in both AC models (CUT AC model: CUT Estimate = -0.94 ± 0.12 , SOLID Estimate = 3.63 ± 0.41 ; SOLID AC model: CUT Estimate = -0.91 ± 0.13 , SOLID Estimate = 3.70 ± 0.40)

Table 2.4. Individual marker effects from the linear mixed model used to identify significant markers to include in allelic combinations (ACs). Model indicates if the marker is in the cutting score (CUT) or solidity (SOLID) AC, CUT Effect is the estimated effect of the marker on CUT, CUT SE is the standard error for the CUT effect estimate, SOLID Effect is the estimated effect of the marker on SOLID, SOLID SE is the standard error for the SOLID effect estimate, Allele indicates if the resistant allele is the major allele or the minor allele, and the Wald P-Value indicates the significance of the marker's effect on the trait for the model of interest.

Model	Marker	CUT Effect ^a	CUT SE	SOLID Effect ^a	SOLID SE	Frequency	Allele ^b	Wald P-Value
CUT	S2B_5645271	-0.57	0.1	-0.43	0.39	0.76	Major	2.85E-07
	S3A_610265762	-0.39	0.08	-0.1	0.32	0.35	Minor	2.00E-05
	S3B_844887153	-0.94	0.12	3.63	0.41	0.46	Minor	1.01E-14
	S5A_705490182	-0.5	0.11	-0.03	0.39	0.80	Minor	1.08E-05
	S5D_62660804	-0.68	0.14	0.38	0.72	0.92	Major	1.52E-05
SOLID	S2D_654117623	-0.09	0.09	1.21	0.3	0.70	Minor	2.65E-05
	S3B_844887153	-0.91	0.13	3.7	0.4	0.46	Minor	0.00E+00
	S4B_480704002	-0.06	0.16	2.2	0.5	0.78	Minor	1.02E-04

^a CUT and SOLID effect estimates indicate the effect of the resistant allele in reference to the susceptible allele

^b Allele frequency indicates the frequency of the allele in the population of all 2,503 individuals

DISCUSSION

This study utilized GWAS to identify marker associations for WSS cutting (CUT) and stem solidness (SOLID), two traits used to measure and identify WSS resistance. The solid stem locus, *SSt1*, on chromosome 3B (Nilsen et al., 2020) was identified across several models even when used as a fixed effect. However, multiple MTA identified in this study were also identified across multiple models, indicating stability in the population.

Multiple studies have identified novel QTL for WSS resistance, with some loci being identified on the same chromosomes as loci in this study (Supplementary table 2.11). Multiple studies have also identified loci on chromosome 2D. The QTL identified by Varella et al. (2017) (*Q_{wss.msub-2D}*) was associated with antixenosis while Sherman et al. (2010) identified two QTL on chromosome 2D: One associated with HD and the *Ppd-D1* locus and one only associated with CUT and infestation. However, the QTL on 2D identified by Nilsen et al. (2017) (*Q_{ss.usw-2D}*) was associated with solidity. In this manuscript, MTA for both CUT and SOLID were identified on chromosome 2D, however, neither were linked to the *Ppd1* locus. While HD MTA have been found in multiple WSS studies (Sherman et al., 2010; Varella et al., 2015), significant MTA related to CUT or SOLID in this study on the same chromosome as significant HD MTA were found to share near-zero LD, reducing the likelihood of MTA for CUT and SOLID being confounded with HD. Due to lack of available marker sequence data for all similar studies, it could not be determined if the loci identified on the same chromosomes were the same as previous literature.

Nilsen et al. (2017) identified another minor QTL for solidity on chromosome 5A (*Q_{ss.usw-5A}*), which was similar to the CUT MTA identified in this study on chromosome 5A. Additionally, Varella et al. (2019) identified a QTL on chromosome 6A, which may be related to

the 6A locus identified for solidity in the present work. Other studies identified additional MTAs that were located on the same chromosomes as CUT MTA identified in this study. Varella et al. (2015) identified QTL on chromosomes 2A and 3A associated with larval mortality and one on chromosome 5D associated with early stem solidness in spring wheat. Since early stem solidness has not been consistently measured in the population, it could not be directly associated with the same mechanism of resistance. However, early stem solidness is a known form of WSS resistance (Varella et al., 2015; Varella, Zhang, et al., 2019), which could be the driver of the identified MTA in the CSU germplasm.

In addition to previous literature related to WSS resistance, literature related to other agronomic traits such as straw strength and lodging were assessed. Multiple studies identified loci on chromosome 3A for a variety of different agronomic traits including stem strength (*QSS-3A*) (Hai et al., 2005), stem wall width and stem material strength (Ren et al., 2025), and lodging resistance (Dreccer et al., 2022). However, the closest 3A lodging QTL was over 100Mbp distal to the significant marker identified on 3A in the CUT fixed effects models. Ren et al., (2025) also identified a QTL associated with stem wall width on chromosome 6A. Additionally, loci on chromosomes 3A and 6A are associated with plant height, which can lead to reduced lodging (Curtis & Halford, 2014).

Hai et al. (2005) also identified a QTL on chromosome 2D (*QCwt-2D*) associated with culm wall thickness and explained 9.6% of phenotypic variation for stem wall thickness. For lodging, Additional QTL for lodging resistance was identified on chromosome 2D (*Excalibur_rep_c102984_157*) (Dreccer et al., 2022), and chromosomes 2B and 4B (*LI 2-4B-19*) (Ren et al., 2025) with the 2B QTL (*qSSE-2B-7*, *qLI 1-2B-21*, *qLI 3-2B-9*; *qSSE-2B-8*, *qLI 1-2B-22*, *qLI 3-2B-10*) being in the same region as the significant CUT marker from this study

(S2B_5645271). Similar MTA on chromosome 2B were identified in a study examining stem properties that included loci for culm wall thickness, stem area, pith diameter, and pith area (Ávila et al., 2021). These MTA were co-localized with previously identified QTL lodging or stem diameter according to the ‘Svevo’ genome browser QTL track. Understanding agronomic traits that improve WSS resistance is important for helping breeders combine as many beneficial resistance alleles as possible.

In this study, markers that were identified in multiple of the C1, C2, C3 and S1, S2, S3 models were assessed for their effects on CUT and SOLID. Five CUT markers were used to create the CUT AC (S3B_844887153, S2B_564527, S3A_610265762, S5A_705490182, S5D_62660804) and three markers were used to create the SOLID AC (S3B_844887153, S2D_654117623, S4B_480704002).

As expected, when assessing the combination of additive QTL, the CUT AC with all resistant alleles had the lowest predicted CUT followed by the four other ACs that had resistance alleles at four of the five loci. The AC with the 6th lowest cutting score had the hollow stem allele at the *SStI* locus and the susceptible allele at the 5D locus of the AC. However, the estimate had a higher standard error than all five of the estimates before it and the two estimates below it (predicted CUT = 3.46, error = 0.98) (Figure 2.6). The AC with the 8th lowest predicted CUT also had the hollow stem allele at the *SStI* locus and had a much lower standard error (predicted CUT = 3.70, error = 0.28).

While the solid stem allele is important to reduce CUT in this population, the two ACs with lower predicted CUT support evidence that there is non-solid stem resistance for WSS. Additionally, while the CUT ACs with all resistant alleles or four out of five resistant alleles had the lowest CUT scores, they did not have predictions for the highest SOLID ratings. As expected,

the 12 highest SOLID ratings had the resistant allele at the *SStI* locus. However, the CUT AC with all five resistant alleles only had the 9th highest predicted SOLID (predicted SOLID = 11.1, error = 0.63) (Supplementary Table 2.8) and the CUT AC with the highest predicted SOLID had all resistant alleles with the susceptible allele at the 2B locus (predicted SOLID = 13.1, error = 0.81). These results indicate that the markers in the CUT AC provide resistance to WSS outside of the solid stem trait.

Unlike the CUT ACs, the SOLID AC with all three susceptible alleles and the SOLID AC with all three resistant alleles both have the lowest and highest predicted SOLID and the highest and lowest predicted CUT, respectively (Figure 2.6). For both SOLID and CUT, the four SOLID ACs with the solid stem allele at the *SStI* locus showed the highest resistance with the SOLID and CUT predictions (Supplementary Table 2.10). These results imply that the solid stem allele is a reliable form of WSS resistance in this population across years. While there is literature identifying and evaluating superior haplotypes for the solid stem trait (Cook et al., 2017b), literature looking at allelic combinations across different chromosomes could not be identified.

The individual markers used for each AC were also assessed for their additive effect on both CUT and SOLID (Table 2.4). None of the epistatic interactions for the markers were significant, which aligns with previous research pertaining to the genetic architecture of solid stem resistance (Bainsla et al., 2020). Of the five markers in the CUT AC, two of the resistant alleles were the major allele in the population and had the second and third largest estimated effect on CUT (S5D_62660804, CUT effect = -0.68, error = 0.16; S2B_5645271, CUT effect = -0.57, error = 0.10) after the *SStI* solid stem allele (S3B_844887153, CUT effect = -0.94, error = 0.12). All five CUT markers were also assessed for their effect on SOLID, however, only the *SStI* allele had a notable estimated effect on SOLID and a standard error that was not the same size or larger than

the estimated effect. The large standard errors may be due to the reduced sample size of SOLID data compared to CUT data in the population. The inability of the CUT markers to accurately predict SOLID in the population may also indicate the presence of non-solid stem resistance associated with the MTA.

Of the three MTA used in the SOLID AC, the *SStI* marker had the largest effect on SOLID (S3B_844887153, SOLID effect = 3.7, error = 0.4). However, the markers not associated with *SStI* still had notable effects on SOLID with an effect size of a 2.2 increase (error = 0.5) in SOLID for S4B_480704002 and an effect size of a 1.21 increase (error = 0.3) in SOLID for S2D_654117623. Resistance markers for WSS were identified on chromosome 2D by (Varella et al., 2017), but it was associated with host plant attractiveness and oviposition rates instead of solidity. Like the CUT AC markers that were used to estimate SOLID, the SOLID AC markers (outside of the *SStI* locus) had low estimated effects on CUT with higher standard errors (S2D_654117623, CUT effect = -0.09, error = 0.09; S4B_480704002, CUT effect = -0.06, error = 0.16). Since the *SStI* locus constitutes a large portion of the phenotypic variance in solidity ratings of designed populations (Nilsen et al., 2020), the lower effect of the other SOLID markers on CUT may be attributed to them being overshadowed by the large effect of the *SStI* solid stem allele. While the MTA outside of the *SStI* locus are all smaller effect, considering favorable alleles for selections could have implications for improved WSS resistance.

CONCLUSIONS

Overall, the findings of this study combined with the results of previous literature indicate that there are multiple WSS resistance MTA that are stable across populations and environments. This in combination with the AC evaluation show potential for the use of GBS markers to be used for marker assisted selection (MAS) in addition with other markers such as the *SStI* locus on chromosome 3BL to select superior ACs in breeding populations. Since the solid stem trait and WSS resistance are environmentally variable year to year, it is important for breeders to identify and implement additional loci that could confer more consistent resistance across years.

CHAPTER III – ASSESSING PREDICTIVE ABILITY OF GENOMIC PREDICTION MODELS FOR WHEAT STEM SAWFLY RESISTANCE

INTRODUCTION

Wheat stem sawfly (WSS; *Cephus cinctus* Norton) is a grass-feeding insect native to North America (Ainslie, 1920) and is a significant pest in cereals, including wheat (*Triticum aestivum* L.), barley (*Hordeum vulgare*), and triticale (*Triticosecale wittmack*) (Beres et al., 2012). The WSS became a major pest of spring wheat in North America during the early 20th century and has since adapted its lifecycle to align with the winter wheat growing period in the High Plains of the United States (Peirce et al., 2022). Yield losses due to WSS are estimated at \$350 million USD annually in North America (Beres et al., 2007).

In late Spring, adult WSS emerge from their overwintering structures located in the wheat stubble of the previous year (Ainslie, 1920). Female WSS oviposit their eggs inside the stem of the host plants. Once the eggs hatch, the larvae cannibalize each other until a single WSS larva remains. The larva consumes the parenchymal tissue within the stem and compromises the structural integrity of the wheat plant prior to creating a hibernaculum to overwinter. To create the hibernaculum, sawfly larvae “cut” the stem by creating a perforated line around the base of the stem, causing the upper portion of the stem to detach from the base. This cutting causes the wheat to lodge when it is disturbed by weather variables such as wind and rain. Lodging, or cutting, due to WSS disrupts harvest practices, thus reducing grain yield. Since most of the lifecycle occurs within the stem, cultural control of WSS with treatments like non-systemic insecticides is difficult (Beres et al., 2011).

The most well characterized genetic resistance to WSS comes from the solid stem locus *SStI* on chromosome 3BL (Cook et al., 2004a). The phenotype is conferred by copy number variation of a *TdDof* transcription factor at the *SStI* locus, with a single copy resulting in a hollow stem and three copies resulting in a semi-solid or solid stem. (Nilsen et al., 2020). The *SStI* locus has been shown to explain up to 76% of phenotypic variation for solidity in a double haploid population (Cook et al., 2004a). Additionally, there is a homologue, *Qssmsub-3DL*, located on the long arm of chromosome 3D (Lanning et al., 2006).

Expression of the solid stem trait is impacted by the environment, with solidity having lower expression during cool and wet years (Beres et al., 2012; Holmes, 1984; Platt & Farstad, 1946). It may also be difficult to identify additional forms of WSS resistance, such as non-preference, beyond the solid stem trait due to other genomic regions being confounded by the large effect *SStI* locus (Lorenz & Cohen, 2012). Quantifying WSS resistance is also labor intensive, whether it is rating stems for solidity or performing experiments to determine non-solid stem forms of resistance (Buteler et al., 2009; Piesik et al., 2008; Weaver et al., 2009). Thus, alternative means of selection are warranted.

Genomic prediction (GP) utilizes phenotypic and genomic information to predict the genotypic estimated breeding value (GEBVs) of quantitative traits (Meuwissen et al., 2001). Genomic prediction aids breeders in making selection decisions in un-phenotyped populations of lines and in earlier generations prior to extensive field trials (Heffner et al., 2009, 2011). This can be particularly beneficial for traits that are time intensive or difficult to measure (Merrick & Carter, 2021). At the time of this study, literature discussing the use of GP for WSS resistance breeding has not yet been published.

There are frequentist, Bayesian, and non-parametric methods of predicting GEBVs. A common frequentist model used for GP is the genomic best linear unbiased prediction (GBLUP) model (Clark & Van Der Werf, 2013) which utilizes a genomic relationship matrix to model the covariance structure of genotype effects to estimate the additive genetic merit of an individual. The GBLUP model assumes that all markers follow a normal distribution and have equal variances (X. Wang et al., 2018; Xu et al., 2024). Variations of the GBLUP model have been used to increase predictive ability, including the inclusion of large-effect markers as fixed effects in the mixed linear model (Li et al., 2019; Rutkoski et al., 2014).

In contrast to frequentist models, which assume a gaussian distribution with marker effects near zero, Bayesian models assume different prior distributions and converging on empirical posterior distributions. Bayesian A (BayesA) for example, assumes the same prior distribution of all markers but allows for the shrinkage of small effect markers (de Almeida Filho et al., 2019; Meuwissen et al., 2001). Likewise, Bayesian LASSO (BayesLASSO) can perform well for traits with large effect loci by assuming a Laplace marker effect distribution allowing for more flexibility in their estimates (Kärkkäinen & Sillanpää, 2012). The utility of these models have been used to improve predictive ability for a variety of different traits and species (Meher et al., 2023; Onogi, 2023; Song et al., 2023).

Recently, machine learning (ML) models have been utilized for GP. Machine learning models, like random forest models (Breiman, 2001), are robust non-parametric regression models designed to handle complex interaction structures within large datasets (González-Recio & Forni, 2011). Incorporating models that can handle large effect markers as well as the complex relationships between markers could potentially increase predictive ability.

The Colorado State University (CSU) Wheat Breeding and Genetics Program began breeding for resistance to WSS in 2011, initially focusing on the introgression of *SS1* from key semi-solid stemmed cultivars such as ‘Judee’ (PI 665227) (Carlson, Berg, Kephart, et al., 2013), ‘Bearpaw’ (PI 665228) (Carlson, Berg, Stougaard, et al., 2013), and ‘Warhorse’ (PI 670157) (Berg et al., 2014). Likewise, the program has continuously collected phenotypic information for stem cutting (cutting) and solidity (solidity) from 2014 to present day. For this study, phenotypic and genetic information from 2,503 individuals were used to assess the predictive ability of cutting and solidity across genomic prediction models that differ in their assumptions of underlying trait genetic architecture. Additionally, we assessed the limitations of heritability on predictive ability for WSS resistance phenotypes.

MATERIALS AND METHODS

Germplasm

Germplasm, data collection, and genotyping for this manuscript are identical to that in (Hammers et al., 2026, manuscript submitted for publication). The population used in this study included 3,503 lines from the Colorado State University wheat breeding and genetics program that had unbalanced phenotypic data for solidity or cutting across many years. All data was collected between 2014 and 2025 and totaled 11,574 observations in 45 location-years (Table 3.1) on 2,503 individuals.

Table 3.1. The number of genotypes and sites within each year for which phenotypic data was collected for either stem cutting or stem solidness.

Year	No. of lines	No. of locations
2014	237	2
2016	345	2
2017	689	2
2018	643	3
2019	550	3
2020	941	4
2021	788	5
2022	1,135	3
2023	1,298	7
2024	2,555	8
2025	2,321	6

Data Collection

Data was collected in Sidney, NE, Hemingford, NE, Grant, NE, Mead, NE, Orchard, CO, New Raymer, CO, Roggen, CO, Akron, CO, Genoa, CO, and Fort Collins, CO (Supplemental Table 3.1). Cutting (cutting; lodging due to WSS damage) was scored on an ordinal visual scale ranging from one to nine where one is no visible cutting and nine is 90% to 100% cut. Solidity

(solidity) was rated on 20 stems collected randomly from each plot. For each stem, a cross section was taken from each internode and rated on a visual ordinal scale of solidness from one to five where one was completely hollow and five was completely solid. An average is then taken across internodes to rate each stem from five to 25 where a rating of five indicates a hollow stem throughout the entire length of the stem and 25 is completely solid throughout the entire length of the stem. The rating for each plot is then taken by averaging the stem solidness ratings across all 20 stems. Across all years and sites there were 9,475 plot datapoints for cutting, and 2,913 for solidity.

Genotyping

Lines were whole genome sequenced via genotyping-by-sequencing (GBS). Ten seeds were planted for each line and a roughly one cm tissue sample was collected from each plant for deoxyribonucleic acid (DNA) extraction. DNA was extracted and quantified from the samples using ThermoFisher's MagMax plant DNA and PicoGreen kits (ThermoFisher Scientific, Waltham, Massachusetts, USA), respectively. Extracted DNA was normalized to a concentration of 20 ng μ L⁻¹. Sequencing libraries were prepped following protocol from Poland et al. (2012) and sequenced on a NovaSeq 6000 sequencer (Illumina, San Diego, California, USA) at 384-plex density per lane. Reads were aligned to the RefSeq Chinese Spring wheat reference sequence v2.1 (Appels et al., 2018) using the burrow-wheel aligner (H. Li & Durbin, 2009). Aligned reads were processed using the TASSEL 2.0 pipeline (Glaubitz et al., 2014).

Single nucleotide polymorphisms (SNPs) that did not meet the following filtering criteria were removed: SNPs with greater than 20% missing data; 10% heterozygosity a read depth of zero or greater than 100; a site allele frequency less than 5% or greater than 95%; unaligned or indel SNPs; only biallelic SNPs were retained. After filtering, 22,207 high-quality SNPs remained.

Missing data was imputed using Beagle V 5.4 (Browning et al., 2018). A synthetic wheat biparental cross between ‘W7984’ and ‘Opata’ was used to derive the recombination distance-based map for imputation (Gutierrez-Gonzalez et al., 2019).

Statistical Analysis

Phenotypic Data Analysis

Phenotypic data was partitioned out into two independent groups: one containing all data from 2014 through 2024 for the training population (TP) and a second containing all data from 2025 for the forward-validation population (VP). All genotypes present in the TP set were masked in the VP set to reduce bias in estimates of predictive ability. After partitioning, there were 6,556 observations in the TP dataset including 4,677 for cutting and 2,073 for solidity. After filtration for genetic data availability and across-environment adjustment, there was a total of 1,776 genotyped lines with cutting data and 530 genotypes with solidity data. There were 2,787 observations in the VP dataset with 2,187 for cutting and 308 for solidity. After filtration for genetic data availability and across-environment adjustment, there were 599 and 105 genotyped individuals with cutting and solidity data in the VP set, respectively.

Statistical analysis was performed in R version 4.5.1 (R Core Team, 2025) using ASReml-R version 4.1.0.160 (Butler et al., 2023). Best linear unbiased estimates (BLUEs) were obtained for all three traits for each year-location environment using the following univariate model:

$$y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij} \quad (Eq. 1)$$

Where y_{ij} is the observed phenotype; μ is the overall mean; τ_i is the fixed effect of the i^{th} genotype; β_j is the random row by column effect of the trial design; and ε_{ij} is the residual error term, where $\varepsilon_{ij} \sim N(0, I\sigma^2_\varepsilon)$, where I is an identity matrix and σ^2_ε is the residual error variance. To find the best

fit model, five BLUE models were run for each environment. The first model consisted of a simple linear regression without random effects. The second model incorporated row and column random effects as identically and independently distributed (IID) variables. The third model utilized an IID row effect and an autocorrelated column effect modeled with an AR1 distribution. The fourth model featured an autoregressive (AR1) structure for rows and an IID effect with common variance for columns. Finally, the fifth model employed a row-by-column interaction with autocorrelated structures for both dimensions. The model with the lowest Bayesian info criterion (BIC) was selected for predicting BLUEs within that environment.

For each single environment model which converged and had the lowest BIC among assessed models, BLUEs were calculated for use in a second stage model to calculate adjusted means across environments. The following model was used to calculate BLUEs across environments:

$$y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij} \quad (Eq. 2)$$

where y_{ij} is the observed phenotype; μ is the overall mean; τ_i is the fixed effect of the i^{th} genotype; β_j is the random environment effect; and ε_{ij} is the residual error term, where $\varepsilon_{ij} \sim N(0, W\sigma_\varepsilon^2)$, where W is the weighting matrix where $W_{ii} = 1/\sigma_i^2$ and σ_ε^2 is the residual error variance.

Heritability

Heritability of each trait within and across environments was estimated using a genomic best linear unbiased prediction (GBLUP) model:

$$y = X\beta + Zu + \varepsilon \quad (Eq. 3)$$

Where β is a vector of fixed effects, X is the design matrix of fixed effects, u a vector of genotypic values with an assumed normal distribution $u \sim N(0, G\sigma_u^2)$, where G is the genomic VanRaden additive relationship matrix (VanRaden, 2008) and σ_u^2 is the variance of individual genotypes, Z is the design matrix relating genotype effects to phenotypic values (y), and ϵ is the vector of random normal deviates with variance σ_ϵ^2 (Clark and van der Werf, 2013). The residual error with assumed distribution $\epsilon_i \sim N(0, I\sigma_\epsilon^2)$ where I is the identity matrix and σ_ϵ^2 is the variance of residual errors. The previously calculated within-environment BLUE of each trait was included in the model as the fixed effects instead of raw phenotypic information. (Endelman, 2011). Genomic heritability (h_{GBLUP}^2) was calculated as follows:

$$h_{GBLUP}^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_\epsilon^2} \quad (Eq. 4)$$

where σ_g^2 is the genetic variance estimated with the genomic VanRaden additive relationship matrix, G , of a trait and σ_ϵ^2 is the residual error associated with that trait.

Genomic Prediction

GBLUP

Genomic best-linear unbiased prediction (GBLUP) was performed in ASReml-R using the GBLUP model described previously (Eq. 3). Additionally, a GBLUP plus fixed effect model was carried out using the peak marker (S3B_844887153) (Hammers et al. 2026, manuscript submitted for publication) associated with the solid stem locus *SStI* (GBLUP + *SStI*).

Bayes A and Bayes LASSO

Bayesian models were performed using the BGLR R package version 1.1.4 (De Los Campos & Perez Rodriguez, 2013) using the following continuous model:

$$y_i = \eta_i + \varepsilon_i \quad (Eq. 5)$$

Where η_i is a linear predictor for the expected value of y_i and ε_i are residuals with assumed distribution $\varepsilon_i \sim N(0, w_i^2 \sigma_\varepsilon^2)$ where w_i represents the weights specified in the model and σ_ε^2 is a residual variance parameter. The linear predictor, η_i , represents the conditional expectation function:

$$\eta = 1\mu + \sum_{j=1}^J \mathbf{X}_j \boldsymbol{\beta}_j + \sum_{l=1}^L \mathbf{u}_l \quad (Eq. 6)$$

Where μ is the intercept, $\mathbf{X}_j$ is the design matrix for predictors, $\mathbf{X}_j = x_{ijk}$, $\boldsymbol{\beta}_j$ are vectors of effects associated with $\mathbf{X}_j$, and $\mathbf{u}_l$ is the vector of random effects. Based on these assumptions, the conditional distribution of the data is:

$$p(\mathbf{y}|\boldsymbol{\theta}) = \prod_{i=1}^n N \left(\sum_{j=1}^J \sum_{k=1}^{K_j} x_{ijk} \beta_{jk} + \sum_{l=1}^L u_{li}, \sigma_\varepsilon^2 w_i^2 \right) \quad (Eq. 7)$$

Where $\boldsymbol{\theta}$ is the collection of unknowns, including the intercept, regression coefficients, other random effects, and the residual variance. The prior density of the model can be expressed as

$$p(\boldsymbol{\theta}) = p(\mu)p(\sigma_\varepsilon^2) \prod_{j=1}^J p(\boldsymbol{\beta}_j) \prod_{l=1}^L p(\mathbf{u}_l) \quad (Eq. 8)$$

Where the intercept is assigned a flat prior, meaning all parameter values are equally likely, and the variance is assigned a scaled-inverse χ^2 density $p(\sigma_\varepsilon^2) = \chi^{-2}(\sigma_\varepsilon^2 | S_\varepsilon, df_\varepsilon)$ where S_ε is the scale parameter (>0) and df_ε are the degrees of freedom (>0). For all four Bayesian models, the reciprocal of the standard error from the single-environment BLUEs estimate was used as weights

the respective genotype and environment included, environment was included as a linear predictor with a flat prior, or fixed effect ($\mathbf{u}_\beta \sim N(0, \sigma_\beta^2)$), and an $n \times m$ matrix of genotypes by markers was included as a linear predictor with each model utilizing a different prior.

The markers in the BayesA model were given a Student's t-distribution for the marker effects of the model using a mixture of normal distributions across markers $u_\beta \sim N(0, \sigma_{\beta_j}^2)$, which allows the model to utilize larger effect markers to estimate breeding values (Meuwissen et al., 2001). The non-zero marker effects for the BayesLASSO model are assumed to follow a double exponential distribution $u_\beta \sim N(0, \sigma_{\beta_j}^2)$, where the non-zero marker effects have locus specific variance (Meher et al., 2023; Yi & Xu, 2008).

Random Forest

The R package caret (Kuhn, 2008) was used to perform random forest machine learning models. Random forest models (random forest) were performed using caret as a wrapper for the R package “ranger” (Wright, 2015) which performs random forests as described by Breiman (2001). Random forests in the context of genomic prediction have been described by Montesinos López et al. (2022) and is the basis of the following methodology. In random forest models, n trees are grown that each use a different bootstrap sample of all the training data. The portion of the data that is bootstrap sampled is the hyperparameter *mtry*. In a random forest, each decision tree will take bootstrapped subsamples and randomly subsets the data for splitting tree nodes. Each decision tree is grown by recursively following this splitting step at each terminal node of the tree until the minimum node number criterion is met. Once a forest of trees has been grown, the random forest model uses the following equation:

350
351
352
353
354
355
356

357
358
359
360
361
362
363
364
365
366
367
368
369
370

$$\hat{y}_i = \frac{1}{B} \sum_{b=1}^B T_b(x_i) \quad (Eq. 9)$$

Where $\hat{y}_i$ is the predicted continuous variable and x_i is the input for i individuals, and T_b is the ensemble of trees $\{T_b\}_1^B$ for B bootstrap samples. Recursive feature elimination was used to limit the number of markers and reduce computational load. This process was performed using caret to assess what number of markers between 10 markers and $0.25 \cdot p$ where p is the number of markers were most predictive of the dataset. The training and testing datasets were then reduced to only include the markers selected by the recursive feature elimination process.

To fine-tune the model and increase predictive accuracy, different combinations of n trees and $mtry$ values were tested to decide the best model. The array of n trees to test included 200, 400, 600, 800, and 1000 trees. Values for $mtry$ were based on previous publications regarding random forest models for GP. While the standard $mtry$ value recommended by Breiman (2001) would be $\log_2(p) + 1$, $\sqrt[2]{p}$ where p is the default $mtry$ in the original randomForest R package. Others have proposed that the ideal value of $mtry$ should be $> 0.1p$ (Goldstein et al., 2010) or $mtry = p$ (Genuer et al., 2008; Goldstein et al., 2010) based on population structure. To account for the varying reports of optimal $mtry$ values for datasets with large numbers of predictors, the tuning parameters for $mtry$ included $\sqrt[2]{p}$, $0.1 p$, $0.25 p$, $0.5 p$, and p . The combination of n trees and $mtry$ hyperparameters that had the lowest model root mean square error (RMSE) for the training data was selected as the best fit model and was used in cross-validation of validation data and forward-validation on the test dataset. The optimal hyperparameters were then used for all permutations of the model. To mirror GP across environments, the results of each model were averaged across genotypes to create one prediction per genotype.

Cross-Validation Predictive Ability

Model predictive ability was assessed over 100 permutations using an 80:20 cross-validation scheme. For each permutation, multi-environment BLUEs from 2014-2024 were randomly partitioned into an 80% training set and a 20% validation set. Observations (i.e., multi-environment BLUEs) within the validation set were masked. The training set was then used to calculate genomic estimated breeding values (GEBVs) for the masked individuals. Predictive ability was calculated as the Pearson correlation coefficient (r^2) between these GEBVs and the observed multi-environmental BLUEs of the validation set. To assess statistical differences of the performances between different models, a Friedman's test and Nemenyi post hoc test was performed using the R package "PMCMRplus" version 1.9.12 (M. Liang et al., 2022; Pohlert, 2018). This procedure was also used to assess differences in model performance for forward-validation predictive ability.

Forward-Validation

To evaluate forward predictive ability, the same 80% training partitions generated during cross-validation were used to calculate GEBVs for the 2025 VP set. Predictive ability was again determined using the Pearson correlation coefficient (r^2) between the calculated GEBVs and the observed multi-environmental BLUEs of the 2025 genotypes. Both the cross-validation and forward-validation procedures were repeated across 100 permutations to sample a variety of training populations. From these permutations, the maximum, minimum, mean, and standard deviation of predictive ability were calculated to summarize performance.

Heritability Culling

Genomic heritability (h_{GBLUP}^2) was calculated for each location-year and trait using the same GBLUP model described above (Eq. 3). To assess the relationship of heritability with GP

accuracy, location-years were subset to reflect the Q1, Q2, and Q3 quartiles (e.g., 75%, 50%, or 25% percentile) of the calculated h^2_{GBLUP} for both cutting and solidity individually. The new TP designations were TP_Q1, TP_Q2, and TP_Q3. The TP were used to perform cross- and forward-validation on the 2025 VP with the same five models used previously which allowed for comparison across all models.

RESULTS

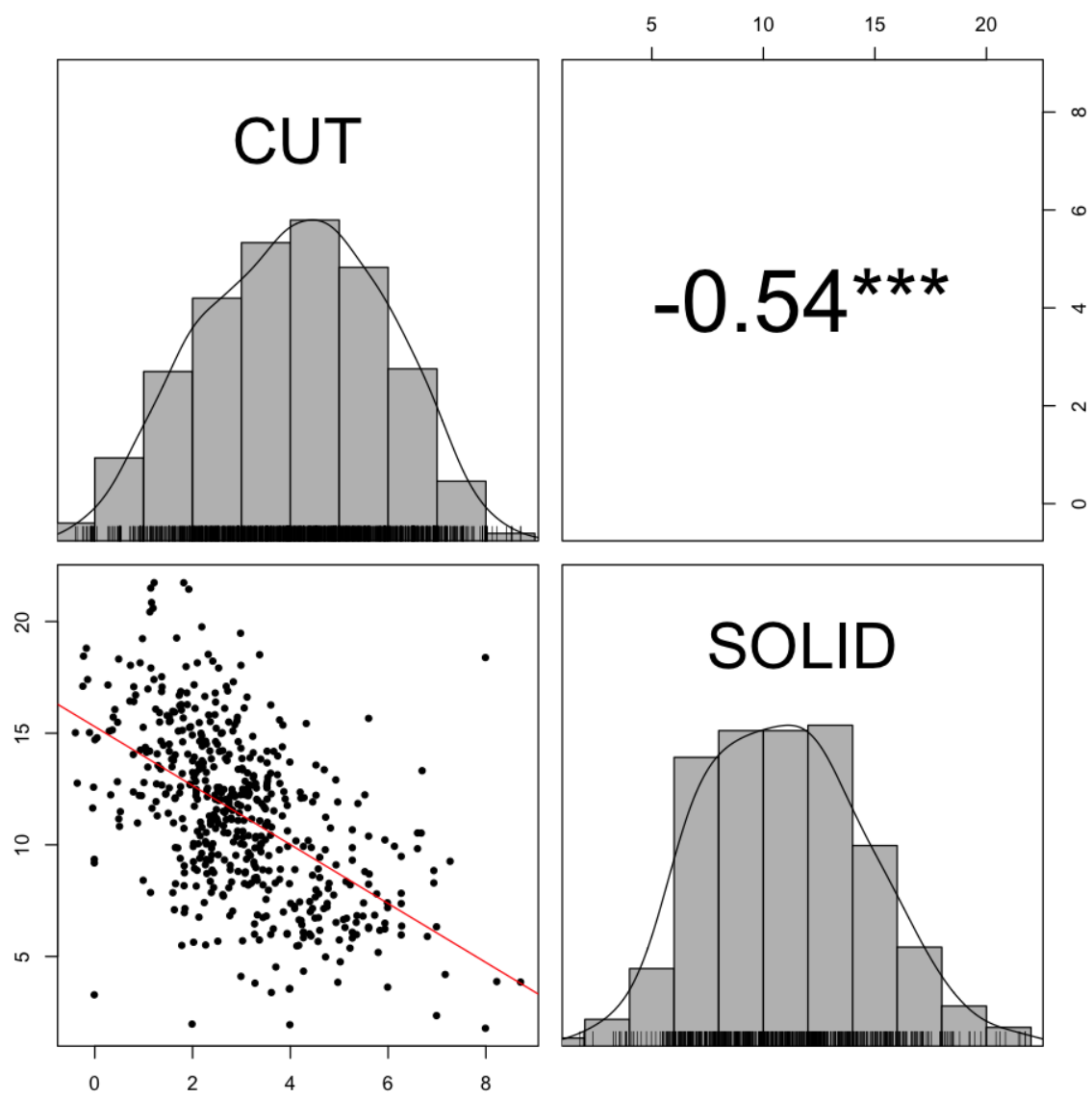
Phenotypic Analysis

Each location-year combination was considered an environment resulting in 39 total environments for the TP dataset and six environments for the VP dataset. Best linear unbiased estimates for the TP ranged from 0.39 to 9.4 for cutting, and 0.6 to 22 for solidity (Table 3.2). Best linear unbiased estimates for the VP ranged from -0.22 to 6.8 for cutting, and 4.9 to 17 for solidity. Genomic heritability (h^2_{GBLUP}) in the TP was higher for solidity ($h^2_{GBLUP} = 0.49$) than for cutting ($h^2_{GBLUP} = 0.38$). However, the highest h^2_{GBLUP} in the VP was cutting ($h^2_{GBLUP} = 0.45$) compared to solidity ($h^2_{GBLUP} = 0.31$). Phenotypic correlation across all environments between cutting and solidity was -0.60 ($p < 0.001$) (Figure 3.1).

412 **Table 3.2.** Descriptive statistics and heritability of BLUEs across all environments for the training and validation populations for cutting
413 score (cutting) and solid stem rating (solidity) where N is the number of individuals, Min is the minimum BLUE across individuals,
414 Max is the maximum BLUE across individuals, h^2_{GBLUP} is the narrow-sense heritability, and SE is the standard error of the h^2_{GBLUP}
415 estimate.

	Trait	N	Mean	Std. Dev.	Min	Max	h^2_{GBLUP}	h^2_{GBLUP} SE
Training Population	cutting	1776	4.5	2	0.05	9.5	0.32	0.02
	solidity	530	11	3.7	0.47	22	0.55	0.03
Validation Population	cutting	586	3.2	1.5	-0.22	6.8	0.42	0.04
	solidity	96	10	2.3	4.9	17	0.30	0.08

416



418
419
420
421
422
423
424
425

Figure 3.1. Pairs plots of univariate, multi-environmental best linear unbiased estimates of cutting score (cutting), and solidity rating (solidity). The lower left quadrant of the grid displays the scatterplot of the two traits of interest with a red linear line of best fit. The upper right quadrant of the grid displays Pearson's correlation coefficient (r) of the who corresponding traits where significance is defined by asterisks (*=0.05, **=0.01, and ***=0.0001). The diagonal plots in the center of the figure are histograms and density plots for each agronomic and wheat stem sawfly trait.

Genomic Prediction using All Environments

In cross-validation for cutting, all models performed similarly, with random forest having the highest mean predictive ability ($\bar{r} = 0.67$), followed by BayesA ($\bar{r} = 0.67$), GBLUP ($\bar{r} = 0.66$), BayesLASSO ($\bar{r} = 0.66$) and GBLUP + *SStl* ($\bar{r} = 0.64$) (Figure. 3.3). In cross-validation of solidity, random forest again outperformed all models, ($\bar{r} = 0.62$) followed by GBLUP ($\bar{r} = 0.53$), BayesA ($\bar{r} = 0.52$), BayesLASSO ($\bar{r} = 0.51$), and GBLUP + *SStl* ($\bar{r} = 0.44$) (Supplemental Table 3.2).

In forward-validation for cutting on the 2025 phenotypes, random forest had the highest mean predictive ability ($\bar{r} = 0.45$) followed by BayesA ($\bar{r} = 0.40$), GBLUP ($\bar{r} = 0.38$), BayesLASSO ($\bar{r} = 0.37$), and GBLUP + *SStl* ($\bar{r} = 0.32$). For forward-validation of solidity the highest mean predictive ability was observed for GBLUP ($\bar{r} = 0.64$), followed by BayesA ($\bar{r} = 0.63$), BayesLASSO ($\bar{r} = 0.61$), random forest ($\bar{r} = 0.58$) and GBLUP + *SStl* ($\bar{r} = 0.58$).

The Friedman's test determined that there were minimal differences in performance for cross-validation and forward-validation across the models. For forward-validation in the cutting models, only the random forest model performed statistically better (p-value < 0.001) than the GBLUP model. There were no other models that performed better than GBLUP for cutting or solidity.

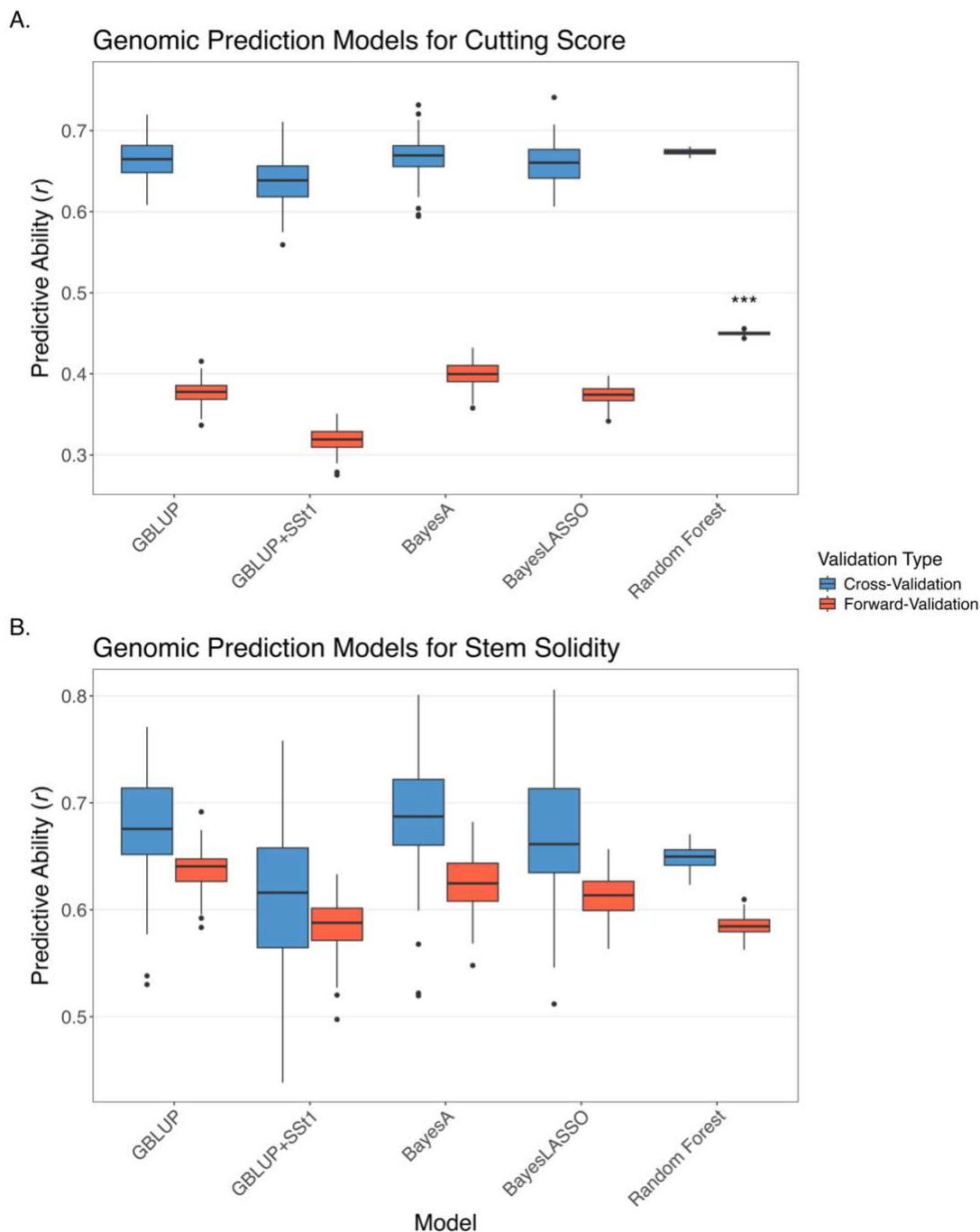


Figure 3.3. Predictive ability in cross and forward-validation of all five models for (A) cutting and (B) solidity. Predictive ability represents the Pearson's correlation coefficient (r) between the predicted value of a genotype and the BLUE calculated for that genotype. Cross-validation predictive ability is represented by the blue boxplots and forward-validation predictive ability is represented by the red boxplots. Asterisks (*) denote models that perform statistically better than the respective cross-validation or forward-validation GBLUP model using all environments and the number of asterisks denote the significance level (* = p -value < 0.05; ** = p -value < 0.01; *** = p -value < 0.001).

452 **Table 3.3.** Within environment heritabilities (h^2_{GBLUP}) and standard errors (SE) for cutting and
453 solidity.

Environment	cutting h^2_{GBLUP}	cutting SE	solidity h^2_{GBLUP}	solidity SE
2014:New Raymer	0.55	0.09	-	-
2016:Orchard	0.51	0.09	0.92	0.09
2016:New Raymer	0.51	0.10	0.00	-
2017:Orchard	0.00	-	0.00	-
2017:New Raymer	0.71	0.05	0.83	0.16
2018:New Raymer	0.57	0.07	0.75	0.07
2018:Orchard	0.54	0.07	0.67	0.09
2019:New Raymer	0.71	0.06	0.36	0.15
2019:Orchard	0.50	0.09	0.57	0.12
2020:Akron	0.49	0.07	-	-
2020:New Raymer	0.68	0.08	-	-
2020:Orchard	0.63	0.08	0.31	0.24
2021:New Raymer	0.67	0.08	-	-
2021:Orchard	0.35	0.14	-	-
2022:Akron	0.32	0.23	-	-
2022:New Raymer	0.58	0.07	0.00	0.00
2023:New Raymer	0.00	0.00	-	-
2023:Akron	0.28	0.11	-	-
2024:Akron	0.38	0.04	0.91	0.03
2024:Roggen	0.39	0.09	-	-
2024:Orchard	0.78	0.16	-	-
*2025:Akron	0.53	0.04	0.29	0.10
*2025:Genoa	0.01	0.05	-	-
*2025:Fort Collins	-	-	0.93	0.09
*2025:New Raymer	0.53	0.06	-	-
*2025:Orchard	0.37	0.02	0.27	0.11
*2025:Roggen	0.45	0.05	-	-

454 * Indicates environments included in the validation population (VP).

Genomic Prediction with Heritability Culling

Partitioning environments based on heritability resulted in threshold h_{GBLUP}^2 estimates of 0.38 for Q1, 0.51 for Q2, and 0.63 for Q3 for cutting across 19 environments. Culling training populations to include environments above each quantile threshold resulted in 14 environments being included in TP_Q1, 10 in TP_Q2, and 5 in TP_Q3. For solidity, threshold h_{GBLUP}^2 estimates were 0.27 for Q1, 0.63 for Q2, and 0.76 for Q3, which included 11 environments. Culling training populations to include environments above each quantile threshold resulted in 8 environments being included in TP_Q1, 5 in TP_Q2, and 3 in TP_Q3. (Supplemental Table 3.3).

In cross-validation, culling the population of low heritability environments had varying effects on mean predictive ability across the models. In cross-validation for cutting, model predictive ability decreased for models in TP_Q1 but increased for models in TP_Q2 and TP_Q3 (Figure 4; Supplemental Table 3.2). The exception to this trend was GBLUP+SS*tl*, which decreased each time the TP was culled ($\bar{r}_{ALL} = 0.64$, $\bar{r}_{TP_Q1} = 0.62$, $\bar{r}_{TP_Q2} = 0.60$, $\bar{r}_{TP_Q3} = 0.49$). Additionally, a total of six different models performed better than the GBLUP using the TP with all environments ($\bar{r}_{ALL} = 0.66$) including GBLUP in TP_Q2 ($\bar{r}_{TP_Q2} = 0.69$), BayesA in TP_Q2 and TP_Q3 ($\bar{r}_{TP_Q2} = 0.71$, $\bar{r}_{TP_Q3} = 0.75$), BayesLASSO in TP_Q3 ($\bar{r}_{TP_Q3} = 0.69$), and random forest in TP_Q2 and TP_Q3 ($\bar{r}_{TP_Q2} = 0.76$, $\bar{r}_{TP_Q3} = 0.77$) (Supplemental Table 4).

The cross-validation results for solidity models did not follow this same pattern (Figure 5). Except for the random forest model in TP_Q1, performance improved across all models as the heritability of the TP was increased. Additionally, a total of 10 different models performed better than the GBLUP using the TP with all environments ($\bar{r}_{ALL} = 0.68$) including GBLUP in TP_Q2 and TP_Q3 ($\bar{r}_{TP_Q2} = 0.77$, $\bar{r}_{TP_Q3} = 0.85$), GBLUP+SS*tl* in TP_Q3 ($\bar{r}_{TP_Q3} = 0.77$), BayesA in

477 TP_Q1, TP_Q2, and TP_Q3 ($\bar{r}_{TP_Q1} = 0.72$, $\bar{r}_{TP_Q2} = 0.79$, $\bar{r}_{TP_Q3} = 0.86$), BayesLASSO in TP_Q2
478 and TP_Q3 ($\bar{r}_{TP_Q2} = 0.77$, $\bar{r}_{TP_Q3} = 0.84$), and random forest in TP_Q2 and TP_Q3 ($\bar{r}_{TP_Q2} = 0.72$,
479 $\bar{r}_{TP_Q3} = 0.85$).

480 For cutting and solidity, forward-validation predictive ability decreased as the heritability
481 of the TP increased except for random forest (Figure 3.4; Figure 3.5). The random forest model
482 saw a small decrease in performance for the models in TP_Q1 but increased back to a similar level
483 as the model using all environments ($\bar{r}_{ALL} = 0.45$, $\bar{r}_{TP_Q1} = 0.43$, $\bar{r}_{TP_Q2} = 0.39$, $\bar{r}_{TP_Q3} = 0.43$) and
484 solidity ($\bar{r}_{ALL} = 0.58$, $\bar{r}_{TP_Q1} = 0.55$, $\bar{r}_{TP_Q2} = 0.56$, $\bar{r}_{TP_Q3} = 0.59$). The random forest models using
485 all environments in the TP, TP_Q2, and TP_Q3 all performed better than GBLUP for forward-
486 prediction of cutting and TP_Q2 and TP_Q3 performed better in forward-prediction for solidity.

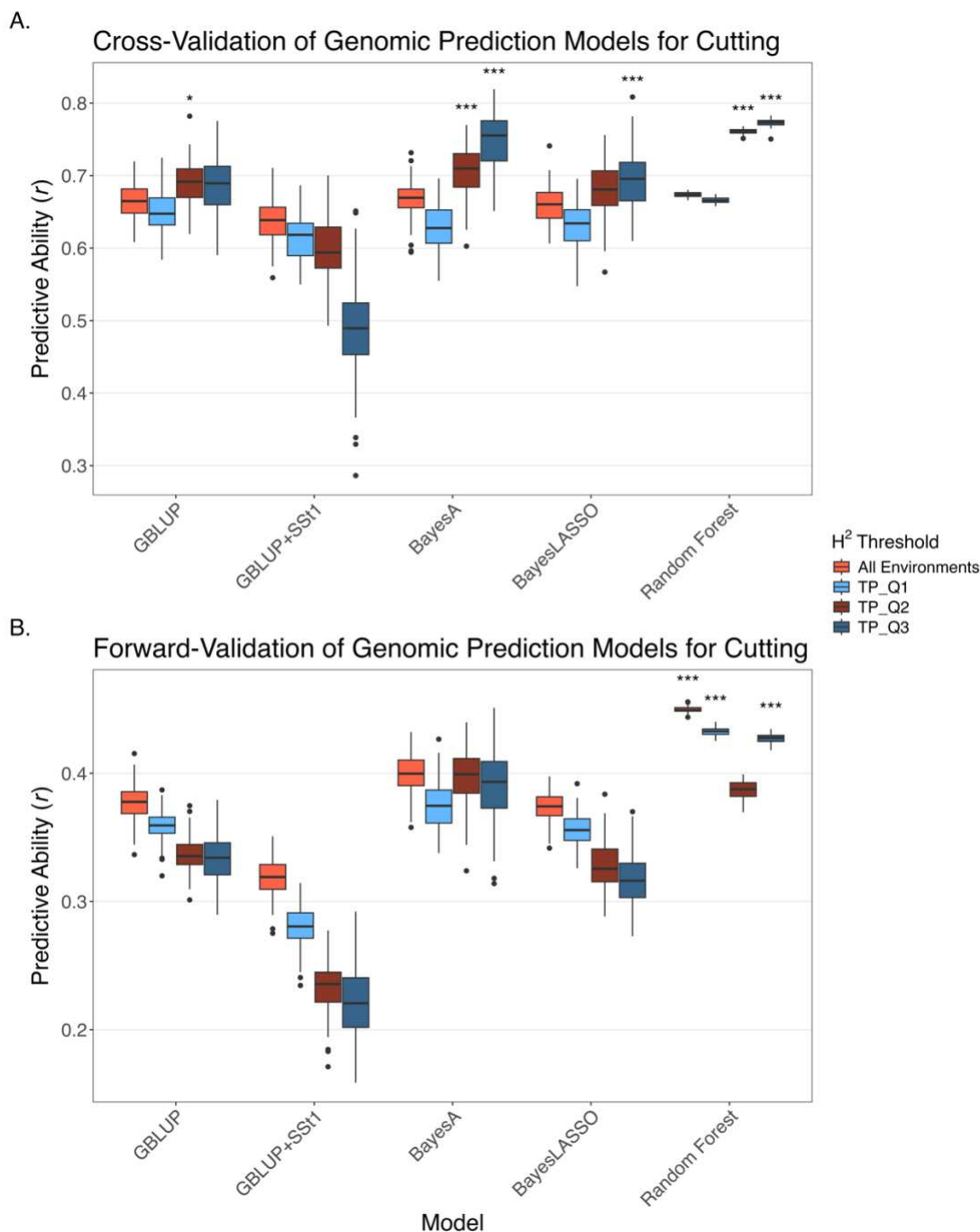


Figure 3.4. Predictive ability for (A) cross-validation and (B) forward-validation of all five models for cutting using environments reduced by heritability thresholds: TP_Q1, TP_Q2, TP_Q3. Predictive ability represents Pearson correlation coefficient (r) between the predicted value of a genotype and the BLUE calculated for that genotype. Cross-validation predictive ability is represented by the blue boxplots and forward-validation predictive ability is represented by the red boxplots. Asterisks (*) denote models that perform statistically better than the respective cross-validation or forward-validation GBLUP model using all environments and the number of asterisks denote the significance level (* = p -value < 0.05; ** = p -value < 0.01; *** = p -value < 0.001).

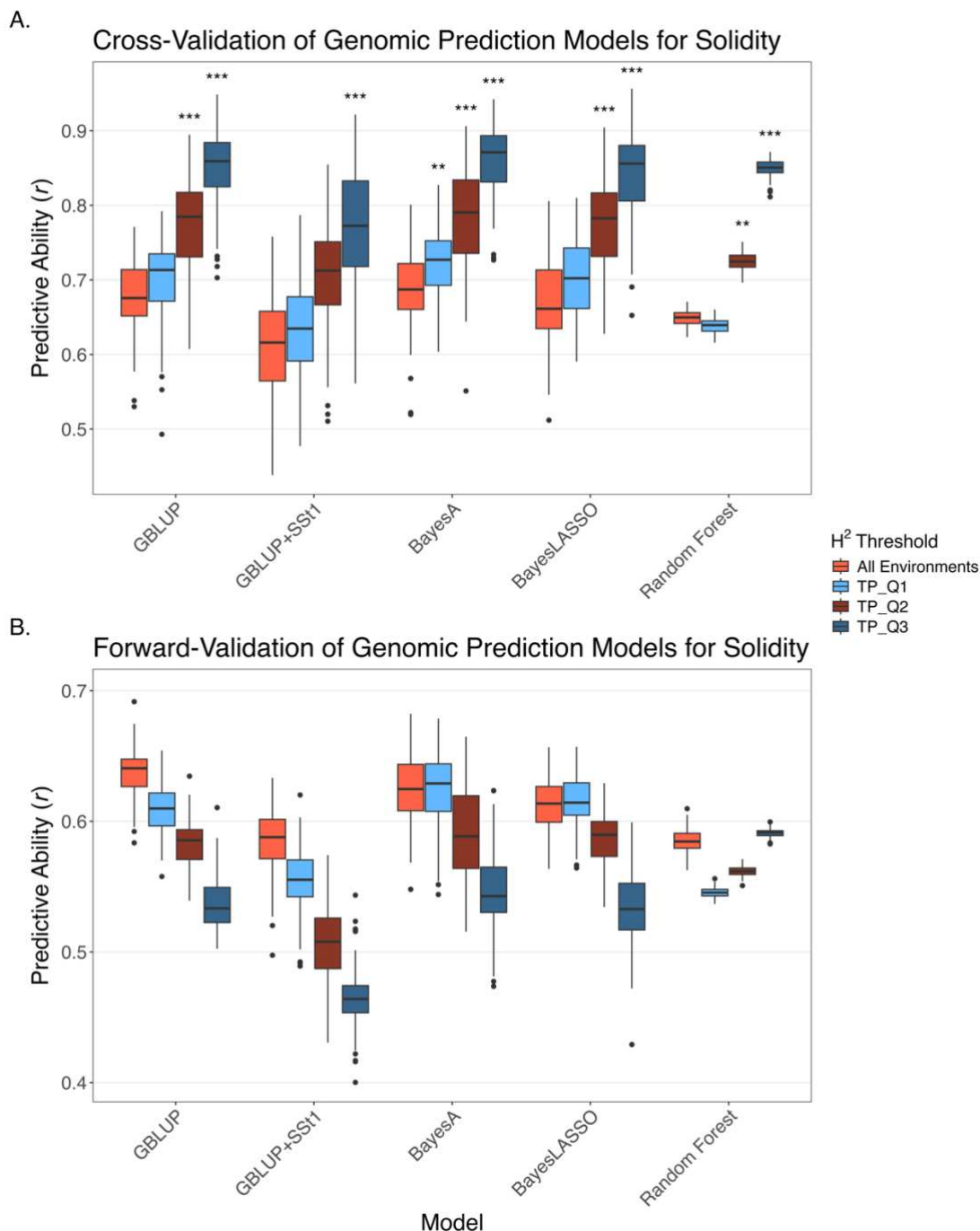


Figure 3.5. Predictive ability for (A) cross-validation and (B) forward-validation of all five models for solidity using environments reduced by heritability thresholds: TP_Q1, TP_Q2, TP_Q3. Predictive ability represents the Pearson correlation coefficient (r) between the predicted value of a genotype and the BLUE calculated for that genotype. Cross-validation predictive ability is represented by the blue boxplots and forward-validation predictive ability is represented by the red boxplots. Asterisks (*) denote models that perform statistically better than the respective cross-validation or forward-validation GBLUP model using all environments and the number of asterisks denote the significance level (* = p -value < 0.05; ** = p -value < 0.01; *** = p -value < 0.001).

DISCUSSION

In this study, five genomic prediction models were used to predict wheat stem sawfly resistance. For cutting, all five models performed well in cross-validation ($\bar{r} > 0.55$). In the forward-validation stage, random forest outperformed all models ($\bar{r} = 0.62$) including GBLUP (p-value < 0.001), while GBLUP+*SStI* performed poorly compared to all the other models ($\bar{r} = 0.32$). This is potentially due to over-fitting as has been reported when using large effect loci with the genomic relationship matrix (Yang et al., 2014). Overfitting could limit the model's ability to account for the more polygenic architecture of the trait.

For solidity, all five models performed well during the cross-validation stage ($\bar{r} > 0.60$) with the Bayesian and GBLUP models the best overall ($\bar{r} > 0.70$) though none significantly outperformed GBLUP in either the cross- or forward-validation groups. Overall, predictability for solidity was higher than cutting in forward validation. The difference in mean predictive ability is likely higher due to the genetic architecture of the solid stem trait. While solidity is largely controlled by the large effect *SStI* locus on chromosome 3B (Cook et al., 2004a; Nilsen et al., 2020), multiple other loci for solidity have been identified (Lanning et al., 2006; Niu et al., 2022; Varella, Zhang, et al., 2019), indicating an oligogenic genetic architecture. However, cutting is affected by loci related to solidity and additional small effect loci related to other factors such as antixenosis (Sherman et al., 2010; Varella et al., 2015). These factors create a higher potential for cutting to have a more complex, polygenic genetic architecture or to be more susceptible to environmental effects.

Increasing the heritability of the training populations was expected to increase the predictive ability of the models (Jannink et al., 2010; Winn et al., 2023; J. Zhang et al., 2022). For

cross-validation for cutting, model performance of BayesA, BayesLASSO, and random forest was statistically better than the performance of the GBLUP (p -value < 0.001) using the full training population. However, in forward-validation, the BayesA and BayesLASSO models did not outperform GBLUP. The Bayesian models were likely able to perform well in cross-validation due to their assumptions and designs. BayesA models assume that all genetic markers have an equal effect on the trait of interest, but that the marker variance is unequal between markers (Meuwissen et al., 2001). BayesLASSO model assumes unequal variances between markers, but that the markers also have unequal effects on the trait of interest, which causes the model to shrink some marker effects towards zero (Kärkkäinen & Sillanpää, 2012; Sun et al., 2010). Since Bayesian models allow for more distinct connections between marker effects and trait values, BayesA and BayesLASSO still likely performed well due to the training data being better suited for the prior assumptions. These models have been shown to increase cross-validation accuracy in other traits such as when combining related populations that have high heritabilities (C. Yin et al., 2024), water stress (Yannam et al., 2025), and micronutrient content in wheat (Meher et al., 2023). While these models performed well in this study in the cross-validation stage, the observed decrease in model efficacy in forward-validation was likely due to the decreased training population size (Fryett et al., 2020).

In forward-validation for cutting, the random forest model still outperformed GBLUP (p -value < 0.001). While the random forest models in training populations with higher heritabilities outperformed GBLUP, they still performed similarly to the random forest model using the full training population. This indicates that culling environments to remove low heritability site years is not advantageous for improving the predictive ability of cutting in this population. However, it demonstrated that random forest is a more useful model for making predictions for the genetic

architecture of cutting. While the random forest model was still likely able to perform well with higher heritability training populations due to the reduced standard error of the training population, the performance of the random forest model using the data from all environments supports the idea that larger datasets are important for machine learning models (Phinzi et al., 2023). Random forest models can be advantageous for highly-dimensional datasets and genomic prediction of complex traits due to its nonparametric tree-based approach (X. Chen & Ishwaran, 2012). This tree-growing nature and bootstrap sampling of variables in a training set allow for random forest models to account for associations and relationships in the dataset (Montesinos López et al., 2022), making it well suited for handling complex genomic interactions such as highly polygenic effects or epistatic interactions.

These results for random forest models have also been observed in simulated and unsimulated data (Montesinos López et al., 2022; Naderi et al., 2016). For non-simulated data, random forest models were able to increase predictive ability for water stress in wheat (Yannam et al., 2025), maize kernel length (Sarkar et al., 2015), and body weight in cattle (B. Li et al., 2018). While all other models likely showed a decrease due to the same training set size limitation, the random forest model was likely able to perform well for cutting in forward prediction, even with a limited training set (X. Chen & Ishwaran, 2012; Ramezan et al., 2021), due to the high control of the *SS11* locus in environments with higher heritability and the ability of random forest models to identify complex relationships between genetic markers and traits of interest (X. Chen & Ishwaran, 2012). This could indicate that the random forest model structure is well suited for the complex genetic architecture (Meher et al., 2022) of these traits while also being less computationally intensive than the other methods (Zheng et al., 2025), which is important for large scale applications in a breeding program.

Cross-validation predictive ability for solidity using the reduced training populations increased for all five models. However, for forward-validation, the predictive ability of all five models decreased and none of the models outperformed GBLUP. Similar results of GBLUP outperforming or being comparable to Bayesian and machine learning methods have been replicated in a variety of studies including mice and cattle studies (Shi et al., 2021), when estimating within independent pig populations (C. Yin et al., 2024), and sugarcane (C. Chen et al., 2023). The model assumptions of the GBLUP are what allows the model to perform well in diverse circumstances for polygenic traits (VanRaden, 2008). A GBLUP model also assumes equal variance of all genetic markers and performs well for traits controlled by additive effects. However, linear models start to lose efficacy with more complex traits that exhibit high genotype by environment interactions or non-linear relationships between genomic loci (Crossa et al., 2017; Montesinos-López et al., 2025), which is likely the reason why GBLUP was not the best model for the genetic architecture of cutting, but was able to perform better for solidity.

Overall, for forward-validation, GBLUP was the best model for solidity and random forest was the best model for cutting in this population. The GBLUP and random forest model predictions are based on the distributions of the training data which can better generalize and extrapolate to new datasets. Additionally, since the data in the study comes from a breeding program, the genotypes in the training and forward-validation datasets are closely related, which would improve predictive ability (Edwards et al., 2019; Goddard, 2009; Van Den Berg et al., 2019). While only using training data that had higher heritabilities was expected to increase predictive ability, it did not have a large impact for cutting and solidity, indicating that using all data available for training GS models could be more advantageous for breeding programs for specific traits.

CONCLUSIONS

Genomic prediction is now commonplace amongst breeding programs for advancing germplasm through their breeding pipelines. While an abundance of work has been done to understand optimization of prediction models on a variety of traits, the use of different genomic prediction models has not yet been explored for WSS resistance. Utilizing models that perform well based on the genetic architecture of WSS resistance could improve the accuracy of selections, even in years where consistent resistance ratings cannot be obtained. Overall, random forest models outperformed all other methods for cutting and GBLUP outperformed all other methods for solidity. Additionally, increasing the heritability of the training population did not improve prediction accuracy for either trait, indicating that it is not an advantageous strategy for these traits. Predictive ability relies on multiple different factors such as population size, marker coverage, relatedness, and heritability, further investigation must be conducted by breeding programs to understand how different genomic prediction methods may be deployed for their traits of interest. Even with cutting and solidity both being largely controlled by *SStI*, these results highlight that effective models vary by trait and population and should be implemented accordingly.

CHAPTER IV – UTILIZING MULTI-OMICS TO UNDERSTAND THE LINK BETWEEN CLEARFIELD AND WHEAT STEM SAWFLY RESISTANCE

INTRODUCTION

Wheat stem sawfly (WSS; *Cephus cinctus* Norton) is a native grass-feeding insect that causes an estimated \$350 million loss in the United States and Canada annually (Beres et al., 2011). The WSS emerges in late spring and oviposits its eggs inside the stem of the plant (Ainslie, 1920), which can result in multiple eggs being present in the same stem. When the larvae hatch, they traverse and feed on the stem and cannibalize each other until one larva remains. This larva will create a hibernaculum called the ‘stub’ at the base of the stem before entering diapause for the winter. During the creation of the hibernaculum, the larvae create a break in the stem that detaches the base from the upper portion of the stem. This predisposes the host plant to lodging due to the damage from the WSS, which is commonly termed ‘cutting’.

Currently, the most well-characterized form of host plant resistance to WSS is the solid stem trait (Cook et al., 2004b). Variation for stem solidness at the *SStI* locus on the long arm of chromosome 3B is controlled by copy number variation of a *TdDof* transcription factor (Nilsen et al., 2020). The solid stem allele is conferred by three copies of the transcription factor while the hollow stem allele is conferred by a single copy. A homeologue of the *SStI* locus is located on the long arm of chromosome 3D (Lanning et al., 2006). Expression of the solid stem trait is variable and is less expressed in cool and wet years (Beres et al., 2012; Holmes, 1984; Platt & Farstad, 1946). Given this, there is a need to better understand non-solid stem resistance to WSS.

Non-preference, also termed antixenosis, refers to any host plant resistance related to deterring insects from colonizing the host plant (Peirce et al., 2022). Previous Y-tube studies (Piesik et al., 2008) and cage choice tests (Buteler et al., 2009; Weaver et al., 2009) have demonstrated that WSS females are attracted to and repulsed by volatile organic compounds (VOCs).

Lavergne et al. (2020) found significant treatment differences in the metabolome and proteome of a diverse set of cultivars infested with WSS. Multiple metabolic pathways that may contribute to WSS resistance were identified, including detoxification, proteinase inhibition, systemic signaling activities, and Ca^{2+} flux. A similar omics study by Hager et al. (2024) utilized two sets of near isogenic lines (NILs) for their study including one developed from the hollow stem variety ‘Reeder’ (PI 613586) and one set developed from the semisolid variety ‘Conan’ (PI 607549) to identify differences between susceptible wheat cultivars and a cultivar associated with solid stem and antixenosis resistance. In the hollow stem NILs, a caffeic acid-oxy-methyltransferase (COMT) involved in lignin biosynthesis (*TraesCS3B02G612000*) and a pectin methylesterase (PME) gene (*TraesCS3B02G610100*) were upregulated in early growth stages of lines with the *3BLa* allele. There was also evidence supporting the connection of these genes to VOCs that impact non-preference previously described by Weaver et al (2009). QTL analyses have identified antixenosis loci such as *Qwss.msub-2D* on chromosome 2D and *Qwss.msub-4A.1* on chromosome 4A which reduced oviposition and attractiveness of wheat plants, respectively (Varella et al., 2017). Other studies have also identified resistance loci not associated with the solid stem trait or *SSt1* locus in a variety of germplasm including hollow stem lines, synthetic wheat, and landraces (Joukhadar et al., 2013; Sherman et al., 2010; Varella et al., 2015; Varella, Weaver, et al., 2019a).

Imidazolinone (IMI) herbicides, such as Beyond® (BASF, Ludwigshafen, Germany), control grass and broadleaf weeds by inhibiting the enzyme acetohydroxyacid synthase (AHAS), also known as acetolactate synthase (ALS). Clearfield® (CL) crops are specifically developed to tolerate these herbicide applications. The ALS enzyme is critical for the synthesis of branched chain amino acids in plants such as valine, leucine, and isoleucine (Mazur et al., 1987). In crops resistant to IMI herbicides, missense mutations within the *ALS* gene induce conformational changes in the resulting protein, preventing IMI herbicides from properly binding to the enzyme (Rojano-Delgado et al., 2015). Three different tolerance genes have been identified in wheat including *imi1*(*AhasL-D1*), *imi2*(*AhasL-B1*), and *imi3*(*AhasL-A1*) on chromosomes 6D, 6B, and 6A, respectively (Pozniak et al., 2004). Multiple studies report that a double mutation in the *imi1* and *imi2* genes confer resistance to IMI herbicides in wheat (Domínguez-Mendez et al., 2017; Jimenez et al., 2016).

Anecdotally, farmers have reported that CL cultivars show less WSS damage compared to non-CL cultivars planted in the same vicinity. This anecdotal finding was confirmed in empirical studies by Vieira et al. (2025) which showed that the CL lines ‘Byrd CL Plus’ (PI 691604), and ‘Brawl CL Plus’ (PI 664255) (Haley, Johnson, Westra, et al., 2012) had lower egg oviposition compared to the susceptible cultivar ‘Byrd’ (PI 664257)(Haley, Johnson, Peairs, Stromberger, Hudson, Seifert, Kottke, Valdez, Rudolph, Bai, et al., 2012) over two years with natural field infestation. Additionally, Byrd CL Plus showed significantly lower infestation rates than Byrd across both years. Based on these findings, we hypothesize that allelic variation in the CL resistance genes, *imi1* and *imi2*, confer variation in WSS resistance observed between Byrd CL Plus and its recurrent parent Byrd. To further investigate this hypothesis, a two-year study of these near-isogenic lines was conducted in the field under naturally high WSS infestation. Proteomic,

metabolomic, and phytohormone analysis was done to identify potential biological pathways leading to variation in WSS resistance.

MATERIALS AND METHODS

Germplasm

Two near-isogenic lines (NILs) from the Colorado State University (CSU) Wheat Breeding and Genetics program in Fort Collins, Colorado were used for this study. Byrd is a hard red winter wheat cultivar released from CSU in 2011. Byrd was derived from a cross between the CSU breeding line ‘CO970547-7’ and the cultivar ‘TAM 112’ (PI 643143)(Rudd et al., 2014) released by Texas A&M University from the Vernon, Texas breeding program in 2005. Byrd CL Plus is a hard red winter wheat cultivar released from CSU in 2018. Byrd CL Plus was developed using a backcross breeding approach to introgress the *imi1* and *imi2* Clearfield® (BASF, Ludwigshafen, Germany) mutations conferring the herbicide tolerant trait into the cultivar Byrd. To develop Byrd CL Plus, the recurrent parent Byrd crossed to the CL donor line CO06072. The resulting Byrd/CO06072 F₁ was then crossed back to the recurrent parent, Byrd, and this was repeated for a total of four generations. Byrd CL Plus is 96.875% Byrd parentage by pedigree. Proprietary molecular markers (BASF, Ludwigshafen, Germany) were used to confirm the presence of the *imi1* and *imi2* herbicide tolerant alleles at each generation and in the final derived line. Tolerance to Beyond Herbicide (BASF, Ludwigshafen, Germany) was also confirmed at multiple points during the breeding process.

Sample Collection

Samples were collected at the United States Division of Agriculture – Agricultural Research Service Central (USDA-ARS) Great Plains Research Station in Akron, Colorado during the 2024 and 2025 field seasons. Three biological replicates of both genotypes were grown in 1.52 m x 3.05 m yield plots in a completely random design that included three replicates of each

genotype. Trials were planted into soils with a N:P ratio of 82.6:7.7 kg and 59.9:21.8 kg for the 2024 and 2025 field seasons, respectively. For the 2025 season, 2.0 kg N and 7.2 kg P were applied for fertilizer at planting. No fertilizer was applied mid-season for either year and no additional management was performed before harvest. Sawfly flight was determined by performing sweeps on a 30 m field section planted with one cultivar. Sawfly caught in the sweep were counted to determine presence of WSS and peak flight. Samples were collected at three time points: before WSS flight, just past peak WSS flight, and two weeks after peak WSS flight. In 2024, samples were collected on May 9th, May 29th, and June 13th, when the wheat was at approximately Zadok's scale 32, 59, and 69, respectively. Due to a sampling error, there were only two replicates of Byrd CL Plus collected at each time point in the 2024 field season. With the two genotypes, three replicates (except for Byrd CL Plus), and three time points, there were a total of 15 samples collected in the 2024 field season. In 2025, samples were collected on May 12th, May 29th, and June 12th, when the wheat was at approximately Zadok's scale 32, 59, and 69, respectively. With the four genotypes, three replicates, and three time points, there were a total of 18 samples collected in the 2025 field season.

In total, across both years, there were 6 replicates of Byrd and 5 replicates of Byrd CL Plus at all three time points. For consistency, all samples were collected between 10 a.m. and 12 p.m. For sample collection, the flag leaf was removed from a single stem before the stem was cut from the plot. For the samples collected before WSS flight, the stems were flash frozen in liquid nitrogen (N₂) immediately after being removed from the plot. For samples collected at peak WSS flight or two weeks post peak WSS flight, the selected stem was then split long-ways from the head down to the first node on the stem. This section was assessed for the presence of WSS eggs, larvae, and/or frass.

Once identified, eggs, larvae, and/or frass were removed from the stem and region between the wheat head and first node of the stem to reduce cross-species contamination prior to flash freezing in liquid nitrogen. Two stems were collected from each plot to provide ample tissue for each genotype, replicate, and timepoint. Tools used in tissue sampling were cleaned with 70% ethanol (C₂H₅OH) between samples to avoid cross-contamination. Samples were then transported from Akron, Colorado to Fort Collins, Colorado on dry ice (CO₂). Samples were transferred from dry-ice to -70°C freezer until sample preparation.

Phytohormone and Metabolomic Sample Preparation

For sample preparation, 30mg +/- 1mg of lyophilized and ground stems were weighed into 2-mL epi-tubes. One mL of 80% MeOH, 10uL of global internal standard (IS), and 10uL of phytohormone IS were added to the sample. Global IS for metabolomics analysis included 600 ug/mL L-alanine-13C3, 60 ug/mL L-phenylalanine-13C6, 30 ug/mL fumaric acid-13C4, 300 ug/mL L-tryptophan-13C11, and 30 ug/mL indole-3-acetic-acid-13C6 in 50% MeOH and phytohormone IS included 500 ng/mL of SA-d4, 250 ng/mL of ABA-d6, 100 ng/mL of JA-d5, 200 ng/mL of IAA-d5, and 5 ng/mL of tZ-15N4 in 50% MeOH.

Samples were briefly vortexed before being added to the VWR VX-2500 MultiTube Vortexer (Marshall Scientific, New Hampshire, USA) in a 6°C cold cabinet for 30 minutes. Samples were then sonicated in a cold-water bath for 15 minutes before moving back to the shaker in the cold cabinet for an additional 30 minutes. Samples were centrifuged at 15,000xg x 4C for 15 minutes before 800uL of supernatant was recovered and transferred to 2-mL glass vials.

Next, 100 uL of supernatant from the glass vials was transferred to new glass vials for untargeted metabolomics analysis. An additional 25 uL was pooled into a new glass vial to create

a quality control (QC). Both the metabolomics samples and QC were stored in a -80°C freezer. To collect more supernatant, 0.5 mL of 80% MeOH was added to the remaining pellet, and the samples were briefly vortexed and centrifuged at 15,000xg x 4C for 15 minutes. An additional 0.45 mL of supernatant was collected from samples and added to glass vials with the previous supernatant collections.

Samples were lyophilized then resuspended in 100 uL of 50% MeOH and tubes were rotated until the entire inner wall of the tube was coated in the 50% MeOH. Samples were briefly centrifuged before being transferred into inserts. Inserts were placed inside the glass vials and then centrifuged at 3,000xg x 4C for five minutes before 75 uL of supernatant was transferred into a new insert, being careful not to extract any of the solids from the bottom of the insert. Remaining samples from the original inserts were pooled into a 0.6 mL epi tube and vortexed to be used as QC for the phytohormone analysis. The QC was centrifuged and 100 uL of QC supernatant was transferred into new inserts until no more supernatant remained.

The authentic standards used in this assay included jasmonic acid (JA), salicylic acid (SA), indole 3-acetic acid (IAA), indole 3-acetonitrile (IAN), indole-3-carboxylic acid (IACA), indole-3-acetamide (IAM), indole-3-propionic acid (IPA), indole-3-acetyl-alanine (IA-Ala), N-(3-indolylacetyl)-L-phenylalanine (IA-Phe), gibberellin A3 (GA3), gibberellin A4 (GA4), and trans-zeatin ribose (tZR) purchased from Sigma-Aldrich (St. Louis, MO, USA), jasmonic acid-d5 (JA-D5), salicylic acid-D4 (SA-D4), and indole 3-acetic acid-D5 (IAA-D5) purchased from CDN Isotopes (Point-Claire, Ontario, Canada), jasmonic acid-isoleucine (JA-Ile) and trans-zeatin (tZ) purchased from Cayman (Ann Arbor, MI, USA), IA-aspartic acid (IA-Asp) purchased from Toronto Research Chemicals (Vaughan, Ontario, Canada), abscisic acid-D6 (ABA-D6), trans-zeatin-15N4 (tZ-15N4), and gibberellin A4-D2 (GA4-D2) purchased from Olchemim (Holice,

Czech Republic), indole-3-butyric acid (IBA) which was purchased from Santa Cruz Biotechnology (Dallas, TX, USA), IA-Ile purchased from GoldBio (St Louis, MO, USA) and abscisic acid (ABA) purchased from ChromaDex (Los Angeles, CA, USA).

Ultra-Performance Liquid Chromatography-Tandem Mass Spectrometry (UPLC-MS/MS) analysis was performed on a Waters ACQUITY Classic UPLC coupled to a Waters Xevo TQ-S triple quadrupole mass spectrometer (Waters Corporation, Massachusetts, USA). Chromatographic separations were carried out on a Waters ACQUITY Premier HSS T3 VanGuard fully integrated technology (FIT) column (2 x 50 mm, 1.7 μ M). Mobile phases were (A) water with 0.1% formic acid and (B) acetonitrile with 0.1% formic acid. The liquid chromatography LC gradient was as follows: time = 0 min, 1% B; time = 0.65 min, 1% B; time = 2.85 min, 99% B; time = 3.5 min, 99% B; time = 3.55 min, 1% B; time = 5 min, 1% B. Flow rate was 0.5 mL/min and injection volume was 3 μ L. Samples were held at 6°C in the autosampler, and the column was operated at 45°C.

Mass detector was operated in electrospray ionization (ESI)+ and ESI- mode. The capillary voltage set to 0.7 kV and the inter-channel delay was set to 3 msec. Source temperature was 150°C and desolvation gas (nitrogen) temperature 450°C. Desolvation gas flow was 1000 L/h, cone gas flow was 150 L/h, and collision gas (argon) flow was 0.15 mL/min, and nebulizer pressure (nitrogen) was set to 7 Bar. The MS acquisition functions were scheduled by retention times. Autodwell feature was set for each function and dwell time was calculated by Masslynx software (Waters Corporation, Massachusetts, USA) to achieve 12 points-across-peak as the minimum data points per peak.

Proteomics Sample Preparation

For proteomics sample prep, 200 mg of frozen stem samples were weighed to use for protein extraction. Using the Pierce Plant Total Protein Kit (ThermoFisher Scientific, Massachusetts, USA) the proteins were extracted. Briefly, the plant material was loaded into a filter and punched 60 times with a pipette tip. 100uL of Protein Denaturing buffer was added along with 1X HALT, then the sample was ground with twisting force 60 times to ensure full saturation of the filter. The sample was allowed to incubate at room temperature for 3 minutes, then centrifuged for 5 minutes at 16,000xg. This process was repeated once for a total of 200uL of lysis extract. For samples 22-33 from the 2024 set, an additional 200-400uL of water was added to assist with lysis since the plants were from a later timepoint and had dried. Total protein content was measured using the Pierce BCA Protein Assay Kit (ThermoFisher Scientific, Massachusetts, USA) following manufacturer instructions. Absorbance at 550nm was measured on a BioRad 680 microplate reader and total protein concentrations calculated based on a bovine serum albumin standard curve fit to a quadratic curve using Microplate Manager® 5.0 Software (BioRad, California, USA).

Samples were digested using the EasyPep Mini MS Sample Prep Kit (ThermoFisher Scientific, Massachusetts, USA) following the manufacturer's instructions. Briefly, 100ug total protein was aliquoted from each sample and raised to 100ul with lysis buffer. Reduction and alkylation solutions were sequentially added followed by incubation at 95°C for 10 minutes. After cooling to room temperature, 50ug of a 0.2ug/uL Trypsin/LysC mixture was added and samples were digested with shaking at 37°C for 2 hours. The enzymes were then deactivated with the Digestion Stop Solution and contaminants removed using mixed mode peptide clean up columns. Peptide eluate was dried in a vacuum evaporator and resuspended in 3% acetonitrile/0.1% formic

acid. Once resolubilized, absorbance at 205nm was measured on a NanoDrop (ThermoScientific, Massachusetts, USA) and total peptide concentration was subsequently calculated using an extinction coefficient of 31 (Scopes, 1974).

Reverse phase chromatography was performed using water with 0.1% formic acid (A) and 80% acetonitrile with 0.1% formic acid (B). A total of 1 μ g of peptides was purified and concentrated using an on-line enrichment column (Thermo Scientific PepMap Neo C18 5 μ m, 300 μ m ID x 0.5cm). Subsequent chromatographic separation was performed on a Vanquish Neo (Thermo Scientific, Massachusetts, USA) on a reverse phase nanospray column with integrated silica emitter (ionopticks Aurora Ultimate Gen 3 C18 1.7 μ m, 75 μ m ID x 25cm column, 45°C) using a 90 minute method at a flow rate of 300 nanoliters/min: 1-6%B over 3 minutes followed by 6-35%B over 70 minutes, 35-45%B over 5 minutes ending in 12 minutes of washing at 500 nanoliters/minute, 99%B. Peptides were eluted directly into the mass spectrometer (Orbitrap Eclipse, Thermo Scientific, Massachusetts, USA) equipped with a Nanospray Flex ion source (Thermo Scientific, Massachusetts, USA) and spectra were collected over a m/z range of 375–2000 under positive mode ionization. Ions with charge state +2 or higher were accepted for tandem-mass spectrometry (MS/MS) using a dynamic exclusion limit of 1 MS/MS spectra of a given m/z value with an exclusion duration of 60 s. The instrument was operated in Fourier Transformation (FT) mode for MS detection (profile; resolution of 240,000) and ion trap mode for MS/MS detection with a normalized Higher-energy Collisional Dissociation (HCD) collision energy set to 30% and data centroided.

Proteome Discoverer (PD) 3.0 was used for data processing (Thermo Scientific, Massachusetts, USA). A precursor detector node with S/N=1.5 was used to identify additional precursors within the isolation window of the precursor spectrum when chimeric or mixed spectra

is present. Spectra from all samples were searched using the Sequest HT (Eng et al., 1994) node, setting methionine oxidation as dynamic and cysteine Carbamidomethylation as fixed modification. Subsequently, the data was submitted to an intensity-based rescoring process (INFERYS node in PD) using a deep learning algorithm to predict fragment ion intensities (Zolg et al., 2025). Raw data was interrogated against the FASTA file for the reference proteome of wheat (*Triticum aestivum* L.) available at Uniprot (UP000019116, downloaded on 10-16-2024). Additionally, the cRAP proteome was included (The common Repository of Adventitious Proteins-cRAP- that contains commonly found contaminant proteins in proteomics experiments) FASTA file “cRAP_100518”. Sequest HT was searched with a fragment ion mass tolerance of 0.60 Da and a parent ion tolerance of 10 PPM. Peptide spectral matches (PSMs) were validated using the Percolator node (Käll et al., 2007). Thresholds were set such that a false discovery rate (FDR) of $\leq 1\%$ and protein identification was defined with at least one peptide.

Metabolomics Sample Processing

For untargeted analysis, samples were run on Reverse Phase (RP) Chromatography using a Waters ACQUITY I-Class FTN-UPLC coupled to a Waters Xevo G2-XS qToF mass spectrometer (Waters Corporation, Massachusetts, USA). Chromatography conditions for RP were as follows: The Mobile Phases used for the RP separation consisted of Solvent A: Water with 0.1% Formic Acid; Solvent B: Acetonitrile with 0.1% Formic Acid. The column used was a Waters Acquity Premiere HSS-T3 column, length x diameter of 100 x 2.1 mm, with 1.7 micron particle size. The column was maintained at a temperature of 45 °C. The flow rate was 0.5 mL / min and the gradient consisted of: 0–1-minute, 1% B; 1-13 minutes linear gradient, 1 to 98% B; 13-16 minutes, Hold at 98% B; 16 minutes Step Gradient 98%-1% B; 16-20 minutes, 1% B. The injection volume was 2 uL per sample.

The column eluent was infused into a Waters Xevo G2-XS qToF-MS with electrospray source in positive ion, sensitivity mode, with extended dynamic range. The mass range 50-1200 m/z was scanned at 0.1 sec / scan alternating between MS1 (6 V Collision Energy) and MSE (15-30 V Ramp) collected in continuum (Profile) mode. Calibration was performed using sodium formate with 1 part per million (ppm) mass accuracy. The electrospray source parameters were as follows: Inlet Capillary voltage was 1.0 kV, Sampling Cone Voltage 30 V, Source Temperature 120 °C, Source Offset 80 V, Desolvation Temperature 600 °C, Cone Gas Flow 50 L/h and Desolvation Gas Flow 1000 L/h. Lockspray solution containing LeuEnk was infused at 10 uL / min and sampled every 40 seconds with reference mass of 556.2771 m/z. To enable downstream processing, profile mode data was converted to centroided mzML 64-bit precision data using MSConvertGUI 3.0 with the Waters Peak Picking filter, and no other filters were selected.

Phytohormone data was processed in Skyline 25.1.0 (MacLean et al., 2010) to obtain a preliminary assessment of data quality. Data was imported with mass accuracy 20 ppm, and the following masses were used for targeted data processing: L-phenylalanine-13C6 [M+H] (172.1064 m/z at RT = 2.4 min), L-tryptophan-13C11 [M+H] (216.1341 m/z at RT = 3.0 min), and indole-3-acetic-acid-13C6 [M+H] (182.0907 m/z at RT = 5.1 min), all of which were easily observed in every injection. The 13C labeled L-Alanine and fumaric acid ISs were not observed, likely due to a combination of early elution (high polarity in RP column) and positive polarity (fumaric acid exists preferentially as an [M-H] ion).

Peak detection and grouping for metabolomics data was performed using R version 4.5.1 (R Core Team, 2025). The xcms package version 4.6.4 (Smith, 2017) and RAMClustR version 1.3.2 (Broeckling et al., 2014). Peak width detection used a peakwidth set to (1.5, 12) and a ppm error set to 20. For peak grouping, the minfrac was set to 0.4 and the ppm error was set to 10.

Detected peaks were analyzed using SIRIUS 4 (Dührkop et al., 2019) using the MS2 mass accuracy set to 10 ppm and the precursor deviation set to 10 ppm. To match peaks to the appropriate metabolite, the dataset was compared to the ARC-BIO specific wheat database, all Chemical Entities of Biological Interest (CHEBI) wheat accessions, and the PubChem database.

Statistical Analysis

Differential Expression Analysis

Differences in proteomic, metabolomic, and phytohormone abundances were separately performed using the ‘limma: Linear Models for Microarray and RNA-Seq Data’ R package (Ritchie et al., 2015). Differential expression of proteins, metabolites, and phytohormones were individually performed using the following linear model (Phipson et al., 2016; Smyth, 2004) where y_{gi} is the log-expression of compound g in sample i :

$$E(y_g) = X\beta_g$$

Where y_g is a vector of log-expressions for n compounds, X is an $n \times p$ design matrix representing the experimental design, and β_g is an unknown coefficient vector that parameterizes average expression levels in each experimental condition. In this analysis, the experimental design was additively modeled to include the *genotype* $\times$ *time* effect and year effects. For each compound, the y_{gi} are assumed independent with:

$$var(y_{gi}) = \frac{\sigma_g^2}{w_{gi}}$$

Where σ_g^2 is an unknown variance and w_{gi} are known weights. Finally, the least squares coefficient estimator is:

$$\hat{\beta}_g = (X^T W_g X)^{-1} X^T W_g y_g$$

Where W_g is the diagonal matrix of weights (Smyth, 2004). Differential expression was assessed at time points one, two, and three for all biological compounds. To correct for multiple testing, all p-values were FDR adjusted.

Statistical Analysis - Proteomics

For proteomics, presence/absence analysis was performed to assess differences in protein expression that cannot be analyzed using the previous linear model. Due to the nature of detection, proteins which were below the detection threshold in a sample were considered “absent” and proteins that had detectable expression levels in at least 60% of replicates for each genotype by time point were considered present. Proteins that were present across multiple time points in one genotype but absent in the other were assessed for significance. A Fisher’s Exact Test (Fisher, 1936) was conducted to assess likelihood that the presence or absence of the protein was meaningful. To further understand metabolomic and phytohormone relationships, a Spearman’s rank correlation coefficient (ρ) was calculated between annotated and unannotated metabolite abundances and annotated metabolite, unannotated metabolite, and phytohormone abundances.

Phenotypic Analysis

Based on proteomic results, differences in heading date between Byrd and Byrd CL Plus, ASReml-R version 4.1.0.160 (Butler et al., 2023) was used to perform the following univariate model:

$$y_{ijk} = \mu + G_i + Y_j + (GY)_{ij} + s_{k(j)} + \varepsilon_{ijk}$$

Where y_{ijk} is the observed heading date; μ is the overall mean; G_i is the fixed effect of the i^{th} genotype; Y_j is the fixed year effect for the j^{th} year; $(GY)_{ij}$ is the genotype by year effect for the i^{th} genotype and j^{th} year; $s_{k(j)}$ is the random row by column interaction with autocorrelated structures for both dimensions of the trial design; and ε_{ijk} is the residual error term, where $\varepsilon_{ijk} \sim N(0, I\sigma^2_\varepsilon)$, where I is an identity matrix and σ^2_ε is the residual error variance. The model was used to determine differences in heading date between Byrd and Byrd CL Plus and between field seasons.

RESULTS

Variation in the Proteome

Liquid-chromatography mass spectrometry analysis of the NILs resulted in the identification of 152,660 peptides annotated as 14,876 protein accessions, including 9,190 accessions identified using at least two unique peptides. These filtered accessions were linked to 12 GO biological processes. Some accessions overlapped between processes, there were a total of 893 accessions associated with transport, 950 with cell organization and biosynthesis, 1,154 with protein metabolism, 571 with RNA metabolism or transcription, 251 with signal transduction, 181 with DNA metabolism, 145 with cell cycle or cell proliferation, 768 with stress response, 156 with developmental processes, 3,805 with other metabolic process, 1,313 with other biological processes, and 3,068 with no identified process.

For multivariate analysis, proteins were filtered to only include accessions that were detected in at least 60% of all replicates for a genotype at each time point, which resulted in 5,909 unique accessions. There was variation in the proteome between Byrd and Byrd CL Plus including

one significantly ($\text{Log}_2\text{FC} \geq |2|$, $\text{p-value}_{\text{FDR}} \leq 0.05$) differentially expressed protein pre-infestation, two during peak infestation, and two post-infestation (Figure 1; Table 1). None of the significant accessions were significant in multiple timepoints. Three of the five accessions had assigned GO biological terms including accession. At peak infestation, A0A3B6JCK1 ($\text{Log}_2\text{FC} = -3.48$, $\text{p-value}_{\text{FDR}} = 0.002$) and A0A3B6NUD3 ($\text{Log}_2\text{FC} = -2.30$, $\text{p-value}_{\text{FDR}} = 0.004$) were associated with RNA metabolism or transcription, unspecified biological processes, and unspecified metabolic processes, respectively. Post-infestation, A0A3B6I0X1 ($\text{Log}_2\text{FC} = -2.74$, $\text{p-value}_{\text{FDR}} = 0.04$) was associated with cell organization and biogenesis.

Presence-absence analysis of protein accessions was performed by identifying accessions which had protein abundances in one genotype and unmeasurable protein abundances in another genotype using a Fisher's test. No accessions were significant in this analysis (Supplemental Table 4.1).

Clearfield NILs Protein Expression

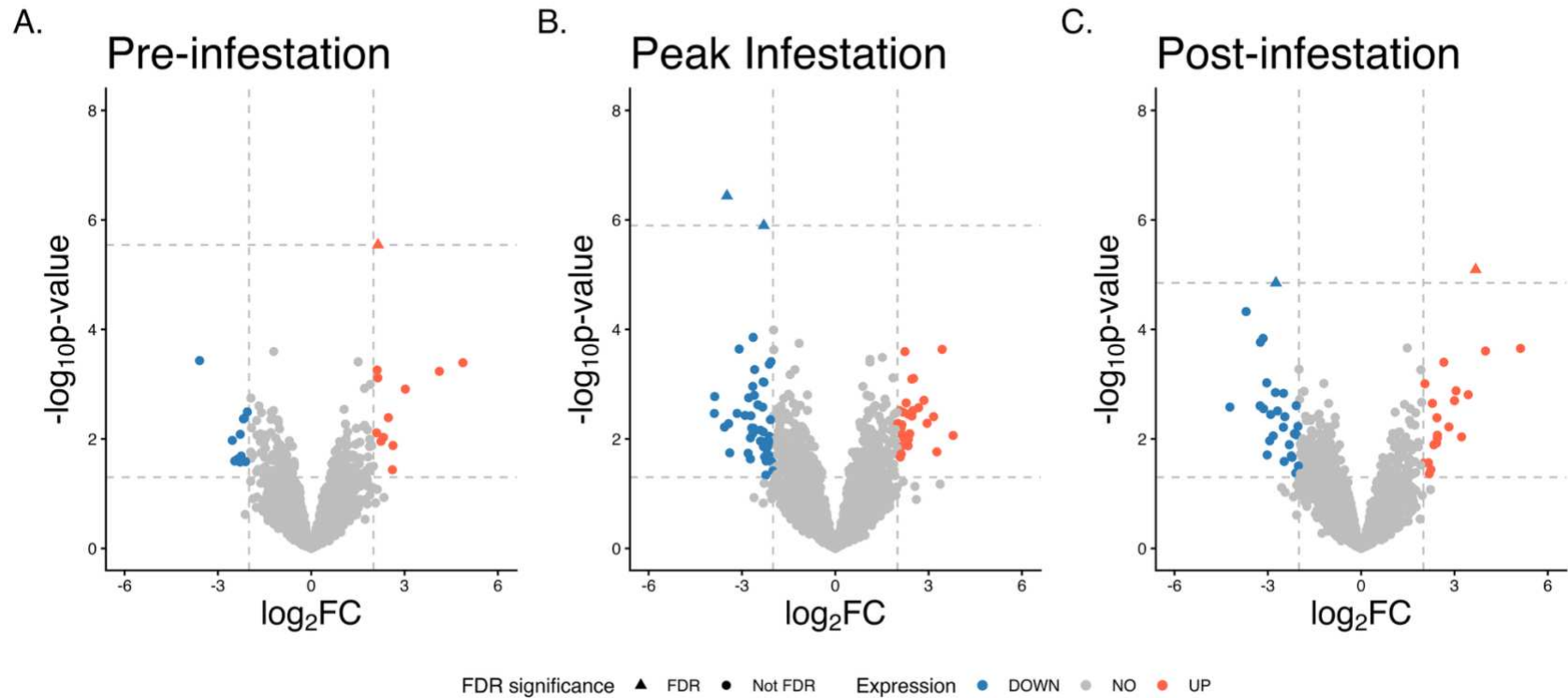


Figure 4.1. Volcano plots of differential protein expression between Byrd and Byrd CL Plus at A) pre-infestation, B) peak infestation, and C) post-infestation. The vertical dashed lines indicate a $\text{Log}_2FC > |2|$ and the lower and upper horizontal dashed lines indicate the $p\text{-value}_{\text{unadjusted}}$ threshold of 0.05 and the $p\text{-value}_{\text{FDR}}$ threshold of 0.05, respectively. Blue points indicate accessions that were downregulated, red indicates accessions that were upregulated, and triangular points indicate accessions that surpass the FDR adjusted threshold.

Table 4.1. Significant differentially expressed protein accessions between Byrd and Byrd CL Plus for pre, peak, and post-infestation where time indicates pre-infestation (1), peak infestation (2), or post-infestation (3), and Log2FC and p-valueFDR indicate the significance of the protein accession in terms of fold change and adjusted p-value, GO biological process indicates identified functions of the protein accession, the Ensembl gene ID for each of the protein accessions, and the chromosome (Chr) where the gene is located.

Time	Accession	Log2FC	p-valueFDR	GO Biological Process	Ensembl Gene ID	Chr
1	A0A3B6SJ32	2.15	0.02	-	TraesCS7B02G255000	7B
2	A0A3B6JCK1	-3.48	0.00	RNA metabolism OR transcription; other metabolic processes; other biological processes	TraesCS4D02G016400	4D
	A0A3B6NUD3	-2.30	0.00	other metabolic processes	TraesCS6A02G302400	6A
3	A0A3B6RQL6	3.68	0.04	-	TraesCS7A02G469800	7A
	A0A3B6I0X1	-2.74	0.04	cell organization and biogenesis	TraesCS4A02G396600	4A

Variation in the Metabolome

A total of 7,590 putative compounds were detected. Of these compounds, 104 were annotated as distinct metabolites with a SIRIUS score greater than 80%. Annotated metabolites were classified into 110 unique classes, 45 unique superclasses, and seven different pathways including alkaloids, amino acids and peptide, carbohydrates, fatty acids, polyketides, shikimates and phenylpropanoids, and terpenoids. Unannotated metabolites were not able to be identified due to only making it through the MS1 phase and not MS2 or due to low confidence matching for the chemical structure at that peak. Differences in metabolomes between genotypes and timepoints were inferred by principal component analysis (PCA) (Figure 4.2; Supplemental Figure 4.1). The first, second, and third PCs accounted 41.9%, 9.5%, and 8.1% of total variation, respectively.

After FDR correction of p-values, one metabolite was significant (ID = C09711, $\text{Log}_2\text{FC} = 3.20$, $\text{p-value}_{\text{FDR}} = 0.02$) at the peak infestation timepoint. No metabolites were significant between Byrd and Byrd CL Plus at the pre- or post-infestation time points. However, notable differentially expressed metabolites ($\text{Log}_2\text{FC} > |2|$, $\text{p-value}_{\text{unadjusted}} < 0.05$) were identified at all three time points (Figure 3). At pre-infestation, there were 10 metabolites that had a $\text{Log}_2\text{FC} < -5$ ($\text{p-value}_{\text{unadjusted}} < 0.05$). While nine of the metabolites were unidentified, one of the metabolites was annotated as 3,4-Bis(sulfanylidene)naphthalene-2-sulfonic acid (SIRIUS score = 0.93) which belonged the tropolones superclass and is associated with shikimates and phenylpropanoids pathways. There were 10 metabolites that had a $\text{Log}_2\text{FC} > 5$ ($\text{p-value}_{\text{unadjusted}} < 0.05$), all of which were unannotated. At peak infestation, there were 28 metabolites that had a $\text{Log}_2\text{FC} < -5$ ($\text{p-value}_{\text{unadjusted}} < 0.05$) and 13 metabolites that had a $\text{Log}_2\text{FC} > 5$ ($\text{p-value}_{\text{unadjusted}} < 0.05$), none of which were annotated. Post-infestation, there were 9 metabolites that had a $\text{Log}_2\text{FC} < -5$ (p-

value_{unadjusted} < 0.05) and 12 metabolites that had a Log₂FC > 5 (p-value_{unadjusted} < 0.05), none of which were annotated.

All Spearman metabolite correlations were significant after FDR correction (p-value_{FDR} < 0.05) (Supplemental Table 4.3). Pre-infestation, there were three unannotated metabolites significantly correlated ($r \geq 0.70$, p-value_{FDR} < 0.05) with one or more annotated metabolites (Supplemental Table 4). The pathways associated with the annotated metabolites included shikimates and phenylpropanoids, amino acids and peptides, alkaloids, terpenoids, polyketides, and fatty acids. Associated classes included triacylglycerols, flavones, flavonols, microlins and mirabimids, phenylalanine-derived alkaloids, carotenoids, capsaicins and capsaicinoids, linear peptides, pyridine alkaloids, simple phenolic acids, macrolide lactones.

At peak infestation, there were twelve unannotated metabolites significantly correlated ($r \geq 0.70$, p-value_{FDR} < 0.05) with one or more annotated metabolites. The pathways associated with the annotated metabolites included shikimates and phenylpropanoids, polyketides, amino acids and peptides, terpenoids, fatty acids, and alkaloids. Associated classes included macrolide lactones, triacylglycerols, linear peptides, diacylglycerols, flavones, flavonols, simple phenolic acids, bisnaphthalenes, linear peptides, amino acids, pyridine alkaloids, carotenoids, and phenylalanine-derived alkaloids. Post-infestation, there were three unannotated metabolites significantly correlated ($r \geq 0.70$, p-value_{FDR} < 0.05) with one or more annotated metabolites. The pathways associated with the annotated metabolites included fatty alcohols, aminosugars, amino acids, purine nucleosides, triacylglycerols, macrolide lactones, pyridine alkaloids, simple phenolic acids, flavones, carotenoids, and linear peptides. For all three time points there were multiple negatively correlated annotated metabolites from all the same pathways.

PCA of Metabolite Expression

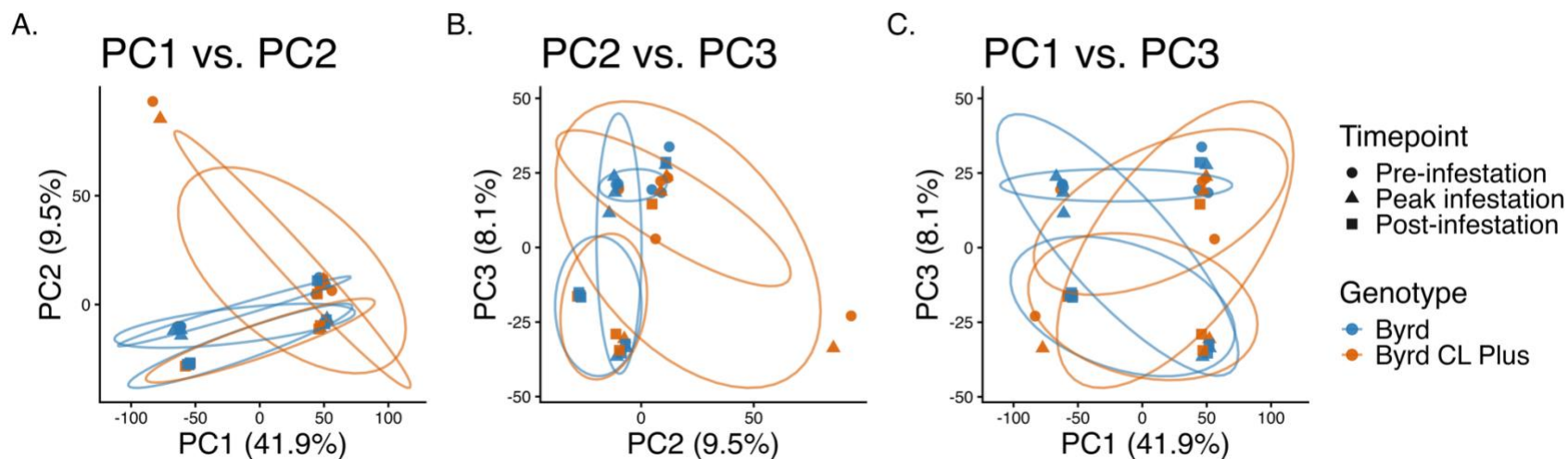


Figure 4.2. Principal component (PC) plots for metabolite expression where (A) shows variance explained by PC1 and PC2, (B) shows PC2 and PC3, and (C) shows PC1 and PC3. Blue ellipses and datapoints indicate Byrd and orange ellipses and datapoints indicate Byrd CL Plus. Timepoint is differentiated by shape where circles indicate pre-infestation samples, triangles indicate peak infestation samples, and squares indicate post-infestation samples. Ellipses indicate the spread of the genotypes at different timepoints.

Clearfield NILs Metabolite Expression

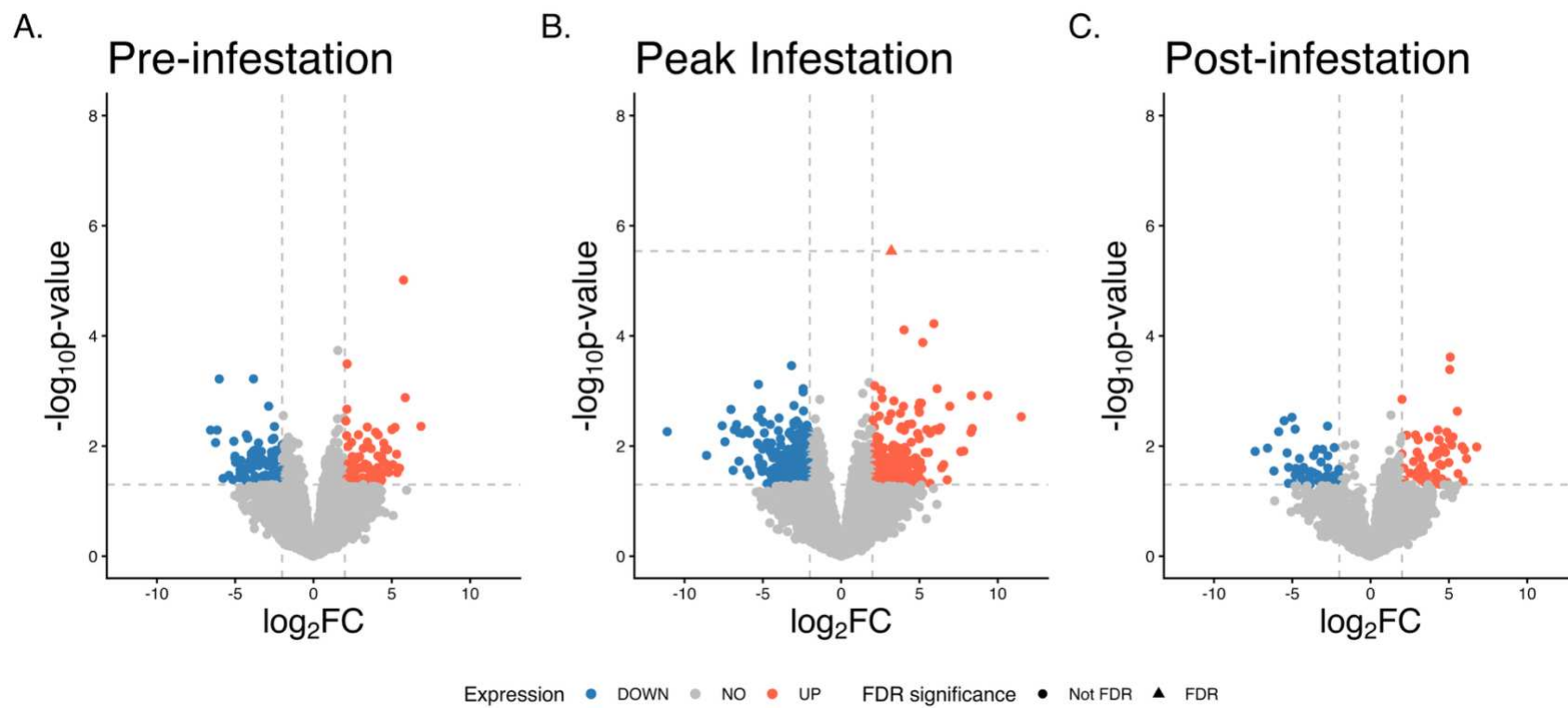


Figure 4.3. Volcano plots indicating differentially expressed metabolites in Byrd CL Plus in reference to Byrd where A) represents samples collected pre-infestation, B) represents samples collected during peak infestation, and C) represents samples collected two weeks after peak infestation. The x-axis indicates the log₂ fold change and the y-axis indicates the -log₁₀(p-value). Each dot represents a metabolite where grey indicates a metabolite with an unadjusted p-value < 0.05 and a Log₂ fold change between -2 and 2, red indicates a metabolite with an unadjusted p-value > 0.05 and a Log₂ fold change greater than 2, and blue indicates a metabolite with an unadjusted p-value > 0.05 and a Log₂ fold change less than -2.

Variation in the Phytohormone Profile

Sixteen different phytohormone abundances were measured for both years of data (Figure 4.4). Eleven of the phytohormones were identified in both years and were considered for multivariate analysis including: salicylic acid (SA), acrylic acid (AA), jasmonic acid (JA), jasmonyl-isoleucine (JA-Ile), indole-3-butyric acid (IBA), abscisic acid (ABA), indole-3-acetic acid (IAA), indole-3-acetic acid-aspartic acid (IAA-Asp), indole-3-acetate monooxygenase (IACA), trans-zeatin (tZ), and trans-zeatin riboside (tZR). After FDR correction of p-values, no statistical difference was identified in the abundances of the phytohormones between Byrd and Byrd CL Plus.

Because only complete cases could be used for Spearman correlations, only the six phytohormones present in the 2024 and 2025 data were included: salicylic acid, jasmonic acid, jasmonoyl-L-isoleucine, abscisic acid, indole-3-acetate monooxygenase, and indole-3-acetic acid. Once correlations between exploratory ($\text{Log}_2\text{FC} \geq 2$, $\text{P-value}_{\text{unadjusted}} \leq 0.05$) unannotated metabolites, annotated metabolites, and phytohormone abundances were combined (Supplemental Table 4.5), there were 19 total correlation combinations that included four unannotated metabolites, three phytohormones, and 12 different annotated metabolites across time points one and two. All correlations were significant after the FDR adjustment ($\text{p-value}_{\text{FDR}} \leq 0.5$). Seven of the 19 correlation groups had strong positive correlations ($r \geq 0.70$) between all three factors (Table 4.2).

Phytohormone log₂FC Across Timepoints

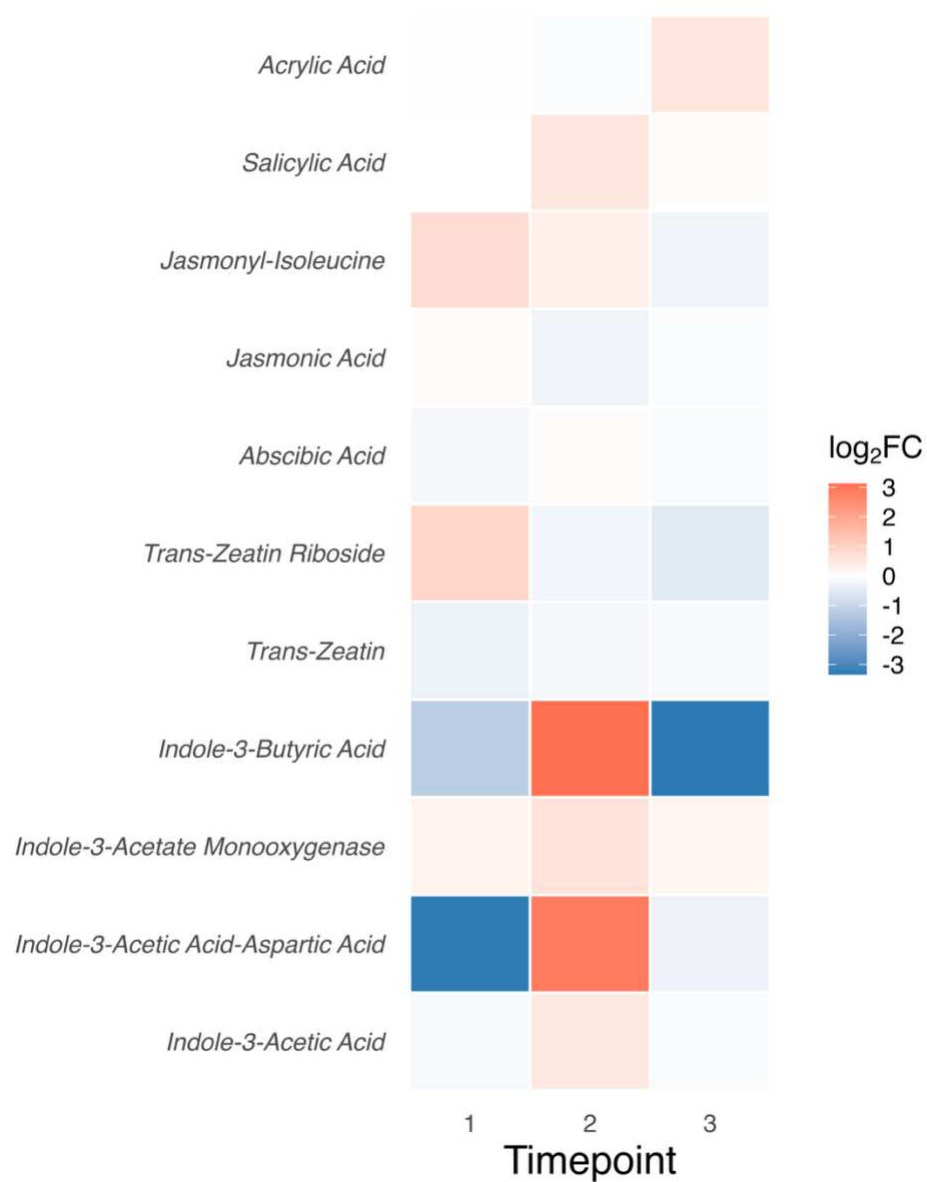


Figure 4.4. Log₂ fold change of phytohormones in Byrd CL Plus with reference to the Byrd. The y-axis indicates the phytohormone and the x-axis indicates the timepoint 1) before sawfly infestation, 2) during peak infestation, or 3) two weeks after infestation. Red indicates a positive fold change and blue indicates a negative fold change.

Table 4.2. Correlations for unannotated metabolites (UA met), phytohormones (PH), annotated metabolites (A Met), pathways and classes associated with the annotated metabolites. The correlations are labeled as follows: correlation for unannotated metabolite and phytohormone (UA Met and PH r), correlation for unannotated metabolite and annotated metabolite (UA Met and A Met r), and correlation for annotated metabolite and phytohormone (A Met and PH r).

UA Met	t	PH	A Met	Pathway	Class	UA Met and PH r	UA and A Met r	A Met and PH r
C4755	1	ABA	C1482	Fatty acids	Triacylglycerols	0.81	0.71	0.81
			C3137	Shikimates and Phenylpropanoids	Flavonols	0.81	0.72	0.82
			C4690	Terpenoids	Carotenoids	0.81	0.75	0.82
			C6226	Polyketides	Macrolide lactones	0.81	0.84	0.82
C7423	2	ABA	C2130	Shikimates and Phenylpropanoids	Flavones	0.93	0.73	0.84
C7423	2	JA-Ile	C2130	Shikimates and Phenylpropanoids	Flavones	0.83	0.73	0.82
			C2663	Terpenoids	Carotenoids	0.83	0.74	0.89

DISCUSSION

This study aimed to understand differences between the proteome, metabolome, and phytohormone profiles of two wheat cultivars differing in their level of resistance to wheat stem sawfly. Samples were collected from two different genotypes, Byrd and Byrd CL Plus, at pre-, peak, and post-infestation to understand how each genotype responded to WSS.

Across all three timepoints, a total of five different proteins were differentially expressed in Byrd CL Plus compared to Byrd. One of the downregulated proteins associated with RNA metabolism and transcription, unspecified biological processes, and unspecified metabolic processes (A0A3B6JCK1) is a homolog to the *Arabidopsis thaliana* CDC73 protein gene (AT3G22590) which is associated with high levels of *FLOWERING LOCUS C (FLC)* expression (Yu & Michaels, 2010) – a repressor of flowering. Interestingly, despite Byrd and Byrd CL Plus being NILs and historically having very similar flowering times, the two cultivars had varying heading dates in the plots used in this study with Byrd CL Plus heading out later than Byrd by 2.5 Julian days. Differences in proteins that effect flowering time could affect WSS resistance due to the known association of heading date with avoidance (Bathini et al., 2023). However, it is important to note that A0A3B6JCK1 was downregulated at peak infestation after the wheat had already flowered when *FLC* is normally downregulated after vernalization (Jean Finnegan et al., 2005).

Another protein accession of interest was A0A3B6NUD3 which was downregulated during peak infestation and annotated as a phosphoglycerate mutase that was linked to glycolysis. The glycolysis pathway has been shown to increase in other species such as *Lycium barbarum* in response to salt stress as a result of accumulating abscisic, jasmonic, and salicylic acids (W. Liang

et al., 2025). The glycolysis pathway has also been induced in multiple different plant herbivory interaction studies (Pan et al., 2020) that include pea aphids on pea plants (Carrillo et al., 2014), small brown planthopper and rice (Dong et al., 2017), and WSS and winter wheat (Lavergne et al., 2020). This highlights the complexity of the signal transduction pathways related to plant defenses and the role of traditionally non-defense associated pathways in response to abiotic and biotic stress.

As expected for NILs, there were little to no significant differences between the metabolomes of the two genotypes which can be seen by the PCA (Figure 4.2, Supplemental Figure 4.1). However, Byrd CL Plus estimates had wide confidence intervals across time points, which indicates high variability in the Byrd CL Plus sample group as compared to the Byrd sample group. While there is considerable overlap across multiple PCA groups, incomplete overlap still highlights potential differences in the metabolome between these two genotypes. This difference in confidence intervals could also be due to the missing Byrd CL Plus samples.

The metabolome profiles revealed one unannotated metabolite that was significant ($\text{Log}_2\text{FC} = 3.20$, $\text{p-value}_{\text{FDR}} = 0.02$) during peak infestation. This metabolite was not putatively annotated and was not correlated with any annotated metabolites identified in this study, indicating a potentially novel metabolite. However, out of the 330 metabolites in the exploratory threshold ($\text{Log}_2\text{FC} \geq |2|$, $\text{p-value}_{\text{unadjusted}} < 0.05$) across all timepoints, eight were able to be annotated and were associated with the following classes: tropolones and derivatives, capsacians and capsaicinoids, flavones, dipeptides, depeptides and depsidones, and microcolins and mirabimids.

Because fold changes are biologically meaningful (Lötsch et al., 2024), metabolites in the exploratory threshold with a $\text{Log}_2\text{FC} \geq 6$ and an unadjusted $\text{p-value} < 0.05$ were considered for additional analysis and correlated with annotated metabolites. Across all time points, the positively

correlated ($r \geq 0.7$, $p\text{-value}_{\text{FDR}} < 0.05$) metabolites were associated with a variety of pathways including shikimates and phenylpropanoid, flavonoid, terpenoid, phenolic and alkaloid pathways and flavones, flavonols, simple phenolic acids, amino acids, and carotenoids metabolite classes which are often associated with plant defenses (Anjali et al., 2023; Iriti & Faoro, 2009; Kumar et al., 2025). Terpene biosynthetic pathways have associated with antixenosis between tomato (*Sanum habrochaites*) and whitefly (*T. vaporariorum*) and carmine spider mites (*T. cinnabarinus*) (Y. Wang et al., 2025) and is likely related to leaf trichomes.

Alkaloids are synthesized by amino acids have been identified as a deterrent to insect herbivory in nightshade potato (*Solanum demissum*) in response to Colorado beetle (*Leptinotarsa decemlineata*) (Fürstenberg-Hägg et al., 2013). One form of alkaloids, pyrrolizidine alkaloids (PAs), can turn toxic in the digestive tracts of certain insects which also serves as a direct defense. Additionally, phenylpropanoids and flavonoids have been associated with plant defenses and responses to herbivory, however, their exact role has not been fully elucidated (Ramaroson et al., 2022). While amino acids are part of the plant defense system and a product of the pathway associated with the ALS gene for IMI herbicide resistance, nearly no positively correlated phytohormones or metabolites were associated directly with the amino acids class. However, this was expected because the CL defense mechanism still allows resistant plants to synthesize the same necessary amino acids for growth. These metabolomic results in tandem with previous literature support that, while their functions are not fully understood, secondary metabolites could provide antixenosis and antibiosis to insects.

Samples collected at peak and post-infestation would be indicators of differential responses to WSS infestation. Other studies have found differences in similar pathways in response to infestation in hollow stem wheat including the phenylpropanoid pathways, lignin content, and

phospholipids in spring wheat (*Triticum aestivum* L.) and oat (*Avena sativa* L.) (Hager et al., 2024). Due to the genetic similarity of NILs and as expected, fewer differentially expected proteins were identified between cultivars in this study compared to previous studies such as (Lavergne et al., 2020). However, some similar functioning proteins were identified between both studies such as the downregulation of protein biosynthesis molecules in response to WSS infestation. The identified metabolites in the Lavergne et al. (2020) study for the resistant hollow stem ‘Hatcher’ (PI 638512) (Haley et al., 2005) also partially mirrored the results of this study due to the significance of the alkaloid, benzenoid, organic acids, nucleosides, and phenylpropanoid/polyketide pathways. These results support the findings in this study that there are different metabolomic and proteomic responses to WSS infestation amongst cultivars, including between the lines Byrd and Byrd CL Plus.

There were no significantly different ($p\text{-value}_{\text{FDR}} < 0.05$) phytohormones identified between the two genotypes at any timepoint. Correlations between annotated metabolites, unannotated metabolites, and identified phytohormones can be used to elucidate potential pathway connections between the metabolome and phytohormone profiles. This showed that at the peak infestation timepoint, the hormones ABA and JA-Ile were positively correlated with unannotated metabolites associated with shikimates and phenylpropanoid pathway as well as terpenoid pathways. The annotated metabolites associated with the shikimates and phenylpropanoid pathway were both in the flavone class while the annotated metabolite associated with the terpenoid pathway was associated the carotenoids class.

While carotenoids are often associated with micronutrient content and photosynthesis in plants, they also provide the precursors of to the biosynthesis of abscisic acid (ABA) and can act as signaling molecules in response to environmental cues (T. Sun et al., 2022). Abscisic acid is a

major plant hormone involved in development and dormancy as well as plant responses to abiotic and biotic stresses (Mo et al., 2024). Absciscic acid plays a role in pre-infestation defense through cuticle reinforcement as well as post-infestation defense through the antagonization of immune signaling pathways related to important defense hormones such as salicylic acid (SA) and jasmonic acid (JA) (Makris et al., 2026). Jasmonyl-isoleucine (JA-Ile), the most bioactive conjugate of JA, binds to COI1 and promotes proteome mediated degradation of JASMONATE ZIM-DOMAIN (JAZ) proteins. This suppressed downstream transcription factors and JA responses, which allows for the activation of other JA-dependent processes and defense responses (P. Zhang et al., 2025). These findings highlight the relationships between the metabolites and phytohormones in this study and potential low-level differences in defense signaling pathways between Byrd and Byrd CL Plus.

Overall, Byrd CL Plus showed only slight differences in expression compared to Byrd in this study. There were no significant differentially expressed phytohormones and one differentially expressed metabolite. The unannotated metabolite could not be correlated with any other annotated metabolites to discern its potential function. Exploratory analyses of metabolites below the FDR threshold indicated potential low-level differences between the two lines but cannot be determined as conclusive evidence of differences. There were five total differentially expressed proteins between the pre-, peak, and post-infestation samples. Of the three of the proteins were well characterized, two were associated with pathways previously associated with plant defense responses. However, the glycolysis pathway proteins were previously found to be upregulated in response to stress, and in this study were downregulated in Byrd CL Plus in reference to Byrd. Understanding genomic regions associated with these features could also help determine potential relationships with the proteins identified in this study. None of the five identified proteins were

associated with the same chromosome as *imi1* or *imi2*, which could indicate that these proteins are linked to other genome regions from the parent that provided the Clearfield® trait. While responses in the given germplasm do not pass the most stringent criteria for significance, they can still provide information about differences between Byrd and Byrd CL Plus that could be responsible for differences in infestation levels between the two genotypes.

The phytohormone and metabolome results highlight the complexity of biological pathways and the breadth of connections between biological features. While there were prominent and exploratory features that differed between these two genotypes, further analysis would be required to ascertain if these pathways were related to the *imi1* and *imi2* genes associated with Clearfield® tolerance in Byrd CL Plus. Mapping methods such as metabolite quantitative trait loci (mQTL) can be utilized in bi-parental populations to further understand genomic loci associated with both metabolites and phytohormones (Fernie & Schauer, 2009; Kliebenstein, 2009; B. Yin et al., 2024) or in QTL studies to understand if linkage with the herbicide resistant donor line or downstream effects of the silent mutations in the ALS enzyme (Raamsdonk et al., 2001) is conferring additional resistance.

CONCLUSIONS

To investigate potential sources of antixenosis-based resistance to WSS in CL cultivars, proteomic, metabolomic, and phytohormone profiles were compared between Byrd and Byrd CL Plus. While preliminary observations suggest that CL cultivars may exhibit increased resistance to WSS, this study aimed to determine differences in molecular responses to WSS infestation. Across all datasets, differences between Byrd and Byrd CL Plus were limited but biologically informative. The limited number of significant features is consistent with the high genetic similarity of NILs,

143 but the patterns observed support the hypothesis that small-scale differences in metabolic and
144 signaling pathways may influence resistance. Overall, these results highlight the complexity of
145 WSS resistance and reinforce that mechanisms beyond the solid stem trait may contribute to
146 reduced infestation. While the solid stem trait remains a cornerstone of resistance, integrating
147 additional mechanisms could improve the durability and effectiveness of resistant cultivars and is
148 important for the long-term resilience of farmer fields.

149 **CHAPTER V – UTILIZING MULTI-OMICS TO UNDERSTAND MOLECULAR**
150 **MECHANISMS OF SOLID STEM RESISTANCE IN WINTER WHEAT (*Triticum***
151 ***aestivum* L.)**

152
153
154 **INTRODUCTION**

159 Wheat stem sawfly (WSS; *Cephus cinctus* Norton) is a native grass-feeding insect and a
160 major pest of several agronomically important grasses, including common wheat (*Triticum*
161 *aestivum* L.), durum wheat (*Triticum durum*), barley (*Hordeum vulgare*), and triticales
162 (×*Triticosecale* Wittmack) (Beres et al., 2012). Although WSS has been recognized as a pest in
163 the northern United States since the 19th century, shifts in climatic patterns have contributed to
164 emergence earlier in the season, aligning WSS activity with winter wheat development in the Great
165 Plains region of the United States (Peirce et al., 2022). Currently, annual financial losses associated
166 with WSS damage in North America are estimated at approximately \$350 million USD (Beres et
167 al., 2011).

168 Most of the WSS life cycle occurs within the stem of the host plant, making the insect
169 difficult to manage through conventional cultural or chemical control methods (Beres et al., 2011).
170 Adult sawflies emerge from crop residue in late spring and oviposit eggs into host plant stems
171 (Ainslie, 1920). After hatching, larvae feed on the stem pith and cannibalize one another until a
172 single larva remains. The surviving larva continues feeding along the length of the stem before
173 moving to the stem base, where it creates a hibernaculum and enters diapause (Ainslie, 1920; Beres
174 et al., 2007). Formation of the hibernaculum effectively cuts the stem near the base, weakening
175 the tiller and predisposing the plant to lodging. Because traditional pest management strategies are

often ineffective against WSS, host plant genetic resistance is a critical component of crop protection in WSS-infested production systems.

The most well-characterized source of genetic resistance to WSS is stem solidity. This trait is primarily controlled by a locus on the long arm of chromosome 3B (*SStI*) (Cook et al., 2004a; Nilsen et al., 2020) along with a homeologue on chromosome 3D (*Qss.msub-3BL*) (Lanning et al., 2006). The *SStI* locus accounted for approximately 76% of phenotypic variation in stem solidity in a recombinant inbred line (RIL) population (Cook et al., 2004a) and is regulated by copy number variation of a *TdDof* transcription factor (Nilsen et al., 2020). A single copy of this transcription factor confers a hollow stem phenotype and three copies result in semi-solid to solid stem expression.

Stem solidity is determined by the persistence of parenchyma cells within the pith. In hollow-stemmed wheat, these cells degrade as the plant approaches maturity (Ainslie, 1920), forming a large central cavity (Kebrom et al., 2012; Kirby, n.d.). In contrast, solid-stemmed wheat exhibits reduced or incomplete pith degradation, resulting in a smaller cavity and increased stem tissue density. This structural difference contributes to WSS resistance through several mechanisms, including mechanical interference with larval movement (Beres et al., 2012; Holmes & Peterson, 1960, 1964), increased stem moisture (Holmes & Peterson, 1964), and reduced larval feeding efficiency (Evert et al., 2006; Kemp, 1934). Solid stem plants have reduced lodging from WSS damage, mitigating yield loss (Beres et al., 2007).

Despite its effectiveness, expression of the solid stem trait is environmentally sensitive and is often reduced under cool or wet growing conditions (Beres et al., 2012; Holmes, 1984; Platt & Farstad, 1946). This variability limits the consistency of WSS resistance, highlighting the need for

improved understanding of the genetic and physiological mechanisms underlying stem solidity to develop cultivars with stable resistance across environments.

Beyond *SStI*, additional loci contributing to stem solidity have been identified. Minor quantitative trait loci (QTL) have been reported on chromosomes 2D and 5A in common wheat (Nilsen et al., 2017), chromosome 3AL in tetraploid wheat (*Triticum turgidum* L.) (Varella, Zhang, et al., 2019) and chromosome 1B in hexaploid wheat landraces (Varella, Weaver, et al., 2019a). Furthermore, haplotype variation at the 3B locus has been characterized, including the Reeder allele (*Qss.msub-3BLa*), which confers a hollow stem; the Choteau allele (*Qss.msub-3BLb*), which maintains stem solidity throughout development; and the Conan allele (*Qss.msub-3BLc*), which confers early stem solidity followed by a loss of pith integrity as the plant approaches maturity due to assimilate repartitioning (Cook et al., 2019a; Varella et al., 2015; Varella, Weaver, et al., 2019b). While substantial progress has been made in characterizing the genetic architecture of stem solidity, recent work has increasingly focused on integrating multi-omics approaches to better understand the molecular mechanisms underlying WSS resistance.

Lavergne et al. (2020) evaluated four wheat genotypes ‘Hatcher’ (PI 638512) (Haley et al., 2005), ‘Denali’ (PI 664256) (Haley, Johnson, Peairs, Stromberger, Hudson, Seifert, Kottke, Valdez, Rudolph, Martin, et al., 2012), ‘Conan’ (PI 607549), and ‘Reeder’ (PI 613586) using integrated stem proteomic and metabolomic analyses in response to WSS infestation. Significant differences in both the metabolome and proteome were observed between infested and control treatments across all cultivars. Several metabolic pathways potentially contributing to WSS resistance were identified, including detoxification processes, proteinase inhibition, systemic signaling, and calcium (Ca^{2+}) flux. A similar omics study by Hager et al. (2025) further

220 investigated these mechanisms using the hollow-stem cultivar ‘Reeder’, the semi-solid cultivar
221 ‘Conan’, and derived near-isogenic line (NIL) populations. Differences in metabolite abundance
222 were observed among genotypes carrying hollow stem, early solidness, and fully solid stem alleles
223 across multiple developmental stages. Across these groups, pathways related to phenylpropanoid
224 biosynthesis, carbohydrate metabolism, and programmed cell death were consistently associated
225 with WSS resistance. Collectively, these studies demonstrate that multi-omics approaches provide
226 valuable insight into the biochemical and molecular differences underlying resistance between
227 hollow- and solid-stemmed wheat.

228 Despite these advances, key questions remain regarding the physiological mechanisms
229 governing stem solidity and the environmental stability of its expression. To further investigate
230 these processes, proteomic, metabolomic, and phytohormone analyses were conducted using a set
231 of NILs derived from the hard red winter wheat breeding lines. One line, ‘CO20SFD047R.4’, had
232 the hollow stem allele and the other, ‘CO20SFD047R.1’, had the solid stem allele. Field trials were
233 performed across two growing seasons to evaluate differences in these molecular profiles in
234 response to WSS infestation. Improving our understanding of the pathways underlying these
235 contrasting genotypes will provide new insight into the biological basis of stem solidity and
236 support the development of cultivars with more stable resistance to WSS.

MATERIALS AND METHODS

Germplasm

Two near isogenic lines (NILs) derived from the cross CO16SF029/CO16SF070 from the Colorado State University Wheat Breeding and Genetics program in Fort Collins, Colorado were used for this study. These NILs were generated from breeding germplasm in the Colorado State University Wheat Breeding and Genetics program in Fort Collins, Colorado using a marker-assisted selection approach in heterogeneous F_{3:5} lines or doubled haploids. Briefly, single heads were selected from segregating F₃ bulk populations in summer 2019, planted as F_{3:4} rows in Fall 2019, and bulk harvested in summer 2020. The developed F_{3:5} lines were subsequently evaluated in yield trials during the 2020-2021 season. In the F_{3:5} generation or from doubled haploid lines, ~20 single heads were again selected from individual plants under the hypothesis that some lines would be heterogenous at *SStI* thus containing individuals that were WT (non-solid) and MT (solid). Individuals from each line were genotyped with markers for the *SStI* locus (*usw275*) (Nilsen et al., 2020) and selected individuals were planted as F_{5:6} rows, thus creating the NIL groups. Seed of individual rows was bulk harvested to produced seed for subsequent yield trials.

These lines were chosen as part of this NIL set based on solid stem ratings where CO20SFD047R.4 averaged around the lowest solidity score of 5 and CO20SFD047R.1 having an average solidity score of 18.8 in the 2023 season on a visual ordinal scale of 5 to 25. This large difference in solidity and fixing of genetic backgrounds was considered most ideal for understanding differences between the hollow and solid stem trait.

Sample Collection

Samples were collected at the United States Division of Agriculture – Agricultural Research Service (USDA-ARS) Central Great Plains Research Station in Akron, Colorado during the 2024 and 2025 field seasons. Three biological replicates of both genotypes were grown in 1.52 m x 3.05 m yield plots in a completely random design that included three replicates of each genotype. Trials were planted into soils with a N:P ratio of 82.6:7.7 kg and 59.9:21.8 kg for the 2024 and 2025 field seasons, respectively. For the 2025 season, 2.0 kg N and 7.2 kg P were applied for fertilizer at planting. No fertilizer was applied mid-season for either year; no additional management was performed before harvest. Sawfly flight was determined by performing sweeps on a 30 m field section planted with one cultivar. Sawfly caught in the sweep were counted to determine presence of WSS and peak flight. Samples were collected at three time points: before WSS flight, just past peak WSS flight, and two weeks after peak WSS flight. In 2024, samples were collected on May 9th, May 29th, and June 13th, when the wheat was at approximately Zadok's scale 32, 59, and 69, respectively. With the two genotypes, three replicates, and three time points, there were a total of 18 samples collected in the 2024 field season. In 2025, samples were collected on May 12th, May 29th, and June 12th, when the wheat was at approximately Zadok's scale 32, 59, and 69, respectively. With the two genotypes, three replicates, and three time points, there were a total of 18 samples collected in the 2025 field season.

In total, across both years, there were 6 replicates of each line at all three time points. For consistency, all samples were collected between 10 a.m. and 12 p.m. For sample collection, the flag leaf was removed from a single stem before the stem was cut from the plot. For the samples collected before WSS flight, the stems were flash frozen in liquid nitrogen (N₂) immediately after being removed from the plot. For samples collected at peak WSS flight or two weeks post peak

WSS flight, the selected stem was then split long-ways from the head down to the first node on the stem. This section was assessed for the presence of WSS eggs, larvae, and/or frass.

Once identified, eggs, larvae, and/or frass were removed from the stem and region between the wheat head and first node of the stem to reduce cross-species contamination prior to flash freezing in liquid nitrogen. Two stems were collected from each plot to provide ample tissue for each genotype, replicate, and timepoint. Tools used in tissue sampling were cleaned with 70% ethanol (C₂H₅OH) between samples to avoid cross-contamination. Samples were then transported from Akron, Colorado to Fort Collins, Colorado on dry ice (CO₂). Samples were transferred from dry-ice to -70°C freezer until sample preparation.

Sample and Data Preparation and Statistical Analysis

Sample preparation of stem samples for proteomic, metabolomic, and phytohormone profiling in this chapter follow the methods described in chapter three (page 83). Additionally, peak detection, annotation, and statistical analysis were all performed using the same parameters and models as described in the chapter 3 methods.

RESULTS

Variation in the Proteome

Liquid-chromatography mass spectrometry analysis of the NILs resulted in the identification of 152,660 peptides annotated as 14,876 protein accessions, including 9,190 accessions identified using at least two unique peptides. These filtered accessions were linked to 12 GO biological processes. Some accessions overlapped between processes, there were a total of 893 accessions associated with transport, 950 with cell organization and biosynthesis, 1,154 with

protein metabolism, 571 with RNA metabolism or transcription, 251 with signal transduction, 181 with DNA metabolism, 145 with cell cycle or cell proliferation, 768 with stress response, 156 with developmental processes, 3,805 with other metabolic process, 1,313 with other biological processes, and 3,068 with no identified process.

For multivariate analysis, proteins were filtered to only include accessions that were detected in at least 60% of all replicates for a genotype at each time point, which resulted in 6,399 unique accessions. There was variation in the proteome between the hollow stem and solid stem lines including one significantly ($\text{Log}_2\text{FC} \geq |2|$, $\text{p-value}_{\text{FDR}} \leq 0.05$) differentially expressed protein pre-infestation, two during peak infestation, and four post-infestation (Figure 5.1; Table 5.1). None of the significant accessions were significant in multiple timepoints. Three of the seven accessions had assigned GO biological terms identified. At peak infestation A0A3B6DL43 ($\text{Log}_2\text{FC} = 3.20$, $\text{p-value}_{\text{FDR}} = 0.02$) was associated with unspecified metabolic processes. Post-infestation, A0A3B6NPX7 ($\text{Log}_2\text{FC} = -3.32$, $\text{p-value}_{\text{FDR}} = 0.01$) and A0A3B6HPQ7 ($\text{Log}_2\text{FC} = 3.73$, $\text{p-value}_{\text{FDR}} = 0.01$) were associated with transport and unspecified biological processes and cell organization and biogenesis, respectively.

Presence-absence analysis of protein accessions was performed by identifying accessions which had protein abundances in one genotype and unmeasurable protein abundances in another genotype using a Fisher's test. No accessions were significant in this analysis (Supplemental Table 5.1).

Solid Stem NILs Protein Expression

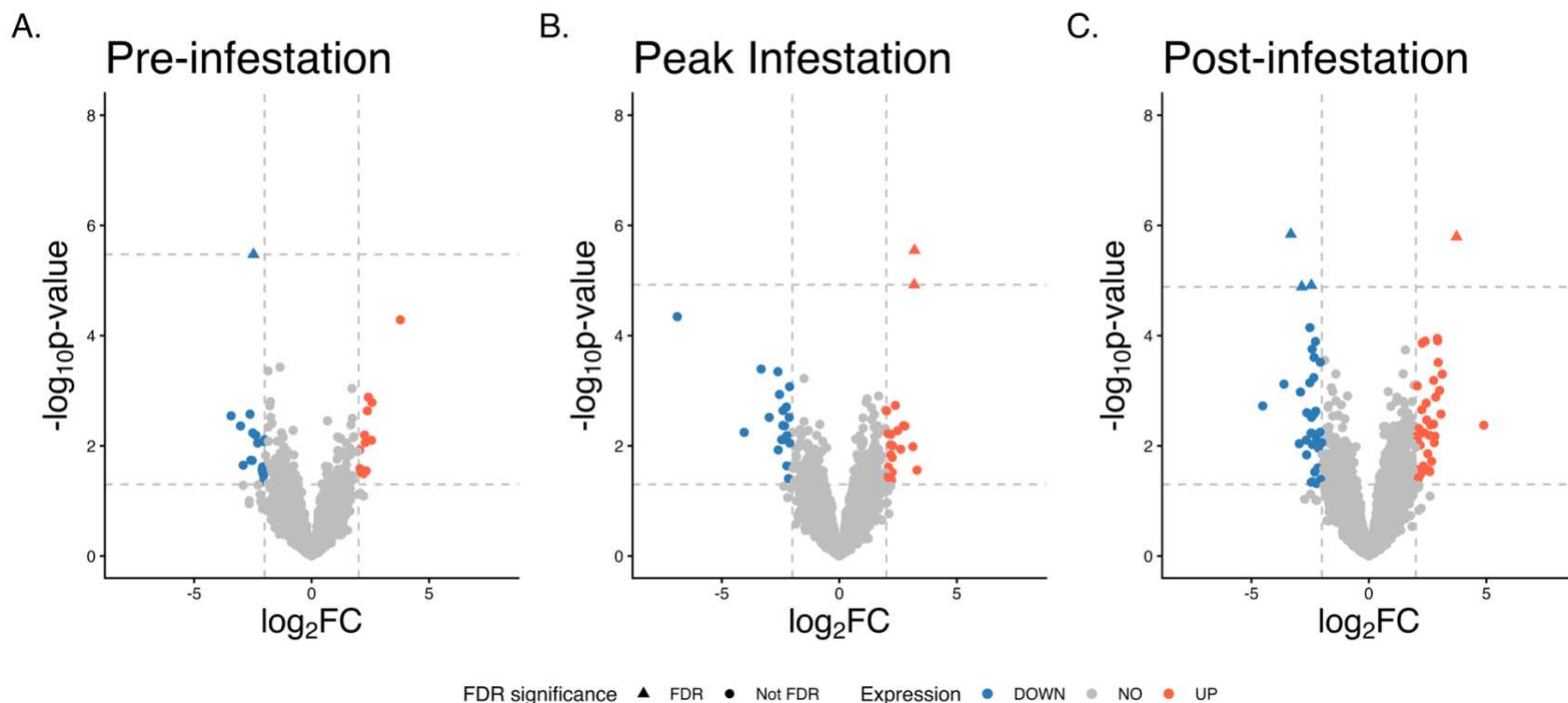


Figure 5.1. Volcano plots of differential protein expression between the hollow and solid stem lines at A) pre-infestation, B) peak infestation, and C) post-infestation. The vertical dashed lines indicate a $\text{Log}_2FC > |2|$ and the lower and upper horizontal dashed lines indicate the $p\text{-value}_{\text{unadjusted}}$ threshold of 0.05 and the $p\text{-value}_{\text{FDR}}$ threshold of 0.05, respectively. Blue points indicate accessions that were downregulated, red indicates accessions that were upregulated, and triangular points indicate accessions that surpass the FDR adjusted threshold.

Table 5.1. Significant differentially expressed protein accessions between the hollow and solid stem lines for pre-, peak, and post-infestation where time indicates pre-infestation (1), peak infestation (2), or post-infestation (3), and Log₂FC and p-value_{FDR} indicate the significance of the protein accession in terms of fold change and adjusted p-value, GO biological process indicates identified functions of the protein accession, the Ensembl gene ID for each of the protein accessions, and the chromosome (Chr) where the gene is located.

Time	Accession	Log ₂ FC	p-value _{FDR}	GO Biological Process	Ensembl Gene ID	Chr
1	A0A3B6UCR7	-2.47	0.02	-	TraesCSU02G235600	Un
2	A0A3B6DL43	3.20	0.02	other metabolic processes	TraesCS2D02G502400	2D
	A0A3B6B301	3.19	0.04	-	TraesCS2A02G367700	2A
3	A0A3B6NPX7	-3.32	0.01	transport; other biological processes	TraesCS6A02G197400	6A
	A0A3B6HPQ7	3.73	0.01	cell organization and biogenesis	TraesCS4A02G074200	4A
	A0A3B6SBN0	-2.45	0.02	-	TraesCS7B02G013300	7B
	A0A3B6C4R7	-2.86	0.02	-	TraesCS2B02G184700	2B

Variation in the Metabolome

A total of 7,591 putative compounds were detected. Of these compounds, 104 were annotated as distinct metabolites with a SIRIUS score greater than 80%. Annotated metabolites were classified into 110 unique classes, 45 unique superclasses, and seven different pathways including alkaloids, amino acids and peptide, carbohydrates, fatty acids, polyketides, shikimates and phenylpropanoids, and terpenoids. Unannotated metabolites could not be identified due to only making it through the MS1 phase and not MS2 or due to low confidence matching for the chemical structure at that peak. Differences in metabolomes between genotypes and timepoints were inferred by principal component analysis (PCA) (Figure 5.2; Supplemental Figure 5.1). The first, second, and third PCs accounted 41.9%, 9.5%, and 8.1% of total variation, respectively.

After FDR correction of p-values, no metabolites were significant between the hollow and solid stem line at any of the three time points. However, notable differentially expressed metabolites ($\text{Log}_2\text{FC} > |2|$, $\text{p-value}_{\text{unadjusted}} < 0.05$) were identified at all three time points (Figure 3), none of which were annotated. Pre-infestation, there were three metabolites that had a $\text{Log}_2\text{FC} < -5$ ($\text{p-value}_{\text{unadjusted}} < 0.05$) and 13 metabolites that had a $\text{Log}_2\text{FC} > 5$ ($\text{p-value}_{\text{unadjusted}} < 0.05$). At peak infestation, there were eight metabolites that had a $\text{Log}_2\text{FC} < -5$ ($\text{p-value}_{\text{unadjusted}} < 0.05$) and six metabolites that had a $\text{Log}_2\text{FC} > 5$ ($\text{p-value}_{\text{unadjusted}} < 0.05$). Post-infestation, there were five metabolites that had a $\text{Log}_2\text{FC} < -5$ ($\text{p-value}_{\text{unadjusted}} < 0.05$) and nine metabolites that had a $\text{Log}_2\text{FC} > 5$ ($\text{p-value}_{\text{unadjusted}} < 0.05$).

All Spearman metabolite correlations were significant after FDR correction ($\text{p-value}_{\text{FDR adjusted}} < 0.05$) (Supplemental Table 2). Pre-infestation, there were six unannotated metabolites significantly correlated ($r \geq 0.70$, $\text{p-value}_{\text{FDR adjusted}} < 0.05$) with one or more annotated metabolites. The pathways associated with the annotated metabolites included shikimates and

phenylpropanoids, amino acids and peptides, terpenoids, carbohydrates, terpenoids, and alkaloids. Associated classes included simple phenolic acids, flavones, flavonols, tripeptides, pyridine alkaloids, diacylglycerols, linear peptides, aminosugars, carotenoids, capsaicins and capsaicinoids, microcolins and mirabimids, pyridine alkaloids, phenylalanine-derived alkaloids, and tripeptides.

At peak infestation, there were 12 significantly correlated metabolites ($r \geq 0.70$, $p\text{-value}_{\text{FDR adjusted}} < 0.05$) with one or more annotated metabolites. The pathways associated with the annotated metabolites included shikimates and phenylpropanoids, amino acids and peptides, terpenoids, carbohydrates, alkaloids, polyketides, and fatty acids. Associated classes included anthroquinones and anthrones, dipeptides, tripeptides, microcolins and mirabimids, diacylglycerols, flavones, flavonols, flavonones, pyrimidine nucleosides, limonoids, and linear peptides. Post-infestation, there were three unannotated metabolites significantly correlated ($r \geq 0.70$, $p\text{-value}_{\text{FDR adjusted}} < 0.05$) with one or more annotated metabolites. The pathways associated with the annotated metabolites included shikimates and phenylpropanoids, amino acids and peptides, terpenoids, polyketides, and alkaloids. Associated classes included carotenoids, flavonols, flavones, capsaicins and capsaicinoids, tripeptides, phenylalanine-derived alkaloids, macrolide lactones, simple phenolic acids, and linear peptides. For all three time points there were multiple negatively correlated annotated metabolites from all the same pathways.

PCA of Metabolite Expression

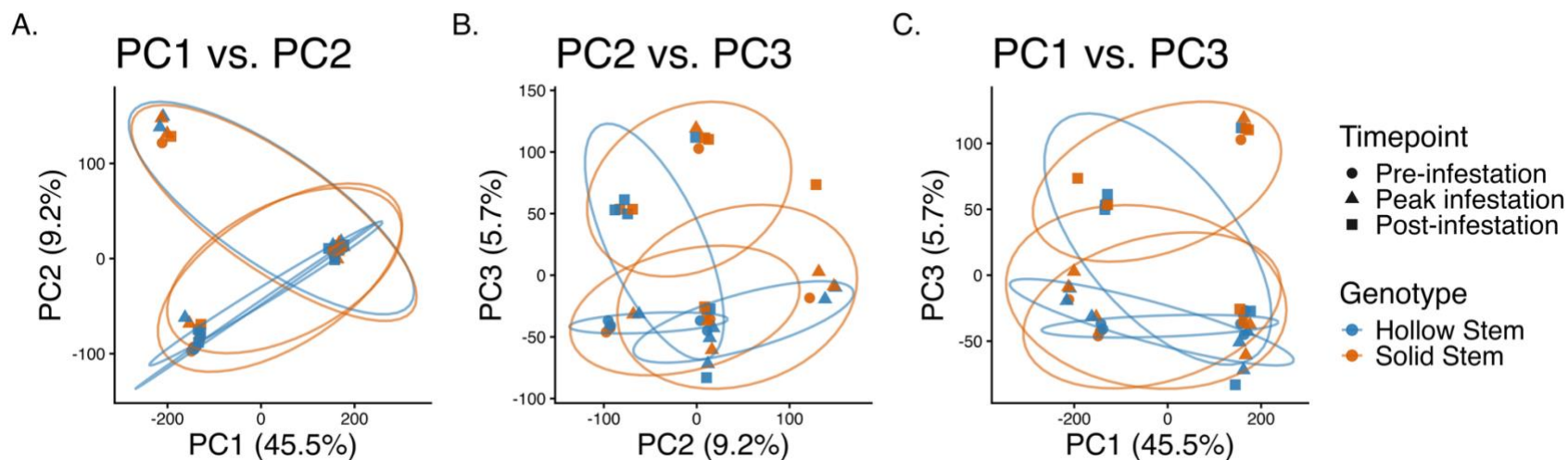


Figure 5.2. Principal component (PC) plots for metabolite expression where (A) shows variance explained by PC1 and PC2, (B) shows PC2 and PC3, and (C) shows PC1 and PC3. Blue ellipses and datapoints indicate the hollow stem line and orange ellipses and datapoints indicate the solid stem line. Timepoint is differentiated by shape where circles indicate pre-infestation samples, triangles indicate peak infestation samples, and squares indicate post-infestation samples. Ellipses indicate the spread of the genotypes at different timepoints.

Solid Stem NILs Metabolite Expression

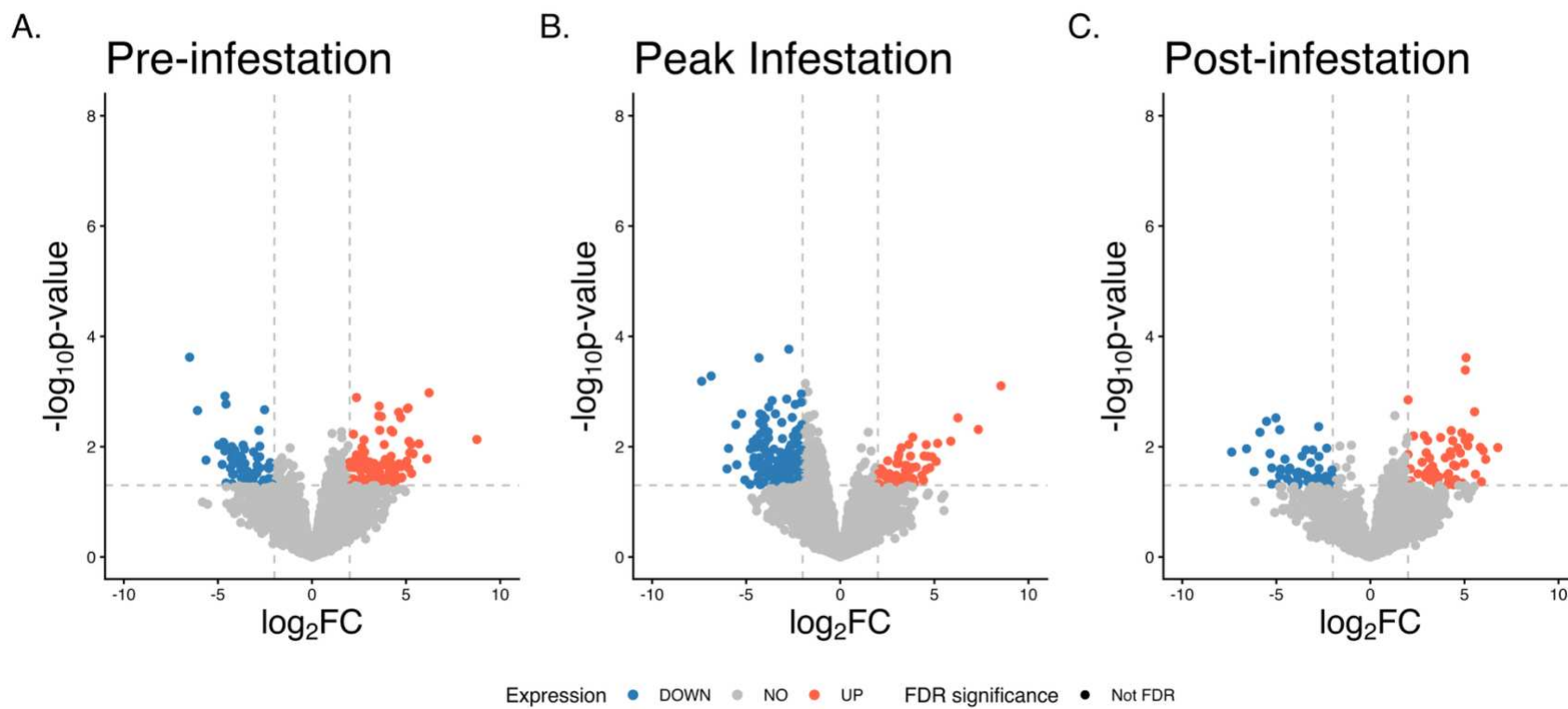


Figure 5.3. Volcano plots indicating differentially expressed metabolites in the solid stem line in reference to the hollow stem lines where A) represents samples collected pre-infestation, B) represents samples collected during peak infestation, and C) represents samples collected two weeks after peak infestation. The x-axis indicates the log₂ fold change and the y-axis indicates the -log₁₀(p-value). Each dot represents a metabolite where grey indicates a metabolite with an unadjusted p-value < 0.05 and a Log₂ fold change between -2 and 2, red indicates a metabolite with an unadjusted p-value > 0.05 and a Log₂ fold change greater than 2, and blue indicates a metabolite with an unadjusted p-value > 0.05 and a Log₂ fold change less than -2.

Variation in the Phytohormone Profile

Sixteen different phytohormone abundances were measured for both years of data (Figure 5.4). Eleven of the phytohormones were identified in both years and were considered for multivariate analysis including: salicylic acid (SA), acrylic acid (AA), jasmonic acid (JA), jasmonyl-isoleucine (JA-Ile), indole-3-butyric acid (IBA), abscisic acid (ABA), indole-3-acetic acid (IAA), indole-3-acetic acid-aspartic acid (IAA-Asp), indole-3-acetate monooxygenase (IACA), trans-zeatin (tZ), and trans-zeatin riboside (tZR). After FDR correction of p-values, only one was significant after the multivariate analysis. In the samples collected just after peak infestation, JA-Ile had a -1.6 Log₂FC between the two lines ($p\text{-value}_{\text{FDR}} = 0.02$), which indicates a 60% decrease in JA-Ile in the solid stem line compared to the hollow stem line. The same phytohormone was significant in samples collected two weeks after peak sawfly infestation, however, at this time point there was a 1.5 Log₂FC between the two lines ($p\text{-value}_{\text{FDR}} = 0.04$), indicating a 50% increase in JA-Ile in the solid stem line compared to the hollow stem line.

Because only complete cases could be used for correlations, only the six phytohormones present in the 2024 and 2025 data were included: salicylic acid, jasmonic acid, jasmonoyl-L-isoleucine, abscisic acid, indole-3-acetate monooxygenase, and indole-3-acetic acid. Once correlations between unannotated metabolites, annotated metabolites, and phytohormone abundances were combined (Supplemental Table 5.3), there were 193 total correlation combinations that included thirteen unannotated metabolites, six phytohormones, and 49 different annotated metabolites across time points one, two, and three. All correlations were significant after the FDR adjustment ($p\text{-value}_{\text{FDR}} \leq 0.5$). Thirty three of the 193 correlation groups had strong positive correlations ($r \geq 0.70$) between all three factors (Table 5.2).

Phytohormone log₂FC Across Timepoints

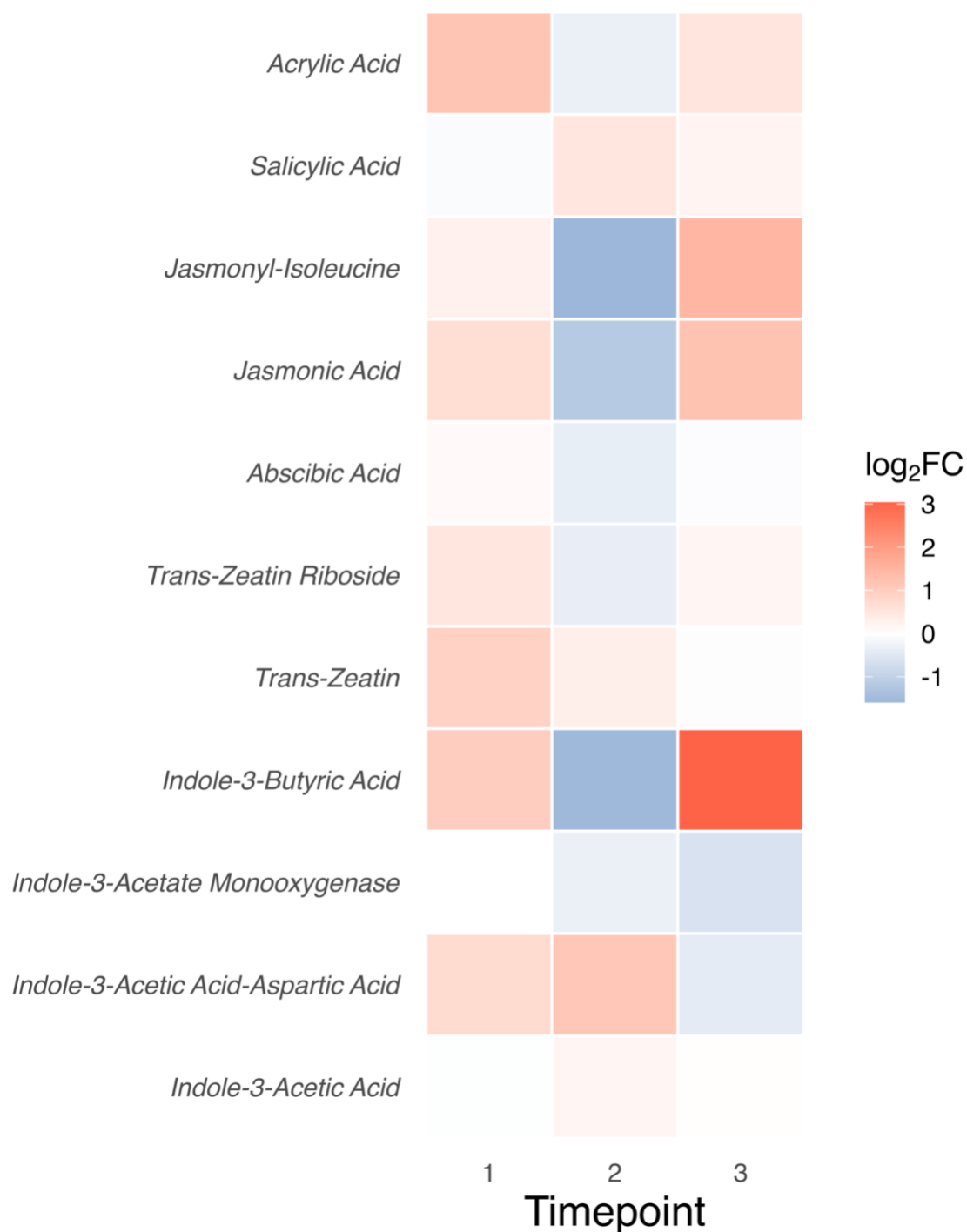


Figure 5.4. Log₂ fold change of phytohormones in the solid stem line with reference to the hollow stem line. The y-axis indicates the phytohormone and the x-axis indicates the timepoint 1) before sawfly infestation, 2) during peak infestation, or 3) two weeks after infestation. Red indicates a positive fold change, and blue indicates a negative fold change.

Table 5.2. Correlations for unannotated metabolites (UA met), phytohormones (PH), annotated metabolites (A Met), pathways and classes associated with the annotated metabolites. The correlations are labeled as follows: correlation for unannotated metabolite and phytohormone (UA Met and PH *r*), correlation for unannotated metabolite and annotated metabolite (UA Met and A Met *r*), and correlation for annotated metabolite and phytohormone (A Met and PH *r*).

UA Met	t	PH	A Met	Pathway	Class	UA Met and PH <i>r</i>	UA Met and A Met <i>r</i>	A Met and PH <i>r</i>
C4755	1	ABA	C6731	Shikimates and Phenylpropanoids	Flavones	0.79	0.74	0.90
			C3916	Amino acids and Peptides	Tripeptides	0.79	0.78	0.74
			C2130	Shikimates and Phenylpropanoids	Flavones	0.79	0.78	0.74
			C1322	Amino acids and Peptides	Pyridine alkaloids	0.79	0.80	0.84
			C3062	Terpenoids	Diacylglycerols	0.79	0.80	0.73
C5600	1	ABA	C5075	Shikimates and Phenylpropanoids	Flavonols	0.78	0.71	0.78
C6882	1	JA-Ile	C1368	Shikimates and Phenylpropanoids	Arylnaphthalene and aryltetralin lignans	0.87	-0.71	-0.71
C7261	1	ABA	C6731	Shikimates and Phenylpropanoids	Flavones	0.90	0.72	0.90
			C5075	Shikimates and Phenylpropanoids	Flavonols	0.90	0.72	0.78
			C6121	Alkaloids	Capsaicins and Capsaicinoids	0.90	0.74	0.78
			C3062	Terpenoids	Diacylglycerols	0.90	0.74	0.73
			C1291	Amino acids and Peptides	Microcolins and mirabimids	0.90	0.74	0.72
			C1322	Amino acids and Peptides	Pyridine alkaloids	0.90	0.76	0.84
			C3916	Amino acids and Peptides	Tripeptides	0.90	0.81	0.74
C5520	2	JA-Ile	C0210	Fatty acids	Wax monoesters	0.75	-0.71	-0.75
			C2032	Carbohydrates	Purine nucleosides	0.75	-0.71	-0.83
			C0767	Carbohydrates	Purine alkaloids	0.75	-0.71	-0.85

C6728	2	JA	C1482	Fatty acids	Triacylglycerols	0.74	0.71	0.76
			C5075	Shikimates and Phenylpropanoids	Flavonols	0.74	0.73	0.72
			C2130	Shikimates and Phenylpropanoids	Flavones	0.74	0.73	0.80
			C3137	Shikimates and Phenylpropanoids	Flavonols	0.74	0.80	0.78
C6099	2	IAA	C1006	Amino acids and Peptides	Dipeptides	0.80	0.72	0.88
C6153	2	IAA	C0611	Polyketides	Limonoids	0.85	0.71	0.83
			C0572	Amino acids and Peptides	Tripeptides	0.85	0.79	0.86
C6198	2	JA-Ile	C2130	Shikimates and Phenylpropanoids	Flavones	0.75	0.73	0.80
C6182	3	SA	C5075	Shikimates and Phenylpropanoids	Flavonols	0.76	0.72	0.73
			C0120	Alkaloids	Phenylalanine-derived alkaloids	0.76	0.74	0.71
			C6226	Polyketides	Macrolide lactones	0.76	0.76	0.72
			C2130	Shikimates and Phenylpropanoids	Flavones	0.76	0.79	0.72
			C5075	Shikimates and Phenylpropanoids	Flavonols	0.71	0.71	0.73
			C3137	Shikimates and Phenylpropanoids	Flavonols	0.71	0.72	0.82
			C2130	Shikimates and Phenylpropanoids	Flavones	0.71	0.83	0.72
C6182	3	ABA	C5075	Shikimates and Phenylpropanoids	Flavonols	0.87	0.72	0.87
			C6121	Alkaloids	Capsaicins and Capsaicinoids	0.87	0.73	0.82
			C6226	Polyketides	Macrolide lactones	0.87	0.76	0.90
			C2130	Shikimates and Phenylpropanoids	Flavones	0.87	0.79	0.94

C5075	Shikimates and Phenylpropanoids	Flavonols	0.75	0.71	0.87
-------	------------------------------------	-----------	------	------	------

DISCUSSION

This study aimed to understand differences between the proteome, metabolome, and phytohormone profiles of hollow stem and solid stem wheat. Samples were collected from two genotypes, CO20SFD047R.4 (hollow stem) and CO20SFD047R.1 (solid stem), at pre-, peak, and post-infestation to evaluate genotype-specific responses to WSS.

Across all three time points, a total of seven proteins were differentially expressed in the solid stem line relative to the hollow stem line. One accession (A0A3B6NPX7) ($\text{Log}_2\text{FC} = -3.32$, $\text{p-value}_{\text{FDR}} = 0.01$), identified as a C2 domain-containing protein, identified as a C2 domain-containing protein, was downregulated post-infestation in solid stem wheat relative to hollow stem wheat. This protein is associated with lipid transport and protein retention and targeting in the Golgi apparatus and vacuole, respectively. While lipids are known to play roles in plant defense (Kuzniak & Gajewska, 2024), reduced lipid transport in the solid stem genotype may reflect decreased metabolic activity as the plant senesces (Thompson et al., 1998) and as excess parenchyma cells in the stem begin to break down. Another accession (A0A3B6SBN0) ($\text{Log}_2\text{FC} = -2.45$, $\text{p-value}_{\text{FDR}} = 0.02$), identified as a hexosyltransferase (β -1,3-galactosyltransferase), was also downregulated post-infestation in the solid stem line. UDP-galactose transporters are important for the synthesis of cell wall components (Norambuena et al., 2002), and reduced expression of this protein may further indicate structural changes associated with pith degradation as the plant approaches maturity. Both significant proteins are biologically relevant for the solid stem trait. Other significant protein accessions were identified but remain less well characterized.

As expected for NILs, relatively few differences were observed between the metabolomes of the two genotypes, as illustrated by PCA (Figure 5.2, Supplemental Figure 5.1). However, the solid stem genotype exhibited larger confidence intervals across time points, suggesting greater variability relative to the hollow stem genotype. Although there was considerable overlap among PCA groupings, incomplete overlap indicates subtle but potentially meaningful metabolic differences between the genotypes.

No metabolites met the significance threshold at any time point ($|\text{Log}_2\text{FC}| \geq 2$, FDR-adjusted p-value < 0.05). However, among the 255 metabolites identified using an exploratory threshold ($|\text{Log}_2\text{FC}| \geq 2$, unadjusted p-value < 0.05), 10 were annotated and associated with diverse compound classes, including open-chain polyketides, capsaicinoids, triacylglycerols, dipeptides, pyrimidine nucleosides, anthraquinones, purine alkaloids, chromones, and phenylethylamines. Because fold changes can provide biologically meaningful insight (Lötsch et al., 2024), metabolites meeting a more stringent exploratory threshold ($|\text{Log}_2\text{FC}| \geq 6$, unadjusted p-value < 0.05) were further analyzed through correlation with annotated metabolites. Across all time points, positively correlated metabolites ($r \geq 0.7$, FDR-adjusted p-value < 0.05) were associated with pathways including shikimate, phenylpropanoid, flavonoid, terpenoid, phenolic, and alkaloid biosynthesis.

These pathways are linked to secondary metabolite classes such as flavones, flavonols, simple phenolic acids, amino acids, and carotenoids, which are frequently implicated in plant defense responses (Anjali et al., 2023; Iriti & Faoro, 2009; Kumar et al., 2025). Supporting this finding in solid stem wheat, Ma et al. (2019) demonstrated in *D. odorifera* that parenchyma cells, similar to those found in wheat stem pith, accumulate flavonoids, particularly as cellular function

declines prior to programmed cell death. This suggests that increased abundance of secondary metabolites in solid stem wheat may be associated with the persistence and eventual degradation of pith cells.

Samples collected at peak and post-infestation are most indicative of differential responses to WSS. Consistent with exploratory results in this study, previous omics analyses have identified changes in phenylpropanoid pathways (Hager et al., 2024, 2025) and the pentose phosphate pathway (Biyiklioglu et al., 2018) in response to infestation. The pentose phosphate pathway produces energy in the form of NADPH which is used in multiple lipid biosynthesis pathways (Stincone et al., 2015). Lipids play a role in passive plant defenses such as structural barriers and antimicrobial activity as well as responses to plant-pathogen interactions through signaling (Kuźniak & Gajewska, 2024). Additionally, Hager et al. (2025) identified metabolites directly involved in plant defense, including flavones and flavonoids, in spring wheat. These findings align with the results of this study, where defense-related metabolites and phytohormones were consistently implicated.

Two phytohormones differ significantly between genotypes ($p\text{-value}_{\text{FDR}} < 0.05$), one at peak infestation and one at post-infestation. In both cases, the hormone was jasmonoyl-isoleucine (JA-Ile), which was downregulated in the solid stem line at peak infestation ($\text{Log}_2\text{FC} = -1.59$, $p\text{-value}_{\text{FDR}} = 0.02$) and upregulated in the solid stem line post-infestation ($\text{Log}_2\text{FC} = 1.46$, $p\text{-value}_{\text{FDR}} = 0.04$). JA-Ile is a conjugate of jasmonic acid and a key component of the jasmonate signaling pathway, which regulates plant defense responses to herbivory (Schuman et al., 2018). The shift from downregulation at peak infestation to upregulation post-infestation suggests dynamic regulation of defense responses over time. Increased JA-Ile levels may activate defense

mechanisms early in infestation (P. Zhang et al., 2025), followed by downregulation as the larva moves toward the base of the stem and the plant reallocates resources in hollow stem wheat. Notably, jasmonic acid itself was not significantly different at any time point.

To further investigate pathway-level relationships, correlations were performed between annotated metabolites, unannotated metabolites, and phytohormones. All correlations between metabolites and JA-Ile were negative, suggesting they are not associated with shared pathways. In contrast, jasmonic acid (JA), abscisic acid (ABA), salicylic acid (SA), and indole-3-acetic acid (IAA) showed positive correlations with both annotated and unannotated metabolites. IAA also exhibited numerous negative correlations between annotated and unannotated metabolites. 6/17/2026_5:53:00 PM demonstrated that wheat may inactivate IAA in the stem in response to WSS infestation, which may explain the negative correlations observed in this study. While ABA is associated with stress response and IAA is an auxin, both have been shown to positively regulate JA when applied exogenously to the woody plants *Cleistocalyx operculatus* and *Syzygium jambos* in long-term water logging conditions (Cisse et al., 2022). Additionally, ABA signaling has been shown to affect JA biosynthesis in *Arabidopsis thaliana* in response to the oomycete pathogen *Pythium irregulare* (Adie et al., 2007), highlighting the interconnected relationship of phytohormones.

Collectively, these results suggest that the *SSt1* locus contributes to WSS resistance through mechanisms that extend beyond structural reinforcement of the stem. Because *SSt1* is associated with the transcription factor *TdDof*, it may impact the expression of other genes and impact greater regulatory networks, causing multiple downstream effects (Ferreira, 2025; Verta et al., 2024). In this pair of NILs, differences in protein expression, metabolite variability, and phytohormone

dynamics indicate that *SStI* may also influence physiological processes such as senescence, metabolic regulation, and defense signaling. The persistence of pith parenchyma cells in solid stem genotypes may provide a biochemical environment conducive to the accumulation of secondary metabolites associated with plant defense, while also contributing to increased metabolic variability. Additionally, the observed temporal regulation of JA-Ile suggests that resistance conferred by the *SStI* locus may involve dynamic modulation of defense responses rather than constitutive activation. By leveraging NILs, this study isolates the effects of stem solidity and provides new insight into the molecular and physiological mechanisms underlying *SStI*-associated resistance to WSS.

CONCLUSIONS

While the solid stem trait improves resistance to WSS, important questions remain regarding its underlying mechanisms and the stability of its expression across environments. To address this, proteomic, metabolomic, and phytohormone analyses were conducted using hollow stem and solid stem NILs, allowing for a controlled comparison of physiological differences associated with the *SStI* locus. The results provide evidence that resistance conferred by the solid stem trait extends beyond structural reinforcement of the stem and involves coordinated changes in metabolic pathways and defense signaling. Differences in protein expression and metabolite variability suggest that *SStI* influences processes related to senescence, pith degradation, and secondary metabolite accumulation. Pathways associated with phenylpropanoid and flavonoid biosynthesis, as well as dynamic regulation of the defense-related phytohormone JA-Ile, highlight the role of biochemical and signaling mechanisms in WSS resistance. Because the solid stem trait

114 remains a cornerstone of WSS resistance in wheat, advancing our understanding of its underlying
115 mechanisms will be critical for developing cultivars with more consistent and effective resistance.
116 These findings provide a foundation for future research aimed at improving the stability of *SStI*
117 expression and integrating both structural and biochemical resistance strategies to better meet
118 farmer needs.

119

CHAPTER VI – EFFECTS OF SSt1, SOLIDITY, AND ENVIRONMENT ON YIELD- COMPONENT TRAITS IN HARD WINTER WHEAT

INTRODUCTION

Wheat stem sawfly (WSS; *Cephus cinctus* Norton) is a native grass-feeding insect pest across multiple crops including winter and spring wheat (*Triticum aestivum* L.) (Ainslie, 1920; Beres et al., 2012). Wheat stem sawfly has been a pest in the Northern part of the United States and Canada for over a century, but due to changes in lifecycle and climate, WSS has recently emerged as a pest in Colorado (Peirce et al., 2022). Damage due to WSS causes an estimated loss of \$350 million USD annually in North America (Beres et al., 2011).

Most of the WSS lifecycle occurs within the stem, making it difficult to control through standard agronomic measures such as chemical control (Beres et al., 2011). In late spring, WSS oviposit their eggs inside the stem of the host plant (Ainslie, 1920). Once the eggs have hatched, the larvae feed on the parenchymal tissue inside the stem of the plant and cannibalize each other until one larva remains. The remaining larva continues to feed before moving to the base of the stem to create a hibernaculum to overwinter. Formation of the hibernaculum damages the stem by creating a break in the stem, also known as “cutting”. Once the wheat has senesced, damaged wheat is predisposed to lodging, which disrupts the agronomic practices of the field and reduces yield.

Since most of the WSS lifecycle occurs within the stem, host plant resistance is the most effective tool in limiting WSS damage. The most well characterized form of genetic resistance to WSS is the solid stem trait on the long arm of chromosome 3B (*Qss.msub-3BL*) (Cook et al., 2004a). Nilsen et al. (2020) found that the region on 3B is controlled by copy number variation of

a *TdDof* transcription factor at the *SStI* locus. Three copies of the transcription factor confer a semi-solid or solid stem, and one copy confers a hollow stem. In a hollow stem variety, the parenchyma cells of the pith break down as the plant approaches physiological maturity, creating a cavity in the center of the stem (Kebrom et al., 2012; Kirby, n.d.).

In cultivars with the solid stem allele, the parenchyma cells do not break down as the plant matures, leaving a smaller space in the center of the pith. Stem solidity improves resistance to WSS through mechanical interference (Beres et al., 2012; Holmes & Peterson, 1960, 1961), which hinders the ability of WSS larvae to move around the stem and feed (Evert et al., 2006; Kemp, 1934). Additionally, the increased moisture in the stem can increase egg and neonate mortality (Holmes & Peterson, 1964). These features of the solid stem trait mitigate WSS damage, thus limiting yield loss in infested fields (Beres et al., 2007).

The solid stem phenotype is influenced by environmental variables (Beres et al., 2012) such as temperature and light intensity (Holmes, 1984) or increased moisture (Platt & Farstad, 1946). It has also been hypothesized, and shown in some studies, that solid stem wheat cultivars yield less than hollow stem wheat cultivars due to the repartitioning of assimilate to the stems (Sherman et al., 2015; Weiss & Morrill, 1992). This aligns with source-sink and resistance trade-off theories for crop improvement which state that the change in one trait comes at the cost of others (Dwivedi et al., 2021). However, there are mixed reports on the effect of the solid stem phenotype on grain yield. While multiple studies were unable to identify genetic or phenotypic correlations between grain yield and stem solidity (Cook et al., 2004a, 2019b; Hayat et al., 1995; Sherman et al., 2015), McNeal & Berg, (1979) identified a negative correlation.

Understanding the impact of environment, genetic backgrounds, and the *SStI* locus on yield component traits is important for the development of resistant varieties. The identification of

171 genotypes which consistently express solidity, while suffering minimal agronomic penalties, is
172 necessary to increase the stability of WSS resistance while still maximizing production. To
173 determine the relationship between the solid stem trait and yield component characteristics in
174 Colorado State University winter wheat germplasm, eight sets of near isogenic lines (NILs) with
175 varying degrees of stem solidity were assessed for a variety of WSS and agronomic traits. Each
176 set of NILs contained lines that had either two copies of the solid or hollow stem alleles at the *SSII*
177 locus. The NIL groups were grown over the 2023 and 2024 field seasons in Colorado and Nebraska
178 to understand agronomic and yield related tradeoffs between hollow stem and solid stem lines
179 across environments.

MATERIALS AND METHODS

Germplasm

The population used in this study consisted of eight sets of NILs that comprised a total of 50 lines (Table 6.1). Each line within each set of NILs had either the solid stem allele (MT) or the hollow stem allele (WT) at the *SStI* locus. These NIL groups were generated from breeding germplasm in the Colorado State University Wheat Breeding and Genetics program in Fort Collins, Colorado using a marker-assisted selection approach in heterogeneous F_{3:5} lines (Figure 1). Briefly, single heads were selected from segregating F₃ bulk populations in summer 2019, planted as F_{3:4} rows in Fall 2019, and bulk harvested in summer 2020. The developed F_{3:5} lines were subsequently evaluated in yield trials during the 2020-2021 season. In the F_{3:5} generation, ~20 single heads were again selected from individual F_{3:5} lines under the hypothesis that some lines would be heterogenous at the *SStI* locus. Individuals from each line were genotyped with markers for the *SStI* locus (*usw275*) (Nilsen et al., 2020) and selected individuals were planted as F_{5:6} rows. F_{5:6} headrows were bulk harvested to produced F_{5:7} seed for 2023 yield trials and a small seed increase (F_{5:8}) was utilized to produce additional seed for the 2024 field season.

For both field seasons, trials were planted in six different locations across Colorado and Nebraska including the Agricultural Research Development and Education Center in Fort Collins, CO, The United States Department of Agriculture – Agricultural Research Service (USDA-ARS) Central Great Plains Research Station in Akron, CO, the Stumpf International Wheat Center in Grant, Nebraska, the High Plains Agricultural Laboratory in Sidney, NE, and the University of Nebraska Eastern Nebraska Research and Extension Center in Mead, NE. In 2023, 50 unique lines across all eight NIL groups and five check lines were planted in a partially replicated design

including 70 yield plots at each location. Yield plots ranged from 4.64 m² to 5.55 m² across sites. All plots were managed according to the needs of the respective breeding programs, including supplemental irrigation in Fort Collins. In total, six site years were able to be used as environments in the study and are identified as follows: 2024 Sidney (2024:Sidney), 2024 Mead (2024:Mead), 2024 Grant (2024:Grant), 2024 Akron (2024:Akron), and 2023 and 2024 Fort Collins (2023:Fort Collins; 2024:Fort Collins).

Table 6.1. NIL group, pedigree, and number of lines with the hollow stem (WT) and solid stem (MT) allele at the *SS1* locus.

NIL Group	Pedigree	WT Lines	MT Lines
NIL1	Warhorse/Breck//CO12D1028	3	2
NIL2	Fortify SF/CO13007	3	1
NIL3	Fortify SF/CO13007	6	5
NIL4	CO15SFD030/CO11D1397	4	6
NIL5	CO15SFD030/CO13D1843	4	3
NIL6	CO16SF029/CO16SF078	3	1
NIL7	CO16SF029/CO16SF078	1	4
NIL8	CO16SF029/CO16SF078	1	3

Sample Collection

Traits collected for this study included total biomass (kg ha⁻¹; BM), vegetative biomass (kg ha⁻¹; VBM), harvest index (HI), tillers m⁻²(Tillers), plant height (cm; HT), heading date (Julian days; HD), kernel number spike⁻¹ (KNS), grain weight spike⁻¹ (g; GWS), thousand kernel weight (g; TKW), test weight (kg hL⁻¹; TW), and grain yield (kg ha⁻¹; GY), and stem solidity (5-25 visual ordinal rating).

Two replicates of biomass samples were collected from each plot at physiological maturity. For each of the two samples, a ~15 cm section of one row of wheat was cut from a representative portion of the yield plot. The stems were cut as close to the base as possible and all plant material

including the lower canopy leaves were collected. These samples were weighed to collect BM before the heads in the sample were counted and removed to collect grain yield component traits.

The heads from each total biomass sample were threshed, and the remaining kernels were weighed and counted. The total kernel weight of each sample was subtracted from its associated BM measurement to calculate VBM. Additionally, the kernel number and weight for each replicate was divided by the number of heads for that replicate to determine KNS and GWS, respectively. Averages of both reps were calculated for VBM, KNS, and GWS for each plot. The average kernel weights and numbers for each plot were used to calculate a prorated TKW. Other grain yield component traits, TW and GY, were measured during combine harvest using the Zürn Harvesting HarvestMaster (74214 Schöntal-Westerhausen, Germany). Approximate number of tillers (referred hereto after as tillers) was calculated by dividing GY by GWS and HI was calculated by dividing GY by BM.

Plant height was measured by measuring from the soil surface to the average height of the wheat heads excluding awns in a section of the yield plot representative of the overall plant height. Heading date was determined by the Julian day in which roughly half of the heads in each plot had fully emerged from the flag leaf. Stem solidity was rated by collecting 20 stems randomly from each plot. A cross section was taken from each internode of the stem and rated on a visual ordinal scale of solidness from one to five where one was completely hollow and five was completely solid. An average is then taken across internodes to rate each stem from five to 25 where a rating of five indicates a hollow stem throughout the whole length of the stem and 25 is completely solid throughout the whole length of the stem. The rating for each plot is then taken by averaging the stem solidness across all 20 stems (DePauw et al., 2005).

Genotyping

Lines originating from Colorado State University were whole genome sequenced via genotyping-by-sequencing (GBS) (Poland et al., 2012). Ten seeds were planted for each line and a roughly one cm tissue sample was collected from each plant for deoxyribonucleic acid (DNA) extraction. DNA was extracted and quantified from the samples using ThermoFisher's MagMax plant DNA and PicoGreen kits (ThermoFisher Scientific, Waltham, Massachusetts, USA), respectively. Extracted DNA was normalized to a concentration of 20 ng μL^{-1} . Sequencing libraries were prepped following protocol from Poland et al. (2012) and sequenced on a NovaSeq 6000 sequencer (Illumina, San Diego, California, USA) at 384-plex density per lane. Reads were aligned to the RefSeq Chinese Spring wheat reference sequence v2.1 (Appels et al., 2018) using the burrow-wheel aligner (H. Li & Durbin, 2009). Aligned reads were processed using the TASSEL 2.0 pipeline (Glaubitz et al., 2014).

All single nucleotide polymorphisms (SNPs) were filtered using the following parameters: SNPs with greater than 15% missing data were removed; SNPs with greater than 30% heterozygosity were removed; SNPs with a read depth of zero or greater than 100 were removed; SNPs with a site allele frequency less than 5% or greater than 95% were removed, and unaligned or indel SNPs were removed; Only biallelic SNPs were retained. After filtering, 53,066 high-quality SNPs remained. Missing data was imputed using Beagle V 5.4 (Browning et al., 2018). A synthetic wheat biparental cross between 'W7984' and 'Opata' was used to derive the recombination distance-based map for imputation (Gutierrez-Gonzalez et al., 2019).

Data Analysis

Statistical analysis was performed using R version 4.5.1 (R Core Team, 2025). A principal components analysis (PCA) was performed on the genetic marker matrix using the R stats package

(R Core Team, 2025) to determine the population structure across NIL groups and the R package ‘ape’ version 5.8-1 (Paradis & Schliep, 2019) was used for building a phylogenetic tree. For phenotypic analysis, a Pearson correlation was performed between all traits across environments and NIL groups. Additional models were used to assess the effect of the *SStI* allele and solidity on the same yield component traits.

Multi-environment best linear unbiased estimates (BLUEs) were obtained for all 12 traits using ASReml-R version 4.2.0.355 (Butler et al., 2023). The following model was used to calculate BLUEs of the genotype effect across environments:

$$y_{ij} = \mu + \tau_i + B_j + \varepsilon_{ij} \quad (Eq. 1)$$

where y_{ij} is the observed phenotype; μ is the overall mean; τ_i is the fixed effect of the i^{th} genotype; B_j is the random environment effect; and ε_{ij} is the residual error term, where $\varepsilon_{ij} \sim N(0, I\sigma^2_\varepsilon)$, where I is an identity matrix and σ^2_ε is the residual error variance. The model employed a row-by-column interaction with autocorrelated structures for both dimensions.

For analysis, SNPs included on chromosome 3B in the genotype file were pruned to not include any markers associated with the *SStI* locus to avoid overfitting of the model. To remove any associated SNPs, the peak *SStI* marker was identified using results from Hammers et al. (2026, manuscript submitted for publication), and linkage disequilibrium was calculated between the peak *SStI* marker and all other markers on chromosome 3B using the R package *gaston* (Perdry & Dandine-Roulland, 2015). Markers with a linkage disequilibrium r^2 greater than 0.05 were removed from the genotype file before generating a VanRaden additive genomic relationship matrix (G) (VanRaden, 2008).

All models were performed using the ASReml-R package version 4.2.0.355 (Butler et al., 2023). For the across environment model, raw phenotypic data was used as the input for the following model:

$$y_{ijkl} = \mu + N_i + S_j + E_k + N_i * S_j * E_k + g_l + \varepsilon_{ijkl} \quad (Eq. 2)$$

Where y_{ijkl} is the phenotype of the l^{th} genotype in the i^{th} NIL group with a j^{th} class of *SStI* allele in the k^{th} environment, μ is the overall mean, N_i and S_i are the NIL group and *SStI* allele of the l^{th} genotype, E_k is the environment, $N_i * S_j * E_k$ is the interaction between the NIL group, *SStI* allele, and environment which expands to $N_i * S_j + N_i * E_k + S_j * E_k + N_i * S_j * E_k$, g_l is the random effect of the l^{th} genotype modeled using $\mathbf{G}$ with assumed normal distribution $l \sim N(0, \mathbf{G}\sigma_l^2)$, and ε_{ijkl} is the residual error term with the assumed normal distribution $\varepsilon_{ijkl} \sim N(0, \sigma_u^2(R_{row,k} \otimes R_{column,k}))$ where $R_{row,k}$ and $R_{column,k}$ are the autocorrelated (AR1) correlation matrices of the row and column for environment k . Predictions and 95% confidence intervals were calculated using asremlPlus version 4.4.58 (Brien, 2015) for each model variable and model interaction were performed using the ASReml model to estimate mean performance across environments.

RESULTS

Population and Data Assessment

General assessments of traits across environments

Means and standard deviations of traits for each NIL group across environments were calculated (Figure 6.1). Across all environments and NIL groups, there were no clear patterns between WT and MT lines for any traits except for HT and solidity, where the WT allele had

314 higher mean height across all NIL groups and the MT allele had higher mean solidity. Across all
315 NIL groups and environments, solidity was only significantly correlated with HT (Figure 6.2).

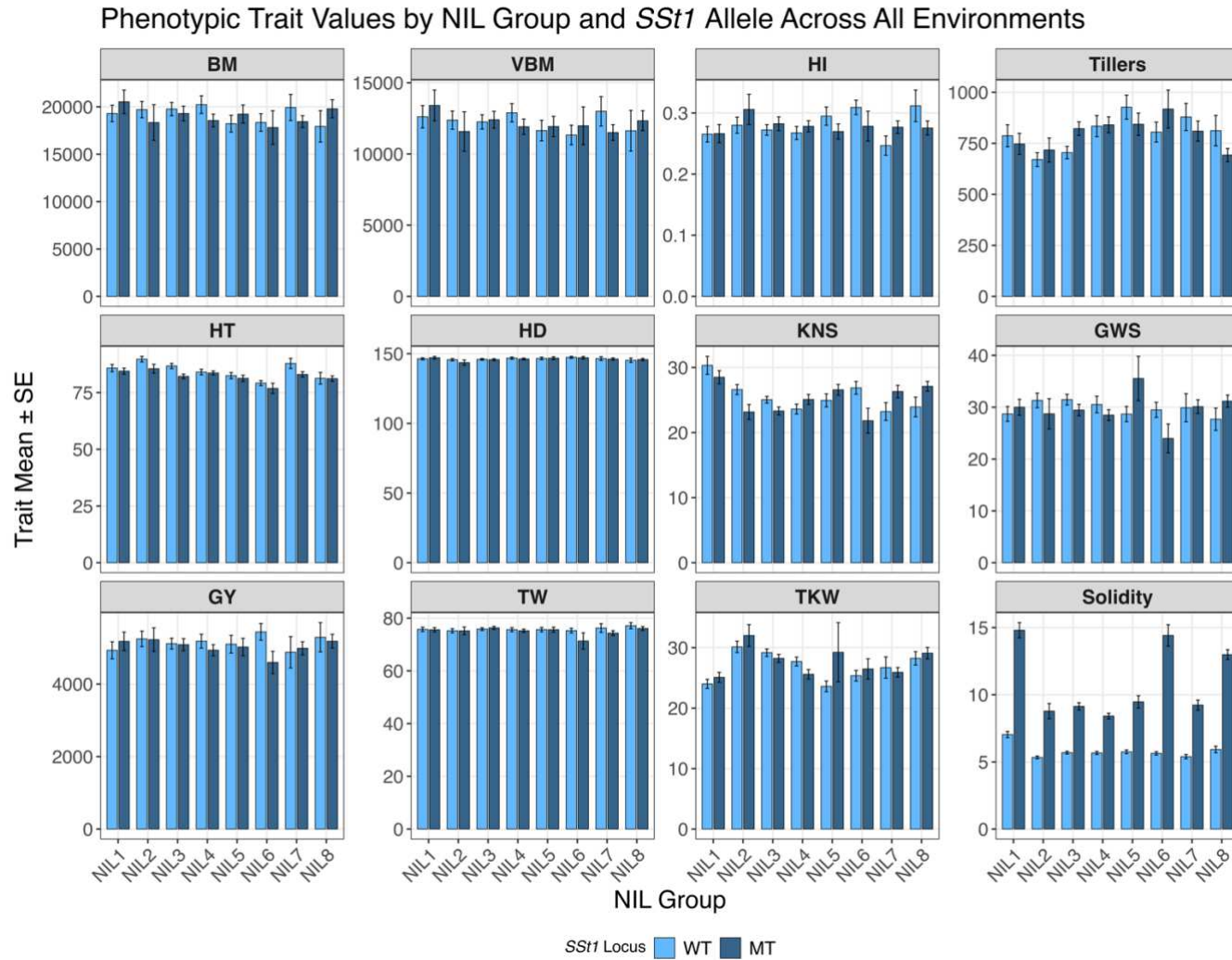


Figure 6.1. Trait means and standard errors for all eight NIL groups across all environments. The light blue boxes represent the hollow stem (WT) allele at the *SS1* locus, and the dark blue boxes indicate the solid stem (MT) locus at the *SS1* locus. The traits are as follows: total biomass (BM), vegetative biomass (VBM), harvest index (HI), tiller number (Tillers), plant height (HT), heading date (HD), kernel number per spike (KNS), grain weight per spike (GWS), grain yield (GY), test weight (TW), thousand kernel weight (TKW), and solidity.

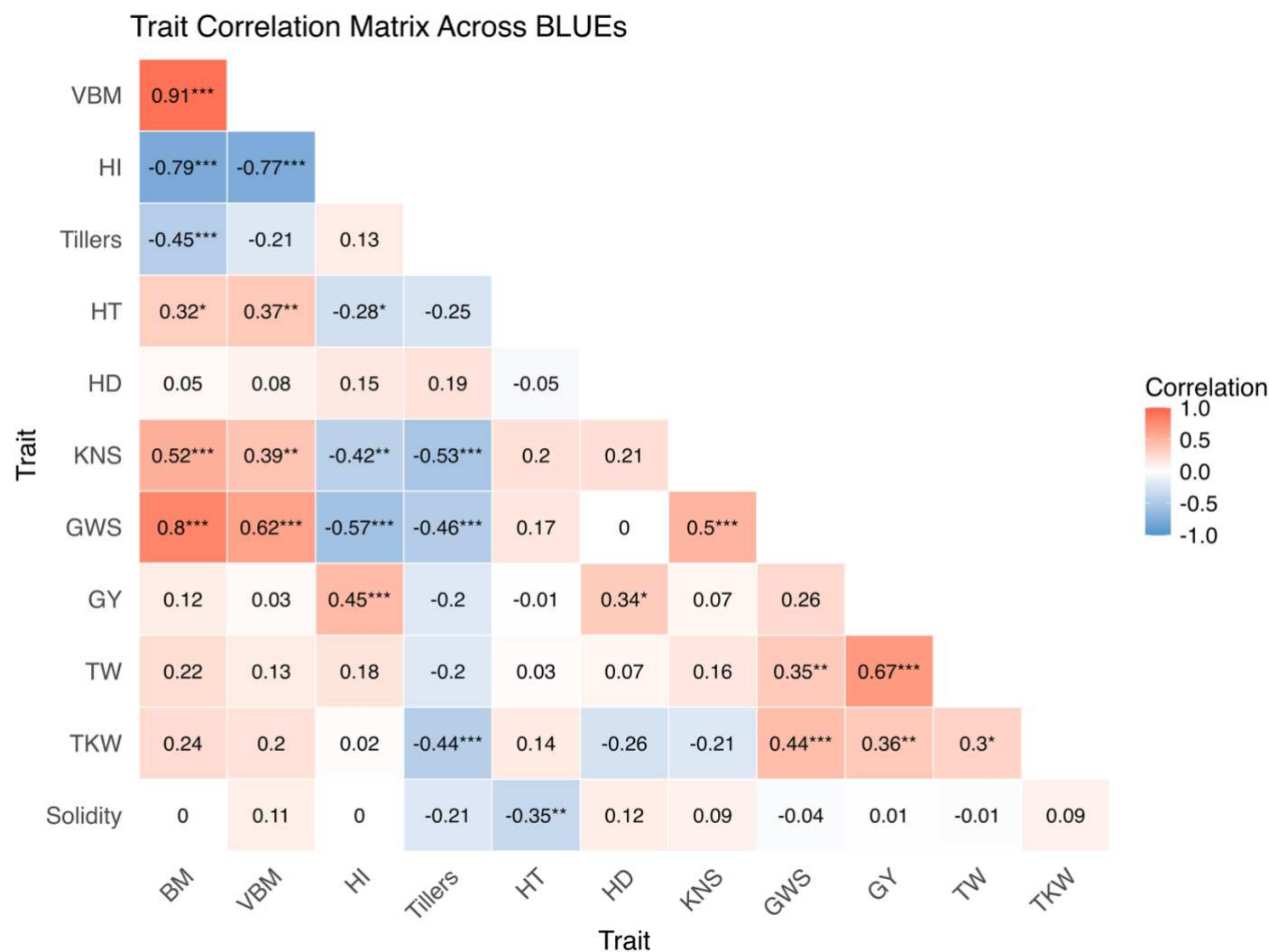


Figure 6.2. Phenotypic correlations between all traits for across NIL groups in all environments. Traits include total biomass (BM), vegetative biomass (VBM), harvest index (HI), tiller number (Tillers), plant height (HT), heading date (HD), kernel number per spike (KNS), grain weight per spike (GWS), grain yield (GY), test weight (TW), thousand kernel weight (TKW), and solidity. P significance is indicated next to each correlation where * indicates a $p \leq 0.05$, ** indicates a $p \leq 0.01$, and *** indicates a $p \leq 0.001$. No asterisk indicates a non-significant p for the correlation.

Phylogenetic Analysis

There were 812 SNPs in linkage disequilibrium with the *SSII* locus that were removed from chromosome 3B, leaving the final genotype file with 52,254 total SNPs (Supplemental Figure 6.1). A principal component analysis of all genetic markers showed that PC1, PC2, and PC3 explained 22.7%, 14.3%, and 11.8% of the variance, respectively (Supplemental Figure 6.2). A phylogenetic tree was generated from the pruned SNP matrix (Figure 6.3). The NIL groups clustered into eight distinct groups except for outliers from NIL3, NIL7, and NIL8. However, the outliers identified in the phylogenetic tree clustered in NIL groups that have the same pedigree (Table 6.1, Supplemental Table 6.1).

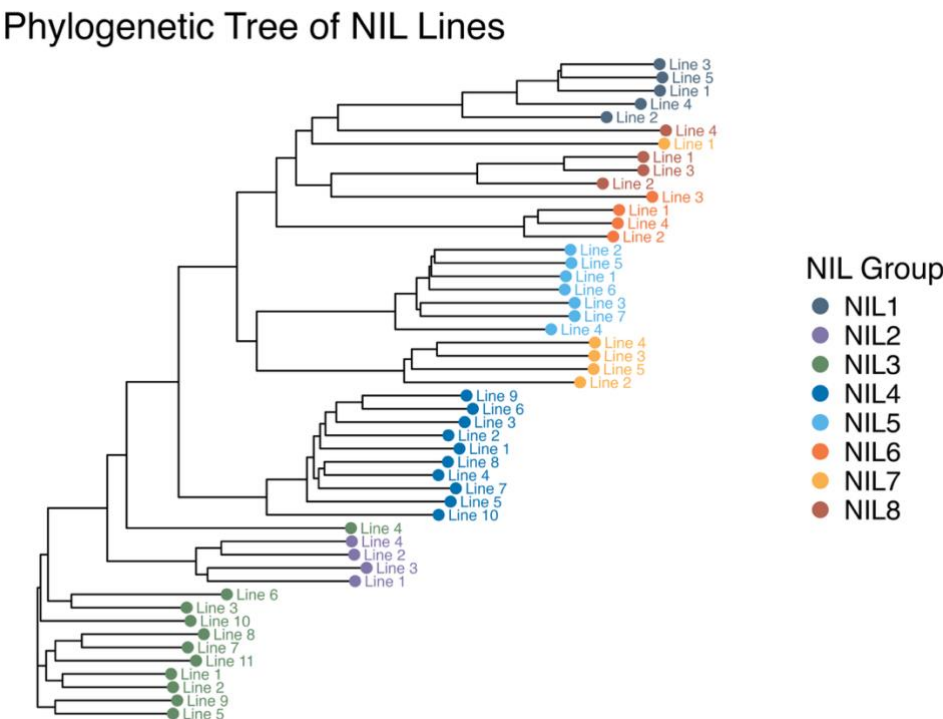
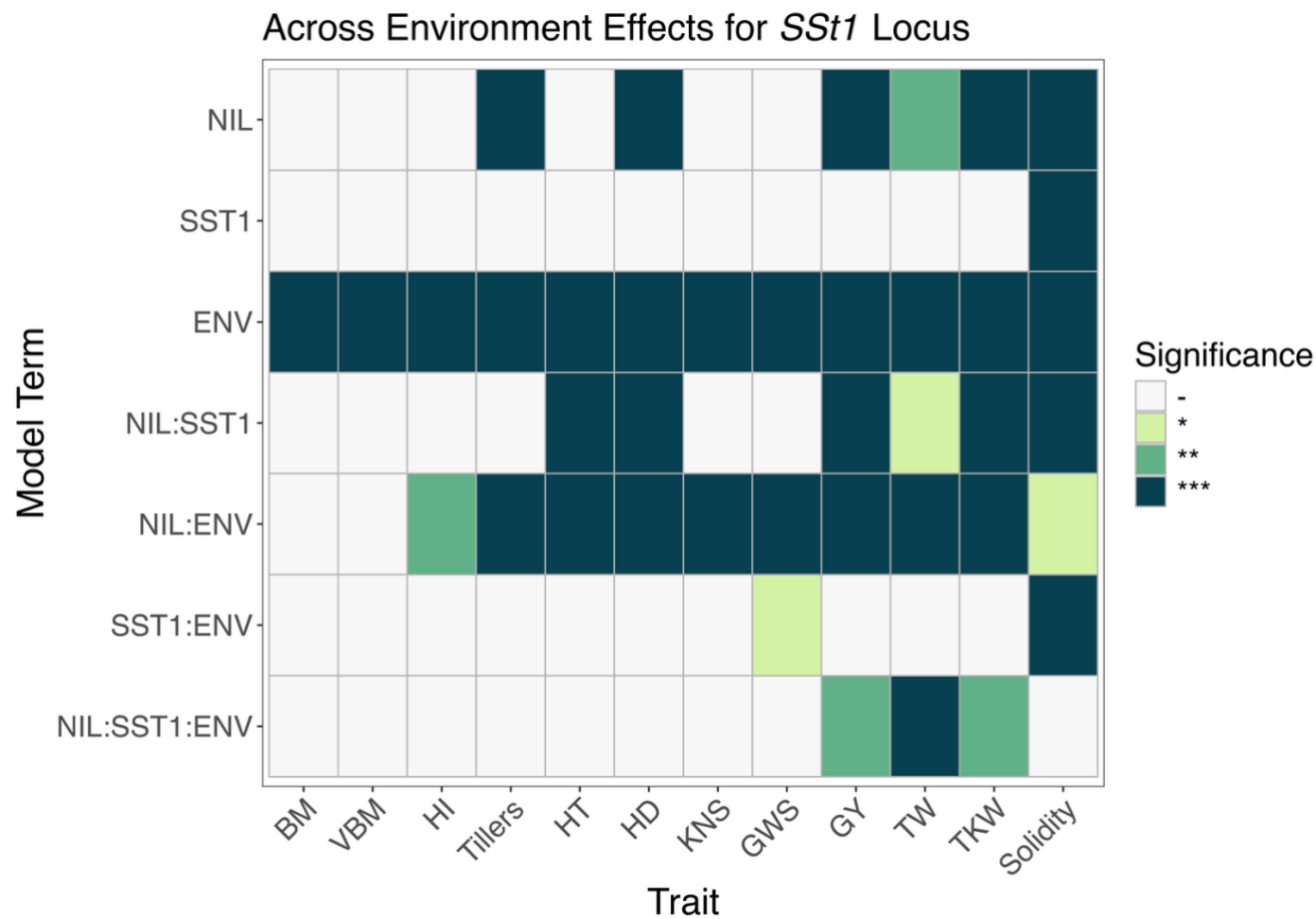


Figure 6.3. Phylogenetic tree showing the relationships between the lines in each NIL group. The color of each leaf indicates which NIL group a leaf belongs to. The distance of a single leaf from the trunk of the tree indicates the relative genetic distance of the leaf from other individuals in the tree.

Effect of Allelic Variation at *SStI* Across Environment

Effect significance across traits

In the across environment analysis, NIL group had a significant effect on TW ($p \leq 0.01$), and Tillers, HD, GY, TKW, and solidity (all $p \leq 0.001$) (Figure 6.4). The *SStI* allele had a significant effect on solidity ($p \leq 0.001$). The interaction of NIL Group-by-*SStI* allele had a significant effect for TW ($p \leq 0.05$), HT, HD, GY, TKW and solidity ($p \leq 0.001$). The environment main effect was significant ($p \leq 0.001$) for all traits assessed. The environment-by-NIL group interaction was significant for solidity ($p \leq 0.05$), HI ($p \leq 0.01$), tillers, HT, HD, KNS, GWS, GY, TW, and TKW ($p \leq 0.001$). The environment-by-*SStI* interaction was significant for GWS ($p \leq 0.05$) and solidity ($p \leq 0.001$). The NIL group-by-*SStI*-by-environment interaction (three-way interaction) was significant for GY and TKW ($p \leq 0.01$) and TW ($p \leq 0.001$) (Supplemental Table 6.2).



357
358 **Figure 6.4.** Across environment Wald table output for significant effects of NIL group, *SSt1* locus, environment, and the interactions of
359 these model components on total biomass (BM), vegetative biomass (VBM), harvest index (HI), tiller number (Tillers), plant height
360 (HT), heading date (HD), kernel number spike⁻¹ (KNS), grain weight spike⁻¹ (GWS), grain yield (GY), test weight (TW), thousand
361 kernel weight (TKW), and solidity. P significance is indicated next to each correlation where light green indicates a $p \leq 0.05$, dark green
362 indicates a $p \leq 0.01$, and dark blue indicates a $p \leq 0.001$. White indicates a non-significant p for the effect.

Biomass Traits

For BM and VBM, environment was the only significant effect, indicating that differences observed were based upon environment rather than NIL group or *SStI* allele. For HI and tillers, the environment by NIL group interaction was significant, but there was no significant effect of the *SStI* locus across NIL groups (Supplementary Figure 6.3).

Physiological Traits

For HD, for the NIL group-by-*SStI* interaction, there were no clear differences between the MT and WT alleles except for NIL2, where the WT allele had an estimated HD 1.5 Julian days after the MT allele ($WT = 145.59 \pm 0.78$; $MT = 144.12 \pm 1.17$)(Figure 6.5A).

For the HT NIL group-by-*SStI* interaction, the mean estimate of the WT allele was higher in all NIL groups except for NIL8. However, there were no major differences between the MT and WT allele for any NIL groups, indicating the *SStI* locus may not have a notable effect on HT (Figure 6.5B).

Estimates for HD and HT for NIL Group-by-*SSt1* Locus

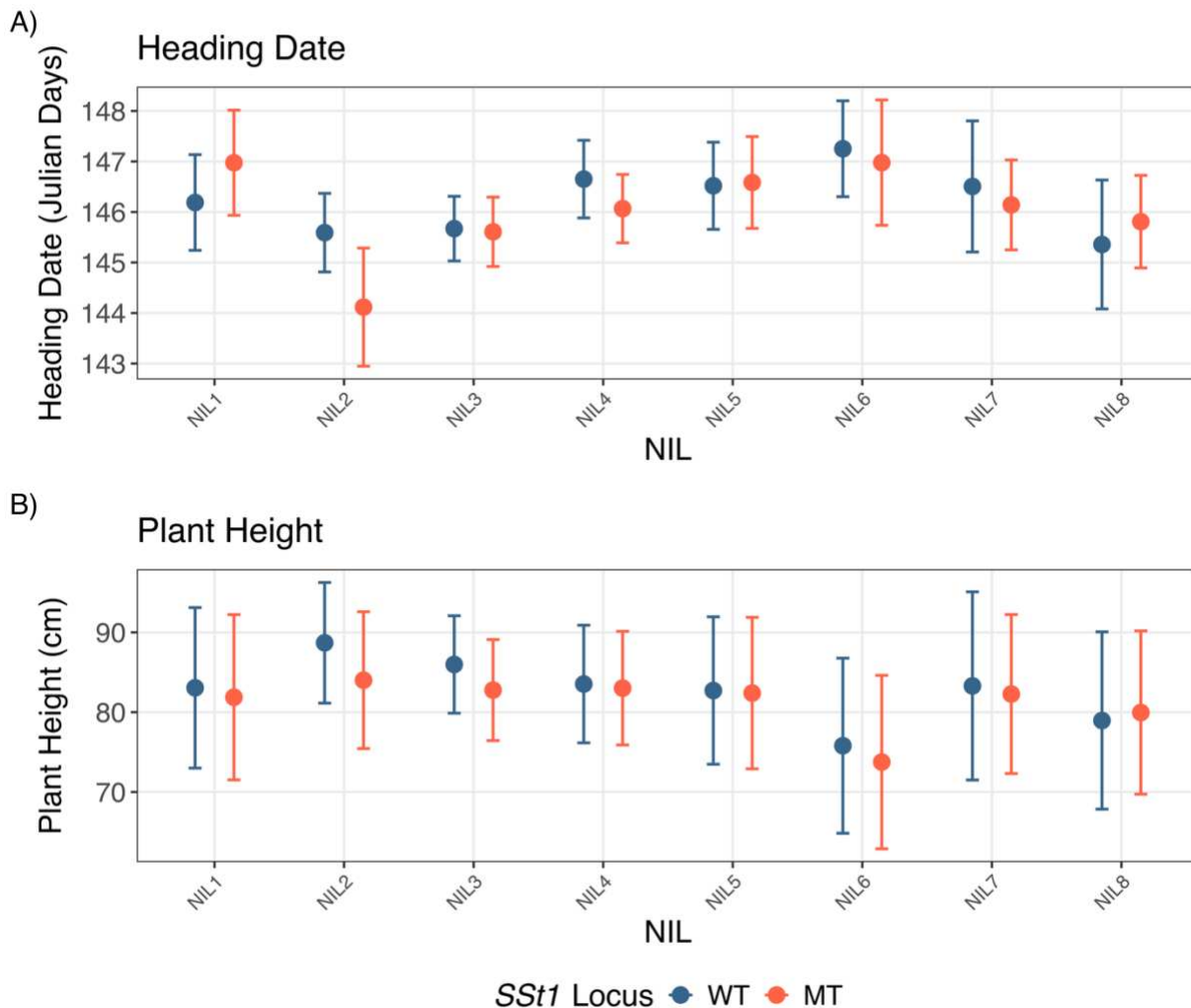


Figure 6.5. Estimates and confidence intervals for NIL group-by-*SSt1* interactions for (A) heading date (Julian days; HD) and (B) plant height (cm; HT). Blue is the hollow stem (WT) allele and red is the solid stem (MT) allele.

Components of Yield

For KNS, there was no significant independent *SStI* effect. For the GWS *SStI*-by-environment interaction, there was a clear separation in environments that formed two groups (Figure 6.6). The WT allele had a lower estimate than the MT allele in two of these environments, one from each group. However, there were no consistent effects across all environments for the *SStI* locus.

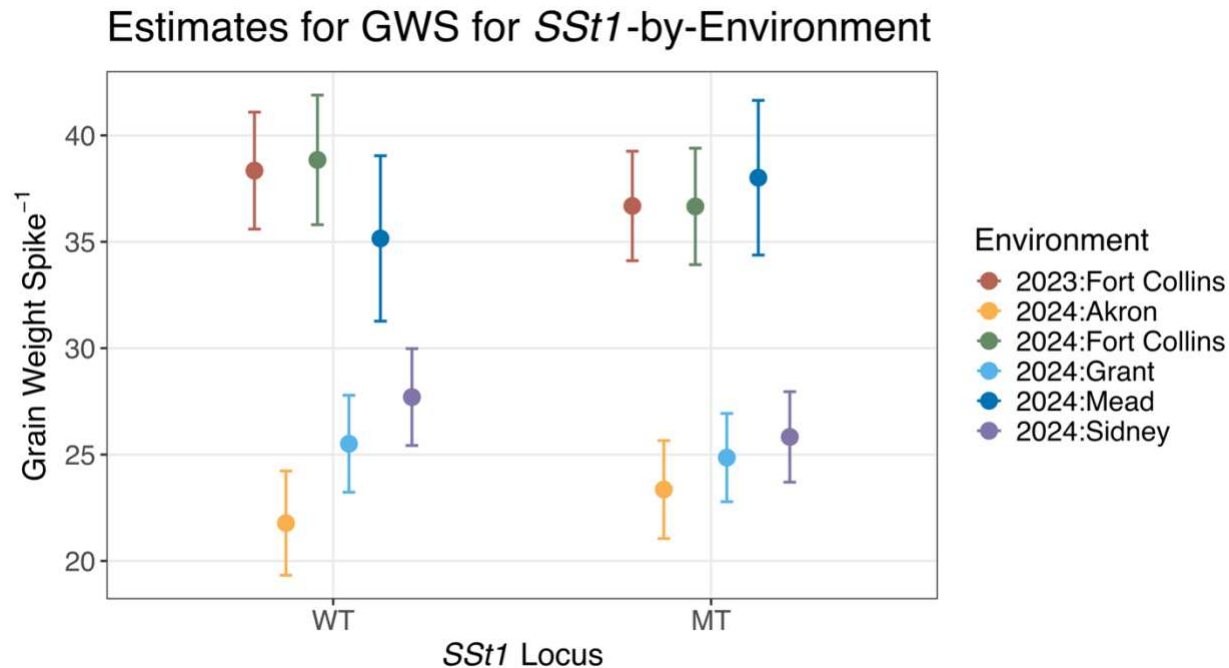


Figure 6.6. Estimates and confidence intervals for *SStI*-by-environment interactions for grain weight spike⁻¹ (g; GWS) for the hollow stem (WT) and solid stem (MT) alleles at all six environments.

For TKW, in the NIL group-by-*SStI* interaction, there was a significant difference in the MT and WT allele for NIL5 where the MT allele had a higher predicted value (WT = 24.0 ± 1.7 g; MT = 30.4 ± 2.1 g). However, the group estimates associated with the NIL group-by-*SStI*-by-environment interaction only showed this difference for NIL5 three environments (2024:Mead, 2024:Fort Collins, 2024:Sidney). Additionally, the WT allele had significantly higher TKW

estimates in NIL4 (2024:Fort Collins, 2024:Sidney) and NIL7 (2023 and 2024 Fort Collins), each in only two environments. This indicates that *SStI* does have an effect on TKW, but it is not consistent.

Yield and Test Weight

For GY, in the NIL group-by-*SStI* interaction, there was a significant difference ($p=0.05$) in group mean estimates between the MT and WT allele for NIL6 where the MT allele had a lower predicted GY (WT = $5606.8 \pm 438.5 \text{ kg ha}^{-1}$; MT = $4859.7 \pm 512.2 \text{ kg ha}^{-1}$)(Figure 6.7A). This finding was supported by the group estimates associated with the NIL group-by-*SStI*-by-environment interaction in five of the six environments (all except 2024:Mead), indicating a decrease in GY for the solid stem trait in the genetic background of NIL6. The estimates from the three-way interaction also showed that the MT allele for NIL1 had higher GY predictions in two environments with no clear difference in others, indicating a potential advantage to having the MT allele in the NIL1 genetic background; this trend, while insignificant in the NIL group-by-*SStI* interaction, is mirrored in the estimates of NIL 1 NIL group-by-*SstI* interaction.

For TW, in the NIL group-by-*SStI* interaction, there was separation in estimates between the MT and WT allele for NIL6 (MT = $73.1 \pm 1.2 \text{ kg haL}^{-1}$; WT = $75.8 \pm 1.0 \text{ kg haL}^{-1}$) and NIL7 (MT = $74.9 \pm 0.9 \text{ kg haL}^{-1}$; WT = $76.9 \pm 1.2 \text{ kg haL}^{-1}$) where the MT allele had a lower predicted value for both NIL groups(Figure 6.7B). However, in the group estimates associated with the NIL group-by-*SStI*-by-environment interaction, this pattern was only visible for NIL6 in three environments. Additionally, for TW in NIL6, the MT allele was clearly higher in one environment. For NIL7, the WT allele still had a higher TW than the MT allele in 4 environments. Finally, for NIL8, there was one environment where the TW for the WT allele was higher and one environment

where the MT allele was higher. These results indicate that genetic background and environment affect the impact of *SStI* on TW.

Solidity

For Solidity, there was clear separation between the MT and WT allele for *SStI* (WT = 5.8 ± 0.4 ; MT = 11.2 ± 0.3). However, in the group estimates associated with the NIL group-by-*SStI* interaction, the MT allele for NIL1, NIL6, and NIL8 were clearly more solid than the MT lines from all other NILs (Figure 6.7C). This illustrates the variability in solidity across lines with different genetic backgrounds.

Estimates for GY, TW, and Solidity for NIL Group-by-*SSt1* Locus

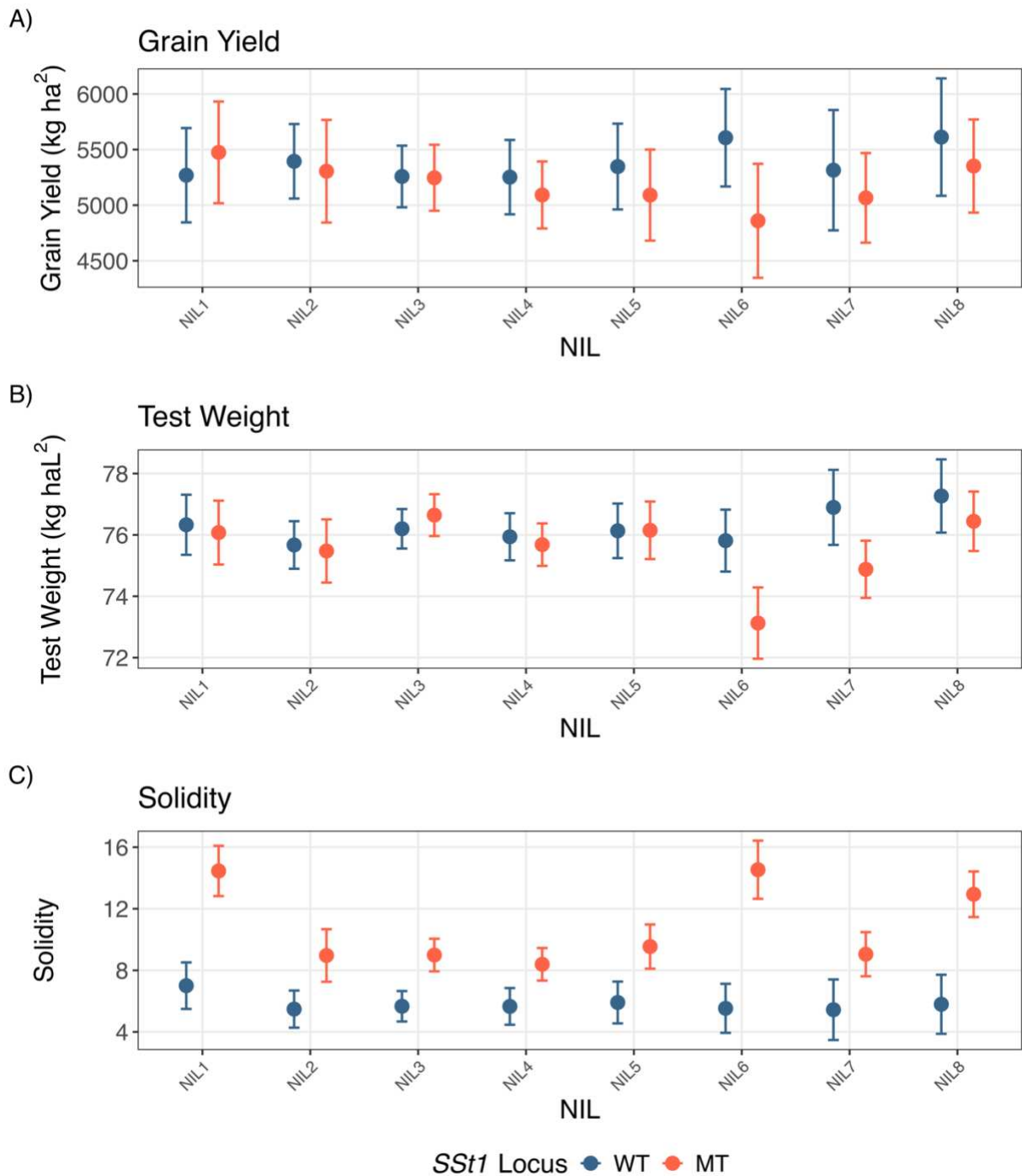


Figure 7. Estimates and confidence intervals for NIL group-by-*SSt1* interactions for (A) yield (kg ha⁻¹; GY), and (B) test weight (kg haL⁻¹; TW), and (C) solidity. Blue is the hollow stem (WT) allele and red is the solid stem (MT) allele.

DISCUSSION

One major difficulty in plant breeding is improving traits that are linked in repulsion (Yan & Wallace, 1995). This challenge is common for traits such as grain yield and quality and is often driven by source–sink dynamics and the allocation of photosynthate within the plant (Smith et al., 2018). There has been mixed evidence regarding trade-offs between disease or insect resistance and yield-related traits in crops (Brown, 2002; Brown & Rant, 2013). Understanding these trade-offs, and identifying genetic backgrounds that minimize them, is essential for improving cultivars. It has been hypothesized that the solid stem trait conferring WSS resistance results in yield penalties. Early work showed that the first solid stem cultivar, Rescue (CI 12435), yielded less than hollow stem cultivars in the absence of WSS pressure (Platt & Farstad, 1946), though more recent studies indicate that solid stem cultivars have improved substantially (Sherman et al., 2015).

Previous research examining correlations between solidity and yield component traits found mixed results. Hayat et al. (1995) found no correlation between solidity and grain yield in spring wheat crosses, but (McNeal & Berg, 1979) identified a positive correlation in one environment and no correlation in another. Saint Pierre et al. (2010) also found a positive correlation between the grain yield and solidity. In contrast, Cook et al. (2004, 2019) studied relationships of yield component traits with different *SStI* solid stem alleles in winter wheat. The studies found no correlation between solidity and yield and no differences between alleles and grain yield, thousand kernel weight, or number of kernels per head, respectively. Similarly, Sherman et al. (2015) utilized NILs that represented multiple different genetic backgrounds across multiple environments. An analysis of variance (ANOVA) determined that the solid stem NILs

had lower grain yield compared to the hollow stem NILs in one of the six genetic backgrounds and that there was a single environment where the solid stem lines had lower grain yield.

Across all environments and NIL groups in this study (Figure 6.3), relationships between solidity and yield components were not uniformly negative. Additionally, only plant height was significantly correlated with solidity ($r = -0.35$, $p\text{-value} < 0.01$). Graham & Frels (2025) also identified a negative correlation between plant height and solidity. However, Lanning et al. (2012) found mixed result for dwarfing genes in wheat where germplasm with the *Rht-B1b* and *Rht-D1b* alleles had increased solidity compared to the wild type. In the same study, germplasm with *Rht8* had decreased solidity compared to the wild type. Mixed linear models indicated that the NIL group-by-*SS1* locus interaction significantly affected plant height, but there was no clear pattern in the across-environment plant height estimates. These results suggested that environmental effects largely drive height variation. Because WSS preferentially oviposits in taller plants (Buteler et al., 2009), reduced height in solid stem lines may contribute an additional mechanism of resistance depending on environmental conditions.

As expected, solidity was significantly affected by the *SS1* locus, NIL group, and environment, but not by the three-way interaction of these variables. This result indicates that differences among NILs are primarily driven by the different genetic backgrounds and their response to the *SS1* allele. This was supported by estimated values, where NIL1, NIL6, and NIL8 consistently exhibited higher solidity across environments with mean NIL group by *SS1* solidity predictions of 14.5, 14.5 and 12.9, respectively. These NIL groups are therefore of particular interest for identifying genetic backgrounds that maintain strong expression of the solid stem trait.

Grain yield was influenced by all model terms except *SStI*-by-environment interactions and *SStI* independently. In NIL6 and NIL8, the hollow stem allele had higher NIL group-by-*SStI* interaction estimates for grain yield than the solid stem allele by 747 kg ha⁻¹ and 260 kg ha⁻¹, respectively, supporting the resistance–yield trade-off hypothesis (Brown, 2002). However, in NIL1, the solid stem allele resulted in a 205 kg ha⁻¹ increase in grain yield relative to the hollow stem allele. This deviation from the expected trade-off suggests that improved solid stem germplasm can mitigate or even overcome yield penalties, highlighting the importance of polygenic genetic background in determining trait outcomes.

Across environments, the *SStI* locus independently only affected solidity, while all other traits were associated through interactions with environment or genetic background. This indicates that the agronomic effects of the solid stem trait are highly context dependent. Biomass traits were only found to be significant in environment or NIL group-by-environment interactions. This result may be due to the fact that biomass samples were collected after physiological maturity, after the solid stem had broken down during senescence (Peirce et al., 2022; Varella et al., 2016).

One limitation of this study is the lack of partial replication of each line within environments, which resulted in wide confidence intervals and minimal detection of significant difference within and among NIL groups. While the results of this study indicate complex interactions between *SStI*, NIL groups, and environment, additional work using key lines of interest in larger studies is required to understand the full effect of the *SStI* allele in the different environments and NIL groups on yield component traits.

Not all NIL groups expressed moderate to high levels of solidity across environments (Figure 6.7). NIL1, NIL6, and NIL8 consistently exhibited high solidity, making them valuable

targets for WSS resistance. Among these, NIL1 showed the most favorable combination of traits, with negligible reductions in test weight in solid stem lines, alongside slight increases in grain yield. Although these increases are modest, the maintenance of yield in consistently solid lines represents a key breeding advantage. In contrast, NIL6 and NIL8 exhibited stronger negative associations between the solid stem allele and grain yield and test weight, suggesting that their genetic backgrounds may be less desirable for breeding applications.

CONCLUSIONS

Overall, the relationships among the solid stem trait, yield component traits, genetic background (NIL group), and environment were highly context dependent. While the *SS1* locus influenced several agronomic traits, these effects were largely mediated by genetic background and environment, with few consistent patterns across all conditions. This was particularly evident for grain yield, reflecting its complex genetic architecture. Among the NIL groups, NIL1 consistently expressed high levels of stem solidity across environments without major penalties in key yield components. This highlights the importance of identifying genetic backgrounds that support stable expression of the solid stem trait while minimizing agronomic trade-offs. Such backgrounds provide valuable opportunities for improving WSS resistance without compromising overall performance.

The solid stem trait remains a critical component of maintaining wheat productivity under WSS pressure. Integrating an improved understanding of genetic background, environmental

interactions, and trait expression will be essential for developing cultivars that combine strong WSS resistance with stable and competitive agronomic performance. Ultimately, improving the consistency of trait expression across environments will enhance yield stability for farmers and support resilient wheat production systems in the High Plains.

CHAPTER VII – CONCLUSIONS

Throughout this dissertation, a wide range of approaches were used to advance our understanding of wheat stem sawfly resistance in hard winter wheat. The overarching objective was to identify strategies that could improve selection and cultivar development within the Colorado State University wheat breeding program. This section synthesizes the major findings from Chapters II–VI and outlines opportunities for incorporating these results into future breeding efforts and research initiatives.

The findings from Chapters II, III, and VI provide important guidance for future breeding objectives. Collectively, these results support the incorporation of small-effect resistance loci into selection decisions and emphasize the importance of identifying genetic backgrounds that consistently express stem solidity while maintaining strong agronomic performance across environments. Although the *SStI* locus remains the primary determinant of stem solidity, additional loci associated with cutting score and stem solidity contributed measurable improvements in resistance. Incorporating these loci into breeding efforts will require the development and validation of molecular markers, such as KASP assays, within program germplasm. Combined with selection for favorable *SStI* alleles, these markers could accelerate the development and release of cultivars with improved wheat stem sawfly resistance.

This work also demonstrated that the effects of *SStI* on stem solidity and yield-related traits vary among genetic backgrounds. Some backgrounds maintained high levels of solidity while exhibiting fewer of the yield penalties traditionally associated with solid-stemmed wheat. Identifying and utilizing these favorable backgrounds represents an important opportunity to improve resistance without compromising yield potential. To further characterize these

interactions, more highly replicated studies using near-isogenic line populations, similar to those described in Chapter VI, should be conducted.

Results from the genomic selection analyses provided support for the breeding strategies currently employed within the program. Although more complex models, such as random forest, showed potential advantages for predicting cutting score, GBLUP remained the most effective model for predicting stem solidity. Additional machine learning approaches may warrant investigation, particularly for traits related to wheat stem sawfly damage. However, the overall findings indicate that the genomic prediction methods currently implemented by the program provide effective support for selection decisions, especially when combined with marker-assisted selection.

The multi-omics analyses presented in Chapters III and IV yielded valuable insights into the molecular basis of stem solidity, although they did not provide a clear explanation for the increased resistance observed in Clearfield® cultivars. Despite these limitations, understanding the mechanisms underlying Clearfield®-associated resistance remains an important objective. Several approaches could be used to identify the genetic factors contributing to this resistance, including the development of biparental populations (e.g., Byrd × Byrd CL Plus) for quantitative trait locus (QTL) mapping or the assembly of a diverse panel of Clearfield® germplasm for genome-wide association studies. In both cases, cutting score would serve as the primary resistance phenotype. These efforts could identify genomic regions associated with enhanced resistance and provide additional tools for breeding improved cultivars.

Beyond the specific objectives addressed in this dissertation, several questions remain that could influence future breeding decisions. For example, how do varying levels of stem solidity affect yield component traits, and how does the distribution of solidity throughout the stem

influence wheat stem sawfly resistance? These questions could be addressed using existing datasets or through the development of new experiments, including the highly replicated NIL populations proposed above. Answering these questions would further refine our understanding of the relationships among stem anatomy, resistance, and agronomic performance.

Wheat stem sawfly remains one of the most important insect pests affecting wheat production across the High Plains. Although numerous research opportunities remain, the findings of this dissertation suggest three priorities for future breeding efforts: incorporating small-effect resistance loci into selection programs, identifying genetic backgrounds that combine strong resistance with favorable agronomic performance, and elucidating the genetic basis of resistance associated with Clearfield® cultivars. Together, these efforts will strengthen the development of cultivars with durable and effective resistance to wheat stem sawfly.

REFERENCES

- Adie, B. A. T., Pérez-Pérez, J., Pérez-Pérez, M. M., Godoy, M., Sánchez-Serrano, J.-J., Schmelz, E. A., & Solano, R. (2007). ABA is an essential signal for plant resistance to pathogens affecting JA biosynthesis and the activation of defenses in Arabidopsis. *The Plant Cell*, 19(5), 1665–1681. <https://doi.org/10.1105/tpc.106.048041>
- Ainslie, C. N. (1920). *The Western Grass-Stem Sawfly*. United States Department of Agriculture Bulletin.
- Al-Kaisi, M., Hanna, M., & Tidman, M. (2002). Does dry weather limit tillage options? In *ICM News Archive*. Iowa State University Extension and Outreach.
<https://crops.extension.iastate.edu/encyclopedia/does-dry-weather-limit-tillage-options>
- Al-Kaisi, M., Hanna, M., & Tidman, M. (2004). Frequent tillage and its impact on soil quality. In *ICM News Archive* (Vol. 492). Iowa State University Extension and Outreach.
<https://crops.extension.iastate.edu/encyclopedia/frequent-tillage-and-its-impact-soil-quality#:~:text=However%2C%20tillage%20has%20all%20along,the%20force%20of%20pounding%20raindrops.>
- Allan, R. E. (1987). Wheat. In *Principles of Cultivar Development* (Vol. 2, pp. 699–748). Macmillan Publishing Company.
- Anjali, Kumar, S., Korra, T., Thakur, R., Arutselvan, R., Kashyap, A. S., Nehela, Y., Chaplygin, V., Minkina, T., & Keswani, C. (2023). Role of plant secondary metabolites in defence and transcriptional regulation in response to biotic stress. *Plant Stress*, 8, 100154.
<https://doi.org/10.1016/j.stress.2023.100154>

- Appels, R., Eversole, K., Stein, N., Feuillet, C., Keller, B., Rogers, J., Pozniak, C. J., Choulet, F., Distelfeld, A., Poland, J., Ronen, G., Sharpe, A. G., Barad, O., Baruch, K., Keeble-Gagnère, G., Mascher, M., Ben-Zvi, G., Josselin, A.-A., Himmelbach, A., ... Wang, L. (2018). Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science*, 361(6403), eaar7191. <https://doi.org/10.1126/science.aar7191>
- Ávila, C. M., Requena-Ramírez, M. D., Rodríguez-Suárez, C., Flores, F., Sillero, J. C., & Atienza, S. G. (2021). Genome-Wide Association Analysis for Stem Cross Section Properties, Height and Heading Date in a Collection of Spanish Durum Wheat Landraces. *Plants*, 10(6), 1123. <https://doi.org/10.3390/plants10061123>
- Bainsla, N. K., Yadav, R., Singh, G. P., & Sharma, R. K. (2020). Additive Genetic Behavior of Stem Solidness in Wheat (*Triticum aestivum* L.). *Scientific Reports*, 10(1), 7336. <https://doi.org/10.1038/s41598-020-64470-x>
- Bathini, A., Mendu, L., Singh, N. P., Cook, J., Weaver, D., Sherman, J., Hager, M., Mondal, S., & Mendu, V. (2023). A “solid” solution for wheat stem sawfly (Hymenoptera: Cephidae) resistance: Genetics, breeding and development of solid stem wheat. *Crop Science*, 63(4), 1934–1948. <https://doi.org/10.1002/csc2.20998>
- Beres, B. L., Cárcamo, H. A., & Bremer, E. (2009). Evaluation of Alternative Planting Strategies to Reduce Wheat Stem Sawfly (Hymenoptera: Cephidae) Damage to Spring Wheat in the Northern Great Plains. *Journal of Economic Entomology*, 102(6), 2137–2145. <https://doi.org/10.1603/029.102.0617>
- Beres, B. L., Cárcamo, H. A., & Byers, J. R. (2007). Effect of Wheat Stem Sawfly Damage on Yield and Quality of Selected Canadian Spring Wheat. *Journal of Economic Entomology*, 100(1), 79–87. <https://doi.org/10.1093/jee/100.1.79>

- Beres, B. L., Dosdall, L. M., Weaver, D. K., Cárcamo, H. A., & Spaner, D. M. (2011). Biology and integrated management of wheat stem sawfly and the need for continuing research. *The Canadian Entomologist*, 143(2), 105–125. <https://doi.org/10.4039/n10-056>
- Beres, B. L., McKenzie, R. H., Cárcamo, H. A., Dosdall, L. M., Evenden, M. L., Yang, R., & Spaner, D. M. (2012). Influence of Seeding Rate, Nitrogen Management, and Micronutrient Blend Applications on Pith Expression in Solid-Stemmed Spring Wheat. *Crop Science*, 52(3), 1316–1329. <https://doi.org/10.2135/cropsci2011.05.0239>
- Berg, J. E., Lamb, P. F., Miller, J. H., Wichman, D. M., Stougaard, R. N., Eckhoff, J. L., Kephart, K. D., Nash, D. L., Grey, W. E., Gettel, D., Larson, R., Jin, Y., Kolmer, J. A., Chen, X., Bai, G., & Bruckner, P. L. (2014). Registration of ‘Warhorse’ Wheat. *Journal of Plant Registrations*, 8(2), 173–176. <https://doi.org/10.3198/jpr2014.01.0001crc>
- Bhatta, M., Morgounov, A., Belamkar, V., & Baenziger, P. S. (2018). Genome-Wide Association Study Reveals Novel Genomic Regions for Grain Yield and Yield-Related Traits in Drought-Stressed Synthetic Hexaploid Wheat. *International Journal of Molecular Sciences*, 19(10), 3011. <https://doi.org/10.3390/ijms19103011>
- Biyiklioglu, S., Alptekin, B., Akpinar, B. A., Varella, A. C., Hofland, M. L., Weaver, D. K., Bothner, B., & Budak, H. (2018). A large-scale multiomics analysis of wheat stem solidness and the wheat stem sawfly feeding response, and syntenic associations in barley, Brachypodium, and rice. *Functional & Integrative Genomics*, 18(3), 241–259. <https://doi.org/10.1007/s10142-017-0585-5>
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>

- Brien, C. (2015). *asremlPlus: Augments “ASReml-R” in Fitting Mixed Models and Packages Generally in Exploring Prediction Differences* (p. 4.4.58) [Dataset].
<https://doi.org/10.32614/CRAN.package.asremlPlus>
- Broeckling, C. D., Afsar, F. A., Neumann, S., Ben-Hur, A., & Prenni, J. E. (2014). RAMClust: A Novel Feature Clustering Method Enables Spectral-Matching-Based Annotation for Metabolomics Data. *Analytical Chemistry*, 86(14), 6812–6817.
<https://doi.org/10.1021/ac501530d>
- Brown, J. K. M. (2002). Yield penalties of disease resistance in crops. *Current Opinion in Plant Biology*, 5(4), 339–344. [https://doi.org/10.1016/S1369-5266\(02\)00270-4](https://doi.org/10.1016/S1369-5266(02)00270-4)
- Brown, J. K. M., & Rant, J. C. (2013). Fitness costs and trade-offs of disease resistance and their consequences for breeding arable crops. *Plant Pathology*, 62(S1), 83–95.
<https://doi.org/10.1111/ppa.12163>
- Browning, B. L., Zhou, Y., & Browning, S. R. (2018). A One-Penny Imputed Genome from Next-Generation Reference Panels. *The American Journal of Human Genetics*, 103(3), 338–348. <https://doi.org/10.1016/j.ajhg.2018.07.015>
- Buteler, M., Weaver, D. K., & Peterson, R. k. D. (2009). Oviposition Behavior of the Wheat Stem Sawfly When Encountering Plants Infested With Cryptic Conspecifics. *Environmental Entomology*, 38(6), 1707–1715. <https://doi.org/10.1603/022.038.0624>
- Butler, D. G., Cullis, B. R., Gilmour, A. R., Gogel, B. J., & Thompson, R. (2023). *ASReml-R Reference Manual Version 4.2* (Version 4.2) [Computer software]. VSN International Ltd/. <https://asreml.kb.vsn.co.uk/wp-content/uploads/sites/3/ASReml-R-Reference-Manual-4.2.pdf>

- Carlson, G. R., Berg, J. E., Kephart, K. D., Wichman, D. M., Lamb, P. F., Miller, J. H., Stougaard, R. N., Eckhoff, J. L., Riveland, N. R., Nash, D. L., Grey, W. E., Jin, Y., Kolmer, J. A., Chen, X., Bai, G., & Bruckner, P. L. (2013). Registration of ‘Judee’ Wheat. *Journal of Plant Registrations*, 7(2), 191–194.
<https://doi.org/10.3198/jpr2012.07.0015crc>
- Carlson, G. R., Berg, J. E., Stougaard, R. N., Eckhoff, J. L., Lamb, P. F., Kephart, K. D., Wichman, D. M., Miller, J. H., Riveland, N. R., Nash, D. L., Grey, W. E., Jin, Y., Kolmer, J. A., Chen, X., Bai, G., & Bruckner, P. L. (2013). Registration of ‘Bearpaw’ Wheat. *Journal of Plant Registrations*, 7(2), 180–183.
<https://doi.org/10.3198/jpr2012.07.0009crc>
- Carrillo, E., Rubiales, D., & Castillejo, M. A. (2014). Proteomic Analysis of Pea (*Pisum sativum* L.) Response During Compatible and Incompatible Interactions with the Pea Aphid (*Acyrtosiphon pisum* H.). *Plant Molecular Biology Reporter*, 32(3), 697–718.
<https://doi.org/10.1007/s11105-013-0677-x>
- Chen, C., Powell, O., Dinglasan, E., Ross, E. M., Yadav, S., Wei, X., Atkin, F., Deomano, E., & Hayes, B. J. (2023). Genomic prediction with machine learning in sugarcane, a complex highly polyploid clonally propagated crop with substantial non-additive variation for key traits. *The Plant Genome*, 16(4), e20390. <https://doi.org/10.1002/tpg2.20390>
- Chen, L., Li, X., Zhang, J., He, T., Huang, J., Zhang, Z., Wang, Y., Hafeez, M., Zhou, S., Ren, X., Hou, Y., & Lu, Y. (2021). Comprehensive Metabolome and Volatilome Analyses in Eggplant and Tomato Reveal Their Differential Responses to *Tuta absoluta* Infestation. *Frontiers in Plant Science*, 12, 757230. <https://doi.org/10.3389/fpls.2021.757230>

- Chen, X., & Ishwaran, H. (2012). Random forests for genomic data analysis. *Genomics*, 99(6), 323–329. <https://doi.org/10.1016/j.ygeno.2012.04.003>
- Chen, X., Xu, Z., Zhao, B., Yang, Y., Mi, J., Zhao, Z., & Liu, J. (2022). Physiological and Proteomic Analysis Responsive Mechanisms for Salt Stress in Oat. *Frontiers in Plant Science*, 13, 891674. <https://doi.org/10.3389/fpls.2022.891674>
- Cisse, E.-H. M., Zhang, J., Li, D.-D., Miao, L.-F., Yin, L.-Y., & Yang, F. (2022). Exogenous ABA and IAA modulate physiological and hormonal adaptation strategies in *Cleistocalyx operculatus* and *Syzygium jambos* under long-term waterlogging conditions. *BMC Plant Biology*, 22(1), 523. <https://doi.org/10.1186/s12870-022-03888-z>
- Clark, S. A., & Van Der Werf, J. (2013). Genomic Best Linear Unbiased Prediction (gBLUP) for the Estimation of Genomic Breeding Values. In C. Gondro, J. Van Der Werf, & B. Hayes (Eds.), *Genome-Wide Association Studies and Genomic Prediction* (Vol. 1019, pp. 321–330). Humana Press. https://doi.org/10.1007/978-1-62703-447-0_13
- Colin A. Smith <Csmith@Scripps. Edu>, R. T. C. (2017). *Xcms* [Computer software]. Bioconductor. <https://doi.org/10.18129/B9.BIOC.XCMS>
- Cook, J. P., Blake, N. K., Heo, H. Y., Martin, J. M., Weaver, D. K., & Talbert, L. E. (2017a). Phenotypic and Haplotype Diversity among Tetraploid and Hexaploid Wheat Accessions with Potentially Novel Insect Resistance Genes for Wheat Stem Sawfly. *The Plant Genome*, 10(1), plantgenome2016.03.0026. <https://doi.org/10.3835/plantgenome2016.03.0026>
- Cook, J. P., Blake, N. K., Heo, H. Y., Martin, J. M., Weaver, D. K., & Talbert, L. E. (2017b). Phenotypic and Haplotype Diversity among Tetraploid and Hexaploid Wheat Accessions with Potentially Novel Insect Resistance Genes for Wheat Stem Sawfly. *The Plant*

Genome, 10(1), plantgenome2016.03.0026.

<https://doi.org/10.3835/plantgenome2016.03.0026>

Cook, J. P., Weaver, D. K., Varella, A. C., Sherman, J. D., Hofland, M. L., Heo, H. -Y., Caron, C., Lamb, P. F., Blake, N. K., & Talbert, L. E. (2019a). Comparison of Three Alleles at a Major Solid Stem QTL for Wheat Stem Sawfly Resistance and Agronomic Performance in Hexaploid Wheat. *Crop Science*, 59(4), 1639–1647.

<https://doi.org/10.2135/cropsci2019.01.0009>

Cook, J. P., Weaver, D. K., Varella, A. C., Sherman, J. D., Hofland, M. L., Heo, H. -Y., Caron, C., Lamb, P. F., Blake, N. K., & Talbert, L. E. (2019b). Comparison of Three Alleles at a Major Solid Stem QTL for Wheat Stem Sawfly Resistance and Agronomic Performance in Hexaploid Wheat. *Crop Science*, 59(4), 1639–1647.

<https://doi.org/10.2135/cropsci2019.01.0009>

Cook, J. P., Wichman, D. M., Martin, J. M., Bruckner, P. L., & Talbert, L. E. (2004a). Identification of Microsatellite Markers Associated with a Stem Solidness Locus in Wheat. *Crop Science*, 44(4), 1397–1402. <https://doi.org/10.2135/cropsci2004.1397>

Cook, J. P., Wichman, D. M., Martin, J. M., Bruckner, P. L., & Talbert, L. E. (2004b). Identification of Microsatellite Markers Associated with a Stem Solidness Locus in Wheat. *Crop Science*, 44(4), 1397–1402. <https://doi.org/10.2135/cropsci2004.1397>

Crossa, J., Pérez-Rodríguez, P., Cuevas, J., Montesinos-López, O., Jarquín, D., De Los Campos, G., Burgueño, J., González-Camacho, J. M., Pérez-Elizalde, S., Beyene, Y., Dreisigacker, S., Singh, R., Zhang, X., Gowda, M., Roorkiwal, M., Rutkoski, J., & Varshney, R. K. (2017). Genomic Selection in Plant Breeding: Methods, Models, and Perspectives. *Trends in Plant Science*, 22(11), 961–975. <https://doi.org/10.1016/j.tplants.2017.08.011>

- Curtis, T., & Halford, N. G. (2014). Food security: The challenge of increasing wheat yield and the importance of not compromising food safety. *Annals of Applied Biology*, 164(3), 354–372. <https://doi.org/10.1111/aab.12108>
- Daba, S. D., Liu, X., Aryal, U., & Mohammadi, M. (2020). A proteomic analysis of grain yield-related traits in wheat. *AoB PLANTS*, 12(5), plaa042. <https://doi.org/10.1093/aobpla/plaa042>
- de Almeida Filho, J. E., Guimarães, J. F. R., Fonsceca E Silva, F., Vilela de Resende, M. D., Muñoz, P., Kirst, M., & de Resende Júnior, M. F. R. (2019). Genomic Prediction of Additive and Non-additive Effects Using Genetic Markers and Pedigrees. *G3*, 9(8), 2739–2748. <https://doi.org/10.1534/g3.119.201004>
- De Los Campos, G., & Perez Rodriguez, P. (2013). *BGLR: Bayesian Generalized Linear Regression* (p. 1.1.4) [Dataset]. <https://doi.org/10.32614/CRAN.package.BGLR>
- DePauw, R. M., Townley-Smith, T. F., Humphreys, G., Knox, R. E., Clarke, F. R., & Clarke, J. M. (2005). Lillian hard red spring wheat. *Canadian Journal of Plant Science*, 85(2), 397–401. <https://doi.org/10.4141/P04-137>
- Domínguez-Mendez, R., Alcántara-de la Cruz, R., Rojano-Delgado, A. M., Fernández-Moreno, P. T., Aponte, R., & De Prado, R. (2017). Multiple mechanisms are involved in new imazamox-resistant varieties of durum and soft wheat. *Scientific Reports*, 7(1), 14839. <https://doi.org/10.1038/s41598-017-13874-3>
- Dong, Y., Fang, X., Yang, Y., Xue, G.-P., Chen, X., Zhang, W., Wang, X., Yu, C., Zhou, J., Mei, Q., Fang, W., Yan, C., & Chen, J. (2017). Comparative Proteomic Analysis of Susceptible and Resistant Rice Plants during Early Infestation by Small Brown Planthopper. *Frontiers in Plant Science*, 8, 1744. <https://doi.org/10.3389/fpls.2017.01744>

- Dreccer, M. F., Macdonald, B., Farnsworth, C. A., Paccapelo, M. V., Awasi, M. A., Condon, A. G., Forrest, K., Lee Long, I., & McIntyre, C. L. (2022). Multi-donor $\times$ elite-based populations reveal QTL for low-lodging wheat. *Theoretical and Applied Genetics*, 135(5), 1685–1703. <https://doi.org/10.1007/s00122-022-04063-6>
- Dührkop, K., Fleischauer, M., Ludwig, M., Aksenov, A. A., Melnik, A. V., Meusel, M., Dorrestein, P. C., Rousu, J., & Böcker, S. (2019). SIRIUS 4: A rapid tool for turning tandem mass spectra into metabolite structure information. *Nature Methods*, 16(4), 299–302. <https://doi.org/10.1038/s41592-019-0344-8>
- Dwivedi, S. L., Reynolds, M. P., & Ortiz, R. (2021). Mitigating tradeoffs in plant breeding. *iScience*, 24(9), 102965. <https://doi.org/10.1016/j.isci.2021.102965>
- Edwards, S. M., Buntjer, J. B., Jackson, R., Bentley, A. R., Lage, J., Byrne, E., Burt, C., Jack, P., Berry, S., Flatman, E., Poupard, B., Smith, S., Hayes, C., Gaynor, R. C., Gorjanc, G., Howell, P., Ober, E., Mackay, I. J., & Hickey, J. M. (2019). The effects of training population design on genomic prediction accuracy in wheat. *Theoretical and Applied Genetics*. <https://doi.org/10.1007/s00122-019-03327-y>
- Eng, J. K., McCormack, A. L., & Yates, J. R. (1994). An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *Journal of the American Society for Mass Spectrometry*, 5(11), 976–989. [https://doi.org/10.1016/1044-0305\(94\)80016-2](https://doi.org/10.1016/1044-0305(94)80016-2)
- Evert, R. F., Eichhorn, S. E., & Esau, K. (2006). *Esau's Plant anatomy: Meristems, cells, and tissues of the plant body : their structure, function, and development* (3rd ed). Wiley-Interscience.

- FAO. (2023). *FAO Cereal Supply and Demand Brief* [Dataset].
<https://www.fao.org/worldfoodsituation/csdb/en/>
- Fenn, M. A., & Giovannoni, J. J. (2021). Phytohormones in fruit development and maturation. *The Plant Journal*, 105(2), 446–458. <https://doi.org/10.1111/tpj.15112>
- Fernie, A. R., & Schauer, N. (2009). Metabolomics-assisted breeding: A viable option for crop improvement? *Trends in Genetics*, 25(1), 39–48. <https://doi.org/10.1016/j.tig.2008.10.010>
- Ferreira, M. S. (2025). Evolution of complexity through regulatory variation at a single gene. *Proceedings of the National Academy of Sciences of the United States of America*, 122(4), e2424050122. <https://doi.org/10.1073/pnas.2424050122>
- Fisher, R. A. (1936). *Statistical methods for research workers*. Genesis Publishing Pvt Ltd.
- Fryett, J. J., Morris, A. P., & Cordell, H. J. (2020). Investigation of prediction accuracy and the impact of sample size, ancestry, and tissue in transcriptome-wide association studies. *Genetic Epidemiology*, 44(5), 425–441. <https://doi.org/10.1002/gepi.22290>
- Fürstenberg-Hägg, J., Zagrobelny, M., & Bak, S. (2013). Plant Defense against Insect Herbivores. *International Journal of Molecular Sciences*, 14(5), 10242–10297. <https://doi.org/10.3390/ijms140510242>
- Ganal, M. W., Plieske, J., Hohmeyer, A., Polley, A., & Röder, M. S. (2019). High-Throughput Genotyping for Cereal Research and Breeding. In *Applications of Genetic and Genomic Research in Cereals* (pp. 3–17). Elsevier. <https://doi.org/10.1016/B978-0-08-102163-7.00001-6>
- Genuer, M., Poggi, J.-M., & Tuleau, C. (2008). *Random Forests: Some methodological insights* (Research Report No. 6729). INRIA. <https://inria.hal.science/inria-00340725/en/>

- Glaubitz, J. C., Casstevens, T. M., Lu, F., Harriman, J., Elshire, R. J., Sun, Q., & Buckler, E. S. (2014). TASSEL-GBS: A High Capacity Genotyping by Sequencing Analysis Pipeline. *PLoS ONE*, 9(2), e90346. <https://doi.org/10.1371/journal.pone.0090346>
- Goddard, M. (2009). Genomic selection: Prediction of accuracy and maximisation of long term response. *Genetica*, 136(2), 245–257. <https://doi.org/10.1007/s10709-008-9308-0>
- Goldstein, B. A., Hubbard, A. E., Cutler, A., & Barcellos, L. F. (2010). An application of Random Forests to a genome-wide association dataset: Methodological considerations & new findings. *BMC Genetics*, 11, 49. <https://doi.org/10.1186/1471-2156-11-49>
- González-Recio, O., & Forni, S. (2011). Genome-wide prediction of discrete traits using bayesian regressions and machine learning. *Genetics Selection Evolution*, 43(1), 7. <https://doi.org/10.1186/1297-9686-43-7>
- Graham, S. E., & Frels, K. (2025). Limitations of phenomic prediction for evaluating wheat stem sawfly resistance in wheat breeding programs. *The Plant Phenome Journal*, 8(1), e70050. <https://doi.org/10.1002/ppj2.70050>
- Gutierrez-Gonzalez, J. J., Mascher, M., Poland, J., & Muehlbauer, G. J. (2019). Dense genotyping-by-sequencing linkage maps of two Synthetic W7984×Opata reference populations provide insights into wheat structural diversity. *Scientific Reports*, 9(1), 1793. <https://doi.org/10.1038/s41598-018-38111-3>
- Hager, M. S., Cook, J. P., Bothner, B., & Weaver, D. K. (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. *Frontiers in Plant Science*, 16, 1497732. <https://doi.org/10.3389/fpls.2025.1497732>

- Hager, M. S., Hofland, M. L., Varella, A. C., Bothner, B., Budak, H., & Weaver, D. K. (2024). Untargeted metabolomics profiling of oat (*Avena sativa* L.) and wheat (*Triticum aestivum* L.) infested with wheat stem sawfly (*Cephus cinctus* Norton) reveals differences associated with plant defense and insect nutrition. *Frontiers in Plant Science*, *15*, 1327390. <https://doi.org/10.3389/fpls.2024.1327390>
- Hai, L., Guo, H., Xiao, S., Jiang, G., Zhang, X., Yan, C., Xin, Z., & Jia, J. (2005). Quantitative trait loci (QTL) of stem strength and related traits in a doubled-haploid population of wheat (*Triticum aestivum* L.). *Euphytica*, *141*(1–2), 1–9. <https://doi.org/10.1007/s10681-005-4713-2>
- Haley, S. D., Johnson, J. J., Peairs, F. B., Stromberger, J. A., Hudson, E. E., Seifert, S. A., Kottke, R. A., Valdez, V. A., Rudolph, J. B., Bai, G., Chen, X., Bowden, R. L., Jin, Y., Kolmer, J. A., Chen, M.-S., & Seabourn, B. W. (2012). Registration of ‘Byrd’ Wheat. *Journal of Plant Registrations*, *6*(3), 302–305. <https://doi.org/10.3198/jpr2011.12.0672crc>
- Haley, S. D., Johnson, J. J., Peairs, F. B., Stromberger, J. A., Hudson, E. E., Seifert, S. A., Kottke, R. A., Valdez, V. A., Rudolph, J. B., Martin, T. J., Bai, G., Chen, X., Bowden, R. L., Jin, Y., Kolmer, J. A., Chen, M.-S., & Seabourn, B. W. (2012). Registration of ‘Denali’ Wheat. *Journal of Plant Registrations*, *6*(3), 311–314. <https://doi.org/10.3198/jpr2011.12.0675crc>
- Haley, S. D., Johnson, J. J., Westra, P. H., Peairs, F. B., Stromberger, J. A., Hudson, E. E., Seifert, S. A., Kottke, R. A., Valdez, V. A., Rudolph, J. B., Bai, G., Chen, X., Bowden, R. L., Jin, Y., Kolmer, J. A., Chen, M.-S., & Seabourn, B. W. (2012). Registration of ‘Brawl CL Plus’ Wheat. *Journal of Plant Registrations*, *6*(3), 306–310. <https://doi.org/10.3198/jpr2011.12.0673crc>

- Haley, S. D., Quick, J. S., Johnson, J. J., Peairs, F. B., Stromberger, J. A., Clayshulte, S. R., Clifford, B. L., Rudolph, J. B., Seabourn, B. W., Chung, O. K., Jin, Y., & Kolmer, J. (2005). Registration of ‘Hatcher’ Wheat. *Crop Science*, 45(6), 2654–2656.
<https://doi.org/10.2135/cropsci2005.0030>
- Hamblin, M. T., Buckler, E. S., & Jannink, J.-L. (2011). Population genetics of genomics-based crop improvement methods. *Trends in Genetics*, 27(3), 98–106.
<https://doi.org/10.1016/j.tig.2010.12.003>
- Hao, S., Lou, H., Wang, H., Shi, J., Liu, D., Baogerile, Tao, J., Miao, S., Pei, Q., Yu, L., Wu, M., Gao, M., Zhao, N., Dong, J., You, M., & Xin, M. (2022). Genome-Wide Association Study Reveals the Genetic Basis of Five Quality Traits in Chinese Wheat. *Frontiers in Plant Science*, 13, 835306. <https://doi.org/10.3389/fpls.2022.835306>
- Hayat, M. A., Martin, J. M., Lanning, S. P., McGuire, C. F., & Talbert, L. E. (1995). Variation for stem solidness and its association with agronomic traits in spring wheat. *Canadian Journal of Plant Science*, 75(4), 775–780. <https://doi.org/10.4141/cjps95-131>
- Heffner, E. L., Jannink, J., & Sorrells, M. E. (2011). Genomic Selection Accuracy using Multifamily Prediction Models in a Wheat Breeding Program. *The Plant Genome*, 4(1), 65–75. <https://doi.org/10.3835/plantgenome2010.12.0029>
- Heffner, E. L., Sorrells, M. E., & Jannink, J.-L. (2009). Genomic Selection for Crop Improvement. *Crop Science*, 49(1), 1–12. <https://doi.org/10.2135/cropsci2008.08.0512>
- Holmes, N. D. (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(5), 677–684.
<https://doi.org/10.4039/Ent116677-5>

- Holmes, N. D., & Peterson, L. K. (1960). THE INFLUENCE OF THE HOST ON OVIPOSITION BY THE WHEAT STEM SAWFLY, *CEPHUS CINCTUS* NORT. (HYMENOPTERA: CEPHIDAE). *Canadian Journal of Plant Science*, 40(1), 29–46. <https://doi.org/10.4141/cjps60-004>
- Holmes, N. D., & Peterson, L. K. (1961). Resistance of Spring Wheats to the Wheat Stem Sawfly, *Cephus cinctus* Nort. (Hymenoptera: Cephidae): I. Resistance to the Egg. *The Canadian Entomologist*, 93(4), 250–260. <https://doi.org/10.4039/Ent93250-4>
- Holmes, N. D., & Peterson, L. K. (1964). Resistance to the Wheat Stem Sawfly, *Cephus cinctus* Nort. *The Canadian Entomologist*, 96(1–2), 120–120. <https://doi.org/10.4039/Ent96120-1>
- Iriti, M., & Faoro, F. (2009). Chemical diversity and defence metabolism: How plants cope with pathogens and ozone pollution. *International Journal of Molecular Sciences*, 10(8), 3371–3399. <https://doi.org/10.3390/ijms10083371>
- Jamil, M., Ali, A., Gul, A., Ghafoor, A., Napar, A. A., Ibrahim, A. M. H., Naveed, N. H., Yasin, N. A., & Mujeeb-Kazi, A. (2019). Genome-wide association studies of seven agronomic traits under two sowing conditions in bread wheat. *BMC Plant Biology*, 19(1), 149. <https://doi.org/10.1186/s12870-019-1754-6>
- Jannink, J.-L., Lorenz, A. J., & Iwata, H. (2010). Genomic selection in plant breeding: From theory to practice. *Briefings in Functional Genomics*, 9(2), 166–177. <https://doi.org/10.1093/bfgp/elq001>
- Jean Finnegan, E., Kovac, K. A., Jaligot, E., Sheldon, C. C., James Peacock, W., & Dennis, E. S. (2005). The downregulation of *FLOWERING LOCUS C* (*FLC*) expression in plants with low levels of DNA methylation and by vernalization occurs by distinct mechanisms. *The Plant Journal*, 44(3), 420–432. <https://doi.org/10.1111/j.1365-313X.2005.02541.x>

- Jimenez, F., Rojano-Delgado, A. M., Fernández, P. T., Rodríguez-Suárez, C., Atienza, S. G., & De Prado, R. (2016). Physiological, biochemical and molecular characterization of an induced mutation conferring imidazolinone resistance in wheat. *Physiologia Plantarum*, 158(1), 2–10. <https://doi.org/10.1111/ppl.12445>
- Joukhadar, R., El-Bouhssini, M., Jighly, A., & Ogbonnaya, F. C. (2013). Genome-wide association mapping for five major pest resistances in wheat. *Molecular Breeding*, 32(4), 943–960. <https://doi.org/10.1007/s11032-013-9924-y>
- Käll, L., Canterbury, J. D., Weston, J., Noble, W. S., & MacCoss, M. J. (2007). Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nature Methods*, 4(11), 923–925. <https://doi.org/10.1038/nmeth1113>
- Kang, H. M., Sul, J. H., Service, S. K., Zaitlen, N. A., Kong, S., Freimer, N. B., Sabatti, C., & Eskin, E. (2010). Variance component model to account for sample structure in genome-wide association studies. *Nature Genetics*, 42(4), 348–354. <https://doi.org/10.1038/ng.548>
- Kärkkäinen, H. P., & Sillanpää, M. J. (2012). Back to basics for Bayesian model building in genomic selection. *Genetics*, 191(3), 969–987. <https://doi.org/10.1534/genetics.112.139014>
- Kebrom, T. H., Chandler, P. M., Swain, S. M., King, R. W., Richards, R. A., & Spielmeyer, W. (2012). Inhibition of Tiller Bud Outgrowth in the *tin* Mutant of Wheat Is Associated with Precocious Internode Development. *Plant Physiology*, 160(1), 308–318. <https://doi.org/10.1104/pp.112.197954>
- Kemp, H. (1934). Studies of solid stem wheat varieties in relation to wheat stem sawfly control. *Scientific Agriculture*, 15(1), 30–38.

- Kirby, E. J. M. (n.d.). *Botany of the wheat plant*. Retrieved October 13, 2023, from <https://www.fao.org/3/Y4011E/y4011e05.htm>
- Kliebenstein, Daniel J. (2009). Advancing Genetic Theory and Application by Metabolic Quantitative Trait Loci Analysis. *The Plant Cell*, 21(6), 1637–1646. <https://doi.org/10.1105/tpc.109.067611>
- Korte, A., & Farlow, A. (2013). The advantages and limitations of trait analysis with GWAS: A review. *Plant Methods*, 9, 29. <https://doi.org/10.1186/1746-4811-9-29>
- Kuhn, M. (2008). Building Predictive Models in R Using the **caret** Package. *Journal of Statistical Software*, 28(5). <https://doi.org/10.18637/jss.v028.i05>
- Kumar, P., Kumar, D., Pal, S., & Singh, S. (2025). Plant secondary metabolites in defense against phytopathogens: Mechanisms, biosynthesis, and applications. *Physiological and Molecular Plant Pathology*, 138, 102639. <https://doi.org/10.1016/j.pmpp.2025.102639>
- Kuźniak, E., & Gajewska, E. (2024). Lipids and Lipid-Mediated Signaling in Plant-Pathogen Interactions. *International Journal of Molecular Sciences*, 25(13), 7255. <https://doi.org/10.3390/ijms25137255>
- Lanning, S. P., Fox, P., Elser, J., Martin, J. M., Blake, N. K., & Talbert, L. E. (2006). Microsatellite Markers Associated with a Secondary Stem Solidness Locus in Wheat. *Crop Science*, 46(4), 1701–1703. <https://doi.org/10.2135/cropsci2005.10-0379>
- Lanning, S. P., Martin, J. M., Stougaard, R. N., Guillen-Portal, F. R., Blake, N. K., Sherman, J. D., Robbins, A. M., Kephart, K. D., Lamb, P., Carlson, G. R., Pumphrey, M., & Talbert, L. E. (2012). Evaluation of Near-Isogenic Lines for Three Height-Reducing Genes in Hard Red Spring Wheat. *Crop Science*, 52(3), 1145–1152. <https://doi.org/10.2135/cropsci2011.11.0625>

- Larkin, Lozada, & Mason. (2019). Genomic Selection—Considerations for Successful Implementation in Wheat Breeding Programs. *Agronomy*, 9(9), 479.
<https://doi.org/10.3390/agronomy9090479>
- Larson, R. I., & Macdonald, M. D. (1966). CYTOGENETICS OF SOLID STEM IN COMMON WHEAT. V. LINES OF S-615 WITH WHOLE CHROMOSOME SUBSTITUTIONS FROM APEX. *Canadian Journal of Genetics and Cytology*, 8(1), 64–70.
<https://doi.org/10.1139/g66-008>
- Lavergne, F. D., Broeckling, C. D., Brown, K. J., Cockrell, D. M., Haley, S. D., Peairs, F. B., Pearce, S., Wolfe, L. M., Jahn, C. E., & Heuberger, A. L. (2020). Differential Stem Proteomics and Metabolomics Profiles for Four Wheat Cultivars in Response to the Insect Pest Wheat Stem Sawfly. *Journal of Proteome Research*, 19(3), 1037–1051.
<https://doi.org/10.1021/acs.jproteome.9b00561>
- Li, B., Zhang, N., Wang, Y.-G., George, A. W., Reverter, A., & Li, Y. (2018). Genomic Prediction of Breeding Values Using a Subset of SNPs Identified by Three Machine Learning Methods. *Frontiers in Genetics*, 9, 237. <https://doi.org/10.3389/fgene.2018.00237>
- Li, D., Xu, Z., Gu, R., Wang, P., Lyle, D., Xu, J., Zhang, H., & Wang, G. (2019). Enhancing genomic selection by fitting large-effect SNPs as fixed effects and a genotype-by-environment effect using a maize BC1F3:4 population. *PloS One*, 14(10), e0223898.
<https://doi.org/10.1371/journal.pone.0223898>
- Li, F., Wen, W., Liu, J., Zhang, Y., Cao, S., He, Z., Rasheed, A., Jin, H., Zhang, C., Yan, J., Zhang, P., Wan, Y., & Xia, X. (2019). Genetic architecture of grain yield in bread wheat based on genome-wide association studies. *BMC Plant Biology*, 19(1), 168.
<https://doi.org/10.1186/s12870-019-1781-3>

- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, 25(14), 1754–1760.
<https://doi.org/10.1093/bioinformatics/btp324>
- Liang, M., An, B., Li, K., Du, L., Deng, T., Cao, S., Du, Y., Xu, L., Gao, X., Zhang, L., Li, J., & Gao, H. (2022). Improving Genomic Prediction with Machine Learning Incorporating TPE for Hyperparameters Optimization. *Biology*, 11(11), 1647.
<https://doi.org/10.3390/biology11111647>
- Liang, W., Zhang, Z., Yao, N., Wang, B., Yu, W., Zhu, Q., Yang, S., Zeng, J., Wang, L., & Liang, W. (2025). Glycolysis and signal transduction participate in *Lycium barbarum* in response to NaCl stress through protein phosphorylation. *BMC Plant Biology*, 25(1), 405.
<https://doi.org/10.1186/s12870-025-06402-3>
- Liu, J., Islam, M. T., Sapkota, S., Ravindran, P., Kumar, P. P., Artlip, T. S., & Sherif, S. M. (2021). Ethylene-Mediated Modulation of Bud Phenology, Cold Hardiness, and Hormone Biosynthesis in Peach (*Prunus persica*). *Plants*, 10(7), 1266.
<https://doi.org/10.3390/plants10071266>
- Liu, Y., Lin, Y., Gao, S., Li, Z., Ma, J., Deng, M., Chen, G., Wei, Y., & Zheng, Y. (2017). A genome-wide association study of 23 agronomic traits in Chinese wheat landraces. *The Plant Journal*, 91(5), 861–873. <https://doi.org/10.1111/tpj.13614>
- Lorenz, K., & Cohen, B. A. (2012). Small- and large-effect quantitative trait locus interactions underlie variation in yeast sporulation efficiency. *Genetics*, 192(3), 1123–1132.
<https://doi.org/10.1534/genetics.112.143107>

- Lötsch, J., Kringel, D., & Ultsch, A. (2024). Revisiting Fold-Change Calculation: Preference for Median or Geometric Mean over Arithmetic Mean-Based Methods. *Biomedicines*, 12(8), 1639. <https://doi.org/10.3390/biomedicines12081639>
- Lou, H., Zhang, R., Liu, Y., Guo, D., Zhai, S., Chen, A., Zhang, Y., Xie, C., You, M., Peng, H., Liang, R., Ni, Z., Sun, Q., & Li, B. (2021). Genome-wide association study of six quality-related traits in common wheat (*Triticum aestivum* L.) under two sowing conditions. *Theoretical and Applied Genetics*, 134(1), 399–418. <https://doi.org/10.1007/s00122-020-03704-y>
- Ma, J., Ding, P., Liu, J., Li, T., Zou, Y., Habib, A., Mu, Y., Tang, H., Jiang, Q., Liu, Y., Chen, G., Wang, J., Deng, M., Qi, P., Li, W., Pu, Z., Zheng, Y., Wei, Y., & Lan, X. (2019). Identification and validation of a major and stably expressed QTL for spikelet number per spike in bread wheat. *Theoretical and Applied Genetics*, 132(11), 3155–3167. <https://doi.org/10.1007/s00122-019-03415-z>
- MacIntosh, G. C. (2019). Gene pyramids and the balancing act of keeping pests at bay. *Journal of Experimental Botany*, 70(18), 4591–4593. <https://doi.org/10.1093/jxb/erz216>
- MacLean, B., Tomazela, D. M., Shulman, N., Chambers, M., Finney, G. L., Frewen, B., Kern, R., Tabb, D. L., Liebler, D. C., & MacCoss, M. J. (2010). Skyline: An open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics*, 26(7), 966–968. <https://doi.org/10.1093/bioinformatics/btq054>
- Makris, A., Daloglu, B., Gutjahr, C., & Ried-Lasi, M. K. (2026). Absciscic acid at the crossroad of abiotic stress responses and plant-microbe interactions. *PLoS Pathogens*, 22(1), e1013872. <https://doi.org/10.1371/journal.ppat.1013872>

- Mazur, B. J., Chui, C. F., & Smith, J. K. (1987). Isolation and characterization of plant genes coding for acetolactate synthase, the target enzyme for two classes of herbicides. *Plant Physiology*, 85(4), 1110–1117. <https://doi.org/10.1104/pp.85.4.1110>
- McNeal, F. H., & Berg, M. A. (1979). Stem solidness and its relationship to grain yield in 17 spring wheat crosses. *Euphytica*, 28(1), 89–91. <https://doi.org/10.1007/BF00029177>
- Meher, P. K., Gupta, A., Rustgi, S., Mir, R. R., Kumar, A., Kumar, J., Balyan, H. S., & Gupta, P. K. (2023). Evaluation of eight Bayesian genomic prediction models for three micronutrient traits in bread wheat (*Triticum aestivum* L.). *The Plant Genome*, 16(4), e20332. <https://doi.org/10.1002/tpg2.20332>
- Meher, P. K., Rustgi, S., & Kumar, A. (2022). Performance of Bayesian and BLUP alphabets for genomic prediction: Analysis, comparison and results. *Heredity*, 128(6), 519–530. <https://doi.org/10.1038/s41437-022-00539-9>
- Merrick, L. F., & Carter, A. H. (2021). Comparison of genomic selection models for exploring predictive ability of complex traits in breeding programs. *The Plant Genome*, 14(3), e20158. <https://doi.org/10.1002/tpg2.20158>
- Meuwissen, T. H. E., Hayes, B. J., & Goddard, M. E. (2001). Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. *Genetics*, 157(4), 1819–1829. <https://doi.org/10.1093/genetics/157.4.1819>
- Mo, W., Zheng, X., Shi, Q., Zhao, X., Chen, X., Yang, Z., & Zuo, Z. (2024). Unveiling the crucial roles of abscisic acid in plant physiology: Implications for enhancing stress tolerance and productivity. *Frontiers in Plant Science*, 15, 1437184. <https://doi.org/10.3389/fpls.2024.1437184>

- Montesinos López, O. A., Montesinos López, A., & Crossa, J. (2022). Random Forest for Genomic Prediction. In O. A. Montesinos López, A. Montesinos López, & J. Crossa, *Multivariate Statistical Machine Learning Methods for Genomic Prediction* (pp. 633–681). Springer International Publishing. https://doi.org/10.1007/978-3-030-89010-0_15
- Montesinos-López, A., Montesinos-López, O. A., Ramos-Pulido, S., Mosqueda-González, B. A., Guerrero-Arroyo, E. A., Crossa, J., & Ortiz, R. (2025). Artificial intelligence meets genomic selection: Comparing deep learning and GBLUP across diverse plant datasets. *Frontiers in Genetics*, 16, 1568705. <https://doi.org/10.3389/fgene.2025.1568705>
- Naderi, S., Yin, T., & König, S. (2016). Random forest estimation of genomic breeding values for disease susceptibility over different disease incidences and genomic architectures in simulated cow calibration groups. *Journal of Dairy Science*, 99(9), 7261–7273. <https://doi.org/10.3168/jds.2016-10887>
- Newell, M. A., & Jannink, J.-L. (2014). Genomic Selection in Plant Breeding. In D. Fleury & R. Whitford (Eds.), *Crop Breeding* (Vol. 1145, pp. 117–130). Springer New York. https://doi.org/10.1007/978-1-4939-0446-4_10
- Nilsen, K. T., N'Diaye, A., MacLachlan, P. R., Clarke, J. M., Ruan, Y., Cuthbert, R. D., Knox, R. E., Wiebe, K., Cory, A. T., Walkowiak, S., Beres, B. L., Graf, R. J., Clarke, F. R., Sharpe, A. G., Distelfeld, A., & Pozniak, C. J. (2017). High density mapping and haplotype analysis of the major stem-solidness locus SSt1 in durum and common wheat. *PLOS ONE*, 12(4), e0175285. <https://doi.org/10.1371/journal.pone.0175285>
- Nilsen, K. T., Walkowiak, S., Xiang, D., Gao, P., Quilichini, T. D., Willick, I. R., Byrns, B., N'Diaye, A., Ens, J., Wiebe, K., Ruan, Y., Cuthbert, R. D., Craze, M., Wallington, E. J., Simmonds, J., Uauy, C., Datla, R., & Pozniak, C. J. (2020). Copy number variation of

- TdDof* controls solid-stemmed architecture in wheat. *Proceedings of the National Academy of Sciences*, 117(46), 28708–28718. <https://doi.org/10.1073/pnas.2009418117>
- Niu, Y., Chen, T., Zhao, C., Guo, C., & Zhou, M. (2022). Identification of QTL for Stem Traits in Wheat (*Triticum aestivum* L.). *Frontiers in Plant Science*, 13, 962253. <https://doi.org/10.3389/fpls.2022.962253>
- Norambuena, L., Marchant, L., Berninsone, P., Hirschberg, C. B., Silva, H., & Orellana, A. (2002). Transport of UDP-galactose in Plants. *Journal of Biological Chemistry*, 277(36), 32923–32929. <https://doi.org/10.1074/jbc.M204081200>
- Onogi, A. (2023). A Bayesian model for genomic prediction using metabolic networks. *Bioinformatics Advances*, 3(1), vbad106. <https://doi.org/10.1093/bioadv/vbad106>
- Pan, Y., Zhao, S., Wang, Z., Wang, X., Zhang, X., Lee, Y., & Xi, J. (2020). Quantitative proteomics suggests changes in the carbohydrate metabolism of maize in response to larvae of the belowground herbivore *Holotrichia parallela*. *PeerJ*, 8, e9819. <https://doi.org/10.7717/peerj.9819>
- Pang, Y., Liu, C., Wang, D., St. Amand, P., Bernardo, A., Li, W., He, F., Li, L., Wang, L., Yuan, X., Dong, L., Su, Y., Zhang, H., Zhao, M., Liang, Y., Jia, H., Shen, X., Lu, Y., Jiang, H., ... Liu, S. (2020). High-Resolution Genome-wide Association Study Identifies Genomic Regions and Candidate Genes for Important Agronomic Traits in Wheat. *Molecular Plant*, 13(9), 1311–1327. <https://doi.org/10.1016/j.molp.2020.07.008>
- Paradis, E., & Schliep, K. (2019). ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, 35(3), 526–528. <https://doi.org/10.1093/bioinformatics/bty633>

- Peirce, E. S., Cockrell, D. M., Mason, E., Haley, S., Peairs, F., & Nachappa, P. (2022). Solid Stems and Beyond: Challenges and Future Directions of Resistance to Wheat Stem Sawfly (Hymenoptera: Cephidae). *Journal of Integrated Pest Management*, 13(1), 29. <https://doi.org/10.1093/jipm/pmac023>
- Perdry, H., & Dandine-Roulland, C. (2015). *gaston: Genetic Data Handling (QC, GRM, LD, PCA) & Linear Mixed Models* (p. 1.6) [Dataset]. <https://doi.org/10.32614/CRAN.package.gaston>
- Phinzi, K., Ngetar, N. S., Pham, Q. B., Chakilu, G. G., & Szabó, S. (2023). Understanding the role of training sample size in the uncertainty of high-resolution LULC mapping using random forest. *Earth Science Informatics*, 16(4), 3667–3677. <https://doi.org/10.1007/s12145-023-01117-1>
- Phipson, B., Lee, S., Majewski, I. J., Alexander, W. S., & Smyth, G. K. (2016). Robust hyperparameter estimation protects against hypervariable genes and improves power to detect differential expression. *The Annals of Applied Statistics*, 10(2). <https://doi.org/10.1214/16-AOAS920>
- Piesik, D., Weaver, D. K., Runyon, J. B., Buteler, M., Peck, G. E., & Morrill, W. L. (2008). Behavioural responses of wheat stem sawflies to wheat volatiles. *Agricultural and Forest Entomology*, 10(3), 245–253. <https://doi.org/10.1111/j.1461-9563.2008.00380.x>
- Platt, A. W., & Farstad, C. W. (1946). The reaction of wheat varieties to wheat stem sawfly attack. *Scientific Agriculture*, 26(6), 245–253.
- Pohlert, T. (2018). *PMCMRplus: Calculate Pairwise Multiple Comparisons of Mean Rank Sums Extended* (p. 1.9.12) [Dataset]. <https://doi.org/10.32614/CRAN.package.PMCMRplus>

- Poland, J., Endelman, J., Dawson, J., Rutkoski, J., Wu, S., Manes, Y., Dreisigacker, S., Crossa, J., Sánchez-Villeda, H., Sorrells, M., & Jannink, J. (2012). Genomic Selection in Wheat Breeding using Genotyping-by-Sequencing. *The Plant Genome*, 5(3), plantgenome2012.06.0006. <https://doi.org/10.3835/plantgenome2012.06.0006>
- Portman, S. L., Jaronski, S. T., Weaver, D. K., & Reddy, G. V. P. (2018). Advancing Biological Control of the Wheat Stem Sawfly: New Strategies in a 100-yr Struggle to Manage a Costly Pest in the Northern Great Plains. *Annals of the Entomological Society of America*, 111(3), 85–91. <https://doi.org/10.1093/aesa/say002>
- Pozniak, C. J., Birk, I. T., O'Donoghue, L. S., Ménard, C., Hucl, P. J., & Singh, B. K. (2004). Physiological and Molecular Characterization of Mutation-Derived Imidazolinone Resistance in Spring Wheat. *Crop Science*, 44(4), 1434–1443. <https://doi.org/10.2135/cropsci2004.1434>
- Pridmore, T. P., French, A. P., & Pound, M. P. (2012). What lies beneath: Underlying assumptions in bioimage analysis. *Trends in Plant Science*, 17(12), 688–692. <https://doi.org/10.1016/j.tplants.2012.07.003>
- R Core Team. (2025). *R: A language and environment for statistical computing* (Version 4.5.1) [Computer software]. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Raamsdonk, L. M., Teusink, B., Broadhurst, D., Zhang, N., Hayes, A., Walsh, M. C., Berden, J. A., Brindle, K. M., Kell, D. B., Rowland, J. J., Westerhoff, H. V., Van Dam, K., & Oliver, S. G. (2001). A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations. *Nature Biotechnology*, 19(1), 45–50. <https://doi.org/10.1038/83496>

- Ramadan, W. A.-K., Mahmoud, F. E.-S., Abou-Deif, M. H., & Ali, M. (2026). Proteomic analysis of proteins responsive to drought stress in barley. *BMC Plant Biology*, 26(1), 385. <https://doi.org/10.1186/s12870-026-08176-8>
- Ramaroson, M.-L., Koutouan, C., Helesbeux, J.-J., Le Clerc, V., Hamama, L., Geoffriau, E., & Briard, M. (2022). Role of Phenylpropanoids and Flavonoids in Plant Resistance to Pests and Diseases. *Molecules*, 27(23), 8371. <https://doi.org/10.3390/molecules27238371>
- Ramezan, C. A., Warner, T. A., Maxwell, A. E., & Price, B. S. (2021). Effects of Training Set Size on Supervised Machine-Learning Land-Cover Classification of Large-Area High-Resolution Remotely Sensed Data. *Remote Sensing*, 13(3), 368. <https://doi.org/10.3390/rs13030368>
- Rand, T. A., Richmond, C. E., & Dougherty, E. T. (2017). Using matrix population models to inform biological control management of the wheat stem sawfly, *Cephus cinctus*. *Biological Control*, 109, 27–36. <https://doi.org/10.1016/j.biocontrol.2017.03.007>
- Rathan, N. D., Krishna, H., Ellur, R. K., Sehgal, D., Govindan, V., Ahlawat, A. K., Krishnappa, G., Jaiswal, J. P., Singh, J. B., Sv, S., Ambati, D., Singh, S. K., Bajpai, K., & Mahendru-Singh, A. (2022). Genome-wide association study identifies loci and candidate genes for grain micronutrients and quality traits in wheat (*Triticum aestivum* L.). *Scientific Reports*, 12(1), 7037. <https://doi.org/10.1038/s41598-022-10618-w>
- Ren, H., Lu, S., Li, S., Dong, Q., Liu, D., Ben-David, R., Chen, L., & Hu, Y.-G. (2025). Novel QTLs/genes affecting single stem elasticity, stem strength, and three lodging indices in bread wheat (*Triticum aestivum* L.) identified by genome-wide association analysis. *Theoretical and Applied Genetics*, 138(7), 139. <https://doi.org/10.1007/s00122-025-04929-5>

- Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., & Smyth, G. K. (2015). Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*, 43(7), e47–e47. <https://doi.org/10.1093/nar/gkv007>
- Rodriguez, V. M., Padilla, G., Malvar, R. A., Kallenbach, M., Santiago, R., & Butrón, A. (2018). Maize Stem Response to Long-Term Attack by *Sesamia nonagrioides*. *Frontiers in Plant Science*, 9, 522. <https://doi.org/10.3389/fpls.2018.00522>
- Rojano-Delgado, A. M., Priego-Capote, F., Luque De Castro, M. D., & De Prado, R. (2015). Mechanism of imazamox resistance of the Clearfield® wheat cultivar for better weed control. *Agronomy for Sustainable Development*, 35(2), 639–648. <https://doi.org/10.1007/s13593-014-0232-7>
- Rudd, J. C., Devkota, R. N., Baker, J. A., Peterson, G. L., Lazar, M. D., Bean, B., Worrall, D., Baughman, T., Marshall, D., Sutton, R., Rooney, L. W., Nelson, L. R., Fritz, A. K., Weng, Y., Morgan, G. D., & Seabourn, B. W. (2014). ‘TAM 112’ Wheat, Resistant to Greenbug and Wheat Curl Mite and Adapted to the Dryland Production System in the Southern High Plains. *Journal of Plant Registrations*, 8(3), 291–297. <https://doi.org/10.3198/jpr2014.03.0016crc>
- Rutkoski, J. E., Poland, J. A., Singh, R. P., Huerta-Espino, J., Bhavani, S., Barbier, H., Rouse, M. N., Jannink, J., & Sorrells, M. E. (2014). Genomic Selection for Quantitative Adult Plant Stem Rust Resistance in Wheat. *The Plant Genome*, 7(3), plantgenome2014.02.0006. <https://doi.org/10.3835/plantgenome2014.02.0006>
- Saint Pierre, C., Trethowan, R., & Reynolds, M. (2010). Stem solidness and its relationship to water-soluble carbohydrates: Association with wheat yield under water deficit. *Functional Plant Biology*, 37(2), 166–174. <https://doi.org/10.1071/FP09174>

- Sarkar, R. K., Rao, A. R., Meher, P. K., Nepolean, T., & Mohapatra, T. (2015). Evaluation of random forest regression for prediction of breeding value from genomewide SNPs. *Journal of Genetics*, 94(2), 187–192. <https://doi.org/10.1007/s12041-015-0501-5>
- Schuman, M. C., Meldau, S., Gaquerel, E., Diezel, C., McGale, E., Greenfield, S., & Baldwin, I. T. (2018). The Active Jasmonate JA-Ile Regulates a Specific Subset of Plant Jasmonate-Mediated Resistance to Herbivores in Nature. *Frontiers in Plant Science*, 9, 787. <https://doi.org/10.3389/fpls.2018.00787>
- Scopes, R. K. (1974). Measurement of protein by spectrophotometry at 205 nm. *Analytical Biochemistry*, 59(1), 277–282. [https://doi.org/10.1016/0003-2697\(74\)90034-7](https://doi.org/10.1016/0003-2697(74)90034-7)
- Sherman, J. D., Blake, N. K., Martin, J. M., Kephart, K. D., Smith, J., Clark, D. R., Hofland, M. L., Weaver, D. K., Lanning, S. P., Heo, H. -Y., Pumphrey, M., Chen, J., & Talbert, L. E. (2015). Agronomic Impact of a Stem Solidness Gene in Near-Isogenic Lines of Wheat. *Crop Science*, 55(2), 514–520. <https://doi.org/10.2135/cropsci2014.05.0403>
- Sherman, J. D., Weaver, D. K., Hofland, M. L., Sing, S. E., Buteler, M., Lanning, S. P., Naruoka, Y., Crutcher, F., Blake, N. K., Martin, J. M., Lamb, P. F., Carlson, G. R., & Talbert, L. E. (2010). Identification of Novel QTL for Sawfly Resistance in Wheat. *Crop Science*, 50(1), 73–86. <https://doi.org/10.2135/cropsci2009.03.0145>
- Shewry, P. R. (2009). Wheat. *Journal of Experimental Botany*, 60(6), 1537–1553. <https://doi.org/10.1093/jxb/erp058>
- Shewry, P. R., & Hey, S. J. (2015). The contribution of wheat to human diet and health. *Food and Energy Security*, 4(3), 178–202. <https://doi.org/10.1002/fes3.64>

- Shi, S., Li, X., Fang, L., Liu, A., Su, G., Zhang, Y., Luobu, B., Ding, X., & Zhang, S. (2021). Genomic Prediction Using Bayesian Regression Models With Global–Local Prior. *Frontiers in Genetics*, 12, 628205. <https://doi.org/10.3389/fgene.2021.628205>
- Singh, A., Ganapathysubramanian, B., Singh, A. K., & Sarkar, S. (2016). Machine Learning for High-Throughput Stress Phenotyping in Plants. *Trends in Plant Science*, 21(2), 110–124. <https://doi.org/10.1016/j.tplants.2015.10.015>
- Smith, M. R., Rao, I. M., & Merchant, A. (2018). Source-Sink Relationships in Crop Plants and Their Influence on Yield Development and Nutritional Quality. *Frontiers in Plant Science*, 9, 1889. <https://doi.org/10.3389/fpls.2018.01889>
- Smyth, G. K. (2004). Linear Models and Empirical Bayes Methods for Assessing Differential Expression in Microarray Experiments. *Statistical Applications in Genetics and Molecular Biology*, 3(1), 1–25. <https://doi.org/10.2202/1544-6115.1027>
- Song, H., Dong, T., Yan, X., Wang, W., Tian, Z., & Hu, H. (2023). Using Bayesian threshold model and machine learning method to improve the accuracy of genomic prediction for ordered categorical traits in fish. *Agriculture Communications*, 1(1), 100005. <https://doi.org/10.1016/j.agrcom.2023.100005>
- Stincone, A., Prigione, A., Cramer, T., Wamelink, M. M. C., Campbell, K., Cheung, E., Olin-Sandoval, V., Grüning, N.-M., Krüger, A., Tauqeer Alam, M., Keller, M. A., Breitenbach, M., Brindle, K. M., Rabinowitz, J. D., & Ralser, M. (2015). The return of metabolism: Biochemistry and physiology of the pentose phosphate pathway. *Biological Reviews of the Cambridge Philosophical Society*, 90(3), 927–963. <https://doi.org/10.1111/brv.12140>
- Sukumaran, S., Dreisigacker, S., Lopes, M., Chavez, P., & Reynolds, M. P. (2015). Genome-wide association study for grain yield and related traits in an elite spring wheat population

- grown in temperate irrigated environments. *Theoretical and Applied Genetics*, 128(2), 353–363. <https://doi.org/10.1007/s00122-014-2435-3>
- Sun, G., Zhu, C., Kramer, M. H., Yang, S.-S., Song, W., Piepho, H.-P., & Yu, J. (2010). Variation explained in mixed-model association mapping. *Heredity*, 105(4), 333–340. <https://doi.org/10.1038/hdy.2010.11>
- Sun, T., Rao, S., Zhou, X., & Li, L. (2022). Plant carotenoids: Recent advances and future perspectives. *Molecular Horticulture*, 2(1), 3. <https://doi.org/10.1186/s43897-022-00023-2>
- Talbert, L. E., Sherman, J. D., Hofland, M. L., Lanning, S. P., Blake, N. K., Grabbe, R., Lamb, P. F., Martin, J. M., & Weaver, D. K. (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(4), 427–432. <https://doi.org/10.1111/pbr.12184>
- Thompson, J. E., Froese, C. D., Madey, E., Smith, M. D., & Hong, Y. (1998). Lipid metabolism during plant senescence. *Progress in Lipid Research*, 37(2–3), 119–141. [https://doi.org/10.1016/S0163-7827\(98\)00006-X](https://doi.org/10.1016/S0163-7827(98)00006-X)
- Van Den Berg, I., Meuwissen, T. H. E., MacLeod, I. M., & Goddard, M. E. (2019). Predicting the effect of reference population on the accuracy of within, across, and multibreed genomic prediction. *Journal of Dairy Science*, 102(4), 3155–3174. <https://doi.org/10.3168/jds.2018-15231>
- Van Rossum, B.-J., & Kruijer, W. (2020). *statgenGWAS: Genome Wide Association Studies* (p. 1.0.12) [Dataset]. <https://doi.org/10.32614/CRAN.package.statgenGWAS>

- Vänninen, I. (2005). Alternatives to pesticides in fruit and vegetable cultivation. In *Improving the Safety of Fresh Fruit and Vegetables* (pp. 293–330). Elsevier.
<https://doi.org/10.1533/9781845690243.2.293>
- VanRaden, P. M. (2008). Efficient Methods to Compute Genomic Predictions. *Journal of Dairy Science*, 91(11), 4414–4423. <https://doi.org/10.3168/jds.2007-0980>
- Varella, A. C., Talbert, L. E., Hofland, M. L., Buteler, M., Sherman, J. D., Blake, N. K., Heo, H., Martin, J. M., & Weaver, D. K. (2016). Alleles at a quantitative trait locus for stem solidness in wheat affect temporal patterns of pith expression and level of resistance to the wheat stem sawfly. *Plant Breeding*, 135(5), 546–551.
<https://doi.org/10.1111/pbr.12398>
- Varella, A. C., Weaver, D. K., Blake, N. K., Hofland, M. L., Heo, H.-Y., Cook, J. P., Lamb, P. F., Jordan, K. W., Akhunov, E., Chao, S., & Talbert, L. E. (2019a). Analysis of recombinant inbred line populations derived from wheat landraces to identify new genes for wheat stem sawfly resistance. *Theoretical and Applied Genetics*, 132(8), 2195–2207.
<https://doi.org/10.1007/s00122-019-03347-8>
- Varella, A. C., Weaver, D. K., Blake, N. K., Hofland, M. L., Heo, H.-Y., Cook, J. P., Lamb, P. F., Jordan, K. W., Akhunov, E., Chao, S., & Talbert, L. E. (2019b). Analysis of recombinant inbred line populations derived from wheat landraces to identify new genes for wheat stem sawfly resistance. *Theoretical and Applied Genetics*, 132(8), 2195–2207.
<https://doi.org/10.1007/s00122-019-03347-8>
- Varella, A. C., Weaver, D. K., Peterson, R. K. D., Sherman, J. D., Hofland, M. L., Blake, N. K., Martin, J. M., & Talbert, L. E. (2017). Host plant quantitative trait loci affect specific

- behavioral sequences in oviposition by a stem-mining insect. *Theoretical and Applied Genetics*, 130(1), 187–197. <https://doi.org/10.1007/s00122-016-2805-0>
- Varella, A. C., Weaver, D. K., Sherman, J. D., Blake, N. K., Heo, H. Y., Kalous, J. R., Chao, S., Hofland, M. L., Martin, J. M., Kephart, K. D., & Talbert, L. E. (2015). Association Analysis of Stem Solidness and Wheat Stem Sawfly Resistance in a Panel of North American Spring Wheat Germplasm. *Crop Science*, 55(5), 2046–2055. <https://doi.org/10.2135/cropsci2014.12.0852>
- Varella, A. C., Zhang, H., Weaver, D. K., Cook, J. P., Hofland, M. L., Lamb, P., Chao, S., Martin, J. M., Blake, N. K., & Talbert, L. E. (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(6), 1999–2006. <https://doi.org/10.1534/g3.119.400240>
- Verta, J.-P., Moustakas-Verho, J. E., Donner, I., Frapin, M., Ruokolainen, A., Debes, P. V., Erkinaro, J., & Primmer, C. R. (2024). A complex mechanism translating variation of a simple genetic architecture into alternative life histories. *Proceedings of the National Academy of Sciences*, 121(48), e2402386121. <https://doi.org/10.1073/pnas.2402386121>
- Vieira, H. V., Nachappa, P., Bradshaw, J., Mason, E., & Rand, T. (2025). *An integrative approach to combat the wheat stem sawfly, Cephus cinctus (Hymenoptera: Cephidae)* [Born digital, doctoral dissertations]. Colorado State University. Libraries. <https://doi.org/10.25675/3.025695>
- Visscher, P. M., Wray, N. R., Zhang, Q., Sklar, P., McCarthy, M. I., Brown, M. A., & Yang, J. (2017). 10 Years of GWAS Discovery: Biology, Function, and Translation. *The American Journal of Human Genetics*, 101(1), 5–22. <https://doi.org/10.1016/j.ajhg.2017.06.005>

- Wang, X., Xu, Y., Hu, Z., & Xu, C. (2018). Genomic selection methods for crop improvement: Current status and prospects. *The Crop Journal*, 6(4), 330–340.
<https://doi.org/10.1016/j.cj.2018.03.001>
- Wang, Y., Gao, R., Gu, T., Li, X., Wang, M., Wang, A., & Qiu, Y. (2025). Metabolomics and Transcriptomics Reveal the Role of the Terpene Biosynthetic Pathway in the Mechanism of Insect Resistance in *Solanum habrochaites*. *Journal of Agricultural and Food Chemistry*, 73(10), 6253–6269. <https://doi.org/10.1021/acs.jafc.4c10397>
- Wang, Y., Liu, Q., Du, L., Hallerman, E. M., & Li, Y. (2020). Transcriptomic and Metabolomic Responses of Rice Plants to *Cnaphalocrocis medinalis* Caterpillar Infestation. *Insects*, 11(10), 705. <https://doi.org/10.3390/insects11100705>
- Ward, B. P., Brown-Guedira, G., Tyagi, P., Kolb, F. L., Van Sanford, D. A., Sneller, C. H., & Griffey, C. A. (2019). Multienvironment and Multitrait Genomic Selection Models in Unbalanced Early-Generation Wheat Yield Trials. *Crop Science*, 59(2), 491–507.
<https://doi.org/10.2135/cropsci2018.03.0189>
- Weaver, D. K., Buteler, M., Hofland, M. L., Runyon, J. B., Nansen, C., Talbert, L. E., Lamb, P., & Carlson, G. R. (2009). Cultivar Preferences of Ovipositing Wheat Stem Sawflies as Influenced by the Amount of Volatile Attractant. *Journal of Economic Entomology*, 102(3), 1009–1017. <https://doi.org/10.1603/029.102.0320>
- Weaver, D. K., Nansen, C., Runyon, J. B., Sing, S. E., & Morrill, W. L. (2005). Spatial distributions of *Cephus cinctus* Norton (Hymenoptera: Cephidae) and its braconid parasitoids in Montana wheat fields. *Biological Control*, 34(1), 1–11.
<https://doi.org/10.1016/j.biocontrol.2005.04.001>

- Weiss, M. J., & Morrill, W. L. (1992). Wheat Stem Sawfly (Hymenoptera: Cephidae) Revisited. *American Entomologist*, 38(4), 241–245. <https://doi.org/10.1093/ae/38.4.241>
- Winn, Z. J., Lyerly, J. H., Brown-Guedira, G., Murphy, J. P., & Mason, R. E. (2023). Utilization of a publicly available diversity panel in genomic prediction of *Fusarium* head blight resistance traits in wheat. *The Plant Genome*, 16(3), e20353. <https://doi.org/10.1002/tpg2.20353>
- Wright, M. N. (2015). *ranger: A Fast Implementation of Random Forests* (p. 0.17.0) [Dataset]. <https://doi.org/10.32614/CRAN.package.ranger>
- Xu, Y., Zhang, Y., Cui, Y., Zhou, K., Yu, G., Yang, W., Wang, X., Li, F., Guan, X., Zhang, X., Yang, Z., Xu, S., & Xu, C. (2024). GA-GBLUP: Leveraging the genetic algorithm to improve the predictability of genomic selection. *Briefings in Bioinformatics*, 25(5), bbae385. <https://doi.org/10.1093/bib/bbae385>
- Yan, W., & Wallace, D. H. (1995). Breeding for Negatively Associated Traits. In J. Janick (Ed.), *Plant Breeding Reviews* (1st ed., pp. 141–177). Wiley. <https://doi.org/10.1002/9780470650059.ch4>
- Yang, J., Zaitlen, N. A., Goddard, M. E., Visscher, P. M., & Price, A. L. (2014). Advantages and pitfalls in the application of mixed-model association methods. *Nature Genetics*, 46(2), 100–106. <https://doi.org/10.1038/ng.2876>
- Yannam, V. R. R., Lopes, M. S., & Soriano, J. M. (2025). Optimizing genomic selection models for wheat breeding under contrasting water regimes in a mediterranean environment. *Plant Methods*, 21(1), 155. <https://doi.org/10.1186/s13007-025-01467-5>
- Yi, N., & Xu, S. (2008). Bayesian LASSO for Quantitative Trait Loci Mapping. *Genetics*, 179(2), 1045–1055. <https://doi.org/10.1534/genetics.107.085589>

- Yin, B., Jia, J., Sun, X., Hu, X., Ao, M., Liu, W., Tian, Z., Liu, H., Li, D., Tian, W., Hao, Y., Xia, X., Sade, N., Brotman, Y., Fernie, A. R., Chen, J., He, Z., & Chen, W. (2024). Dynamic metabolite QTL analyses provide novel biochemical insights into kernel development and nutritional quality improvement in common wheat. *Plant Communications*, 5(5), 100792. <https://doi.org/10.1016/j.xplc.2024.100792>
- Yin, C., Zhou, P., Wang, Y., Yin, Z., & Liu, Y. (2024). Using genomic selection to improve the accuracy of genomic prediction for multi-populations in pigs. *Animal*, 18(2), 101062. <https://doi.org/10.1016/j.animal.2023.101062>
- Yu, X., & Michaels, S. D. (2010). The Arabidopsis Paf1c Complex Component CDC73 Participates in the Modification of *FLOWERING LOCUS C* Chromatin. *Plant Physiology*, 153(3), 1074–1084. <https://doi.org/10.1104/pp.110.158386>
- Zhang, J., Gill, H. S., Brar, N. K., Halder, J., Ali, S., Liu, X., Bernardo, A., St. Amand, P., Bai, G., Gill, U. S., Turnipseed, B., & Sehgal, S. K. (2022). Genomic prediction of Fusarium head blight resistance in early stages using advanced breeding lines in hard winter wheat. *The Crop Journal*, 10(6), 1695–1704. <https://doi.org/10.1016/j.cj.2022.03.010>
- Zhang, P., Jackson, E., Li, X., & Zhang, Y. (2025). Salicylic acid and jasmonic acid in plant immunity. *Horticulture Research*, 12(7), uhaf082. <https://doi.org/10.1093/hr/uhaf082>
- Zheng, W., Zhang, Q., He, J., Han, B., Zhang, Q., & Sun, D. (2025). Comparative evaluation of SNP-weighted, Bayesian, and machine learning models for genomic prediction in Holstein cattle. *BMC Genomics*, 26(1), 1037. <https://doi.org/10.1186/s12864-025-12218-0>
- Zolg, D. P., Gessulat, S., Paschke, C., Graber, M., Rathke-Kuhnert, M., Seefried, F., Fitzemeier, K., Berg, F., Lopez-Ferrer, D., Horn, D., Henrich, C., Huhmer, A., Delanghe, B., &

Frejno, M. (2025). INFERYS rescoring: Boosting peptide identifications and scoring confidence of database search results. *Rapid Communications in Mass Spectrometry*, 39(S1), e9128. <https://doi.org/10.1002/rcm.9128>

APPENDIX

CHAPTER II SUPPLEMENTAL

Supplementary Table 2.1. Analysis of variance (ANOVA) results for cutting score (CUT) in a population of only individuals with the solid stem allele at the *SS1* locus, solidity (SOLID) in a population of only individuals with the solid stem allele at the *SS1* locus, and CUT in a population of only individuals with the hollow stem allele at the *SS1* locus where Model indicates which results for the previously listed ANOVA models, Variable indicates model results for each of the model parameters, Degrees of Freedom indicates the number of independent datapoints used for estimations, Sum of Squares is the measure of variation attributed to the variable, Mean Square is the average variation (Sum of Squares/Degrees of Freedom), F-value is the test statistic to determine variation of group means, and P-value represents the probability of observing the calculated F-value.

Model	Variable	Degrees of Freedom	Sum of Squares	Mean Square	F-value	P-value
CUT in Solid Stem Population	ID	313	2161.94	6.91	5.94	***
	ENV	12	534.19	44.52	38.29	***
	ID:ENV	283	454.60	1.61	1.38	***
	Residuals	380	441.77	1.16	-	-
SOLID in Solid Stem Population	ID	313	10492.52	33.52	6.41	***
	ENV	12	882.42	73.54	14.06	***
	ID:ENV	283	2117.38	7.48	1.43	***
	Residuals	380	1987.11	5.23	-	-
CUT in Hollow Stem Population	ID	59	662.81	11.23	5.78	***
	ENV	12	399.08	33.26	17.10	***
	ID:ENV	49	99.14	2.02	1.04	0.42
	Residuals	138	268.33	1.94	-	-

* indicates significance at p-value < 0.05

** indicates significance at p-value < 0.01

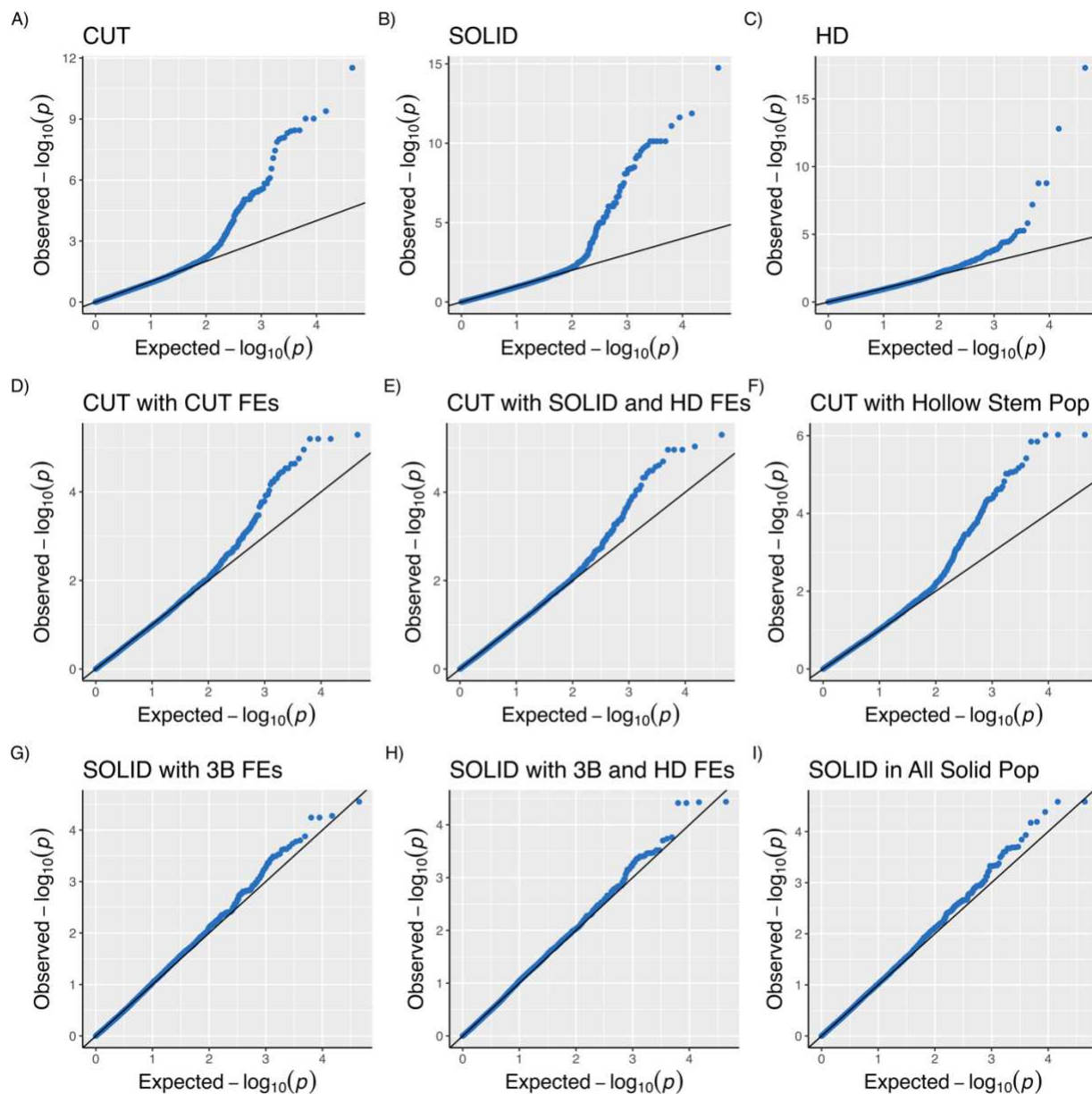
*** indicates significance at p-value < 0.001

Supplementary Table 2.2. Number of phenotypic observations for each year and location for heading date (HD), cutting score (CUT), and solidity (SOLID).

Trait	Location	2014	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025
HD	Akron	-	-	-	-	-	117	149	-	143	1399	464
	Fort Collins	-	-	-	19	52	37	22	184	238	250	132
	Grant	-	-	-	-	-	-	-	-	63	-	-
	New Raymer	-	170	296	236	207	-	150	-	394	-	231
	Orchard	-	175	393	379	291	227	246	-	168	150	-
	Sidney	-	-	-	-	-	-	-	-	66	-	-
CUT	Akron	-	-	-	-	-	397	-	649	227	1433	876
	Genoa	-	-	-	-	-	-	-	-	-	-	150
	New Raymer	144	170	296	236	207	195	150	300	395	-	309
	Orchard	86	175	393	379	291	300	246	-	-	150	502
	Roggen	-	-	-	-	-	-	96	-	-	373	350
SOLID	Akron	-	-	-	-	-	-	148	-	-	148	175
	Alliance	-	-	-	-	-	-	-	-	-	100	-
	Fort Collins	-	-	-	21	52	49	147	186	232	249	17
	Grant	-	-	-	-	-	-	-	-	58	50	-
	Mead	-	-	-	-	-	-	-	-	-	100	-
	New Raymer	-	25	132	112	80	-	-	157	-	-	-
	Orchard	-	30	128	110	80	46	-	-	-	-	116
	Sidney	-	-	-	-	-	-	-	-	66	99	-

Supplementary Table 2.3. Distribution on single nucleotide polymorphisms across all chromosomes and within sub-genomes A, B, and D.

Sub-genome	Chromosome							Total
	1	2	3	4	5	6	7	
A	1,334	1,103	1,012	914	1,177	1,130	1,584	8,254
B	1,578	1,729	1,943	798	1,336	1,846	1,377	10,607
D	621	830	502	145	291	501	456	3,346



Supplementary Figure 2.1. Quantile-quantile (QQ) plots for GWAS in the following order: (A) Cutting score (CUT), (B) stem solidity (SOLID), (C) heading date (HD), (D) CUT with pruned CUT markers as fixed effects, (E) CUT with pruned SOLID and HD markers as fixed effects, (F) CUT using only individuals with the hollow stem allele, (G) SOLID using only pruned SOLID markers as fixed-effects, (H) SOLID with pruned SOLID and HD markers as fixed effects, and (I) SOLID using only individuals with the solid stem allele.

Supplementary Table 2.4. Significant results for nine different GWAS including the GWASs for heading date (HD), cutting score (CUT), and solidity (SOLID) that had no fixed effects (FEs), the GWAS for CUT that included pruned significant CUT markers as Fes (C1), the GWAS for CUT that included pruned significant SOLID and HD markers as Fes (C2), the GWAS for CUT using only individuals with the hollow stem allele at the *SStI* locus (C3), the GWAS for SOLID using pruned significant SOLID markers as Fes (S1), the GWAS for SOLID using pruned significant markers for SOLID and HD as Fes (S2), and the GWAS using only individuals with the solid stem allele at the *SStI* locus (S3). Model indicates which GWAS the results are from, single nucleotide polymorphism (SNP) gives the name of the marker, Alleles indicates the two letter alleles at that SNP locus, CHR indicates what chromosome the SNP marker is located on, position indicates the chromosomal position of the SNP marker, MAF indicates the minor allele frequency at that locus, P-value indicates the significance level for the association of the SNP to the trait, Effect indicates the effect of the SNP marker on the trait of interest in reference to the reference allele, Effect SE is the standard error of the effect estimate, RLR2 is the likelihood-ratio-based R2 as defined by Sun et al. (2010), and LOD is the logarithm of the odds ($-\log_{10}(\text{P-Value})$) for the SNP marker.

Model	SNP	Alleles	CHR	Position	MAF	P-Value	Effect	Effect SE	RLR2	LOD
HD	S7D_70444241	G/A	7D	70444241	0.47	5.04E-18	-0.51	0.06	0.04	17.3
	S7D_80577460	T/C	7D	80577460	0.11	1.60E-13	0.65	0.09	0.03	12.8
	S7D_182010814	G/C	7D	182010814	0.19	1.67E-09	-0.41	0.07	0.02	8.78
	S7D_77767099	A/T	7D	77767099	0.09	1.70E-09	0.59	0.10	0.02	8.77
	S7D_73324993	C/T	7D	73324993	0.41	6.43E-08	-0.31	0.06	0.02	7.19
	S7A_80297844	G/A	7A	80297844	0.14	1.51E-06	-0.47	0.10	0.01	5.82
	S7D_175086569	C/T	7D	175086569	0.13	5.37E-06	-0.35	0.08	0.01	5.27
	S7A_80295746	G/A	7A	80295746	0.14	5.48E-06	-0.46	0.10	0.01	5.26
	S7D_71983459	C/T	7D	71983459	0.11	6.34E-06	-0.40	0.09	0.01	5.20
	S2B_60764872	G/A	2B	60764872	0.39	1.25E-05	0.43	0.10	0.01	4.90
	S3A_623350122	G/C	3A	623350122	0.05	2.05E-05	0.45	0.11	0.01	4.69
	S7D_73923898	C/G	7D	73923898	0.37	2.94E-05	-0.26	0.06	0.01	4.53
	S7A_80297893	C/T	7A	80297893	0.13	3.63E-05	-0.41	0.10	0.01	4.44
	S7A_80297901	G/A	7A	80297901	0.13	3.63E-05	-0.41	0.10	0.01	4.44
	S7A_78548520	G/A	7A	78548520	0.13	3.93E-05	0.40	0.10	0.01	4.41
	S7A_78548527	A/G	7A	78548527	0.13	3.93E-05	0.40	0.10	0.01	4.41
	S3A_483568913	C/T	3A	483568913	0.11	4.87E-05	0.40	0.10	0.01	4.31
	S2B_60764928	T/C	2B	60764928	0.39	7.61E-05	0.39	0.10	0.01	4.12

CUT	S1B_568404583	A/G	1B	568404583	0.13	9.49E-05	-0.51	0.13	0.01	4.02
	S3B_844992127	G/A	3B	844992127	0.47	3.01E-12	-0.54	0.08	0.02	11.52
	S3B_844846869	G/A	3B	844846869	0.46	4.09E-10	-0.45	0.07	0.02	9.39
	S3B_850636960	G/T	3B	850636960	0.46	9.54E-10	-0.45	0.07	0.02	9.02
	S3B_850636975	T/C	3B	850636975	0.46	9.54E-10	-0.45	0.07	0.02	9.02
	S3B_847656942	G/A	3B	847656942	0.48	3.58E-09	-0.43	0.07	0.01	8.45
	S3B_847656944	G/A	3B	847656944	0.48	3.58E-09	-0.43	0.07	0.01	8.45
	S3B_850711969	G/A	3B	850711969	0.54	3.95E-09	0.41	0.07	0.01	8.40
	S3B_848648204	T/C	3B	848648204	0.46	4.90E-09	-0.43	0.07	0.01	8.31
	S3B_844846941	G/A	3B	844846941	0.47	8.17E-09	-0.4	0.07	0.01	8.09
	S3B_850180824	A/G	3B	850180824	0.52	8.70E-09	0.41	0.07	0.01	8.06
	S3B_844887153	T/C	3B	844887153	0.48	9.80E-09	-0.4	0.07	0.01	8.01
	S3B_850642529	C/G	3B	850642529	0.54	1.33E-08	0.42	0.07	0.01	7.88
	S3B_847463798	T/A	3B	847463798	0.47	3.58E-08	-0.4	0.07	0.01	7.45
	S3B_847233049	A/G	3B	847233049	0.49	8.35E-08	-0.39	0.07	0.01	7.08
	S3B_847454154	C/T	3B	847454154	0.48	2.75E-07	-0.37	0.07	0.01	6.56
	S3B_850964363	A/G	3B	850964363	0.11	7.96E-07	0.4	0.08	0.01	6.10
	S3B_847391191	T/C	3B	847391191	0.47	9.26E-07	-0.34	0.07	0.01	6.03
	S3B_851413871	C/G	3B	851413871	0.53	1.48E-06	0.35	0.07	0.01	5.83
	S3B_851413872	T/A	3B	851413872	0.53	1.48E-06	0.35	0.07	0.01	5.83
	S3B_851413875	T/A	3B	851413875	0.53	1.48E-06	0.35	0.07	0.01	5.83
	S2B_5645271	T/C	2B	5645271	0.24	2.62E-06	0.28	0.06	0.01	5.58
	S3B_850973066	C/T	3B	850973066	0.1	2.69E-06	0.39	0.08	0.01	5.57
	S3B_847925770	G/C	3B	847925770	0.15	3.04E-06	0.28	0.06	0.01	5.52
	S3B_839904446	T/C	3B	839904446	0.44	3.12E-06	0.33	0.07	0.01	5.51
	S3B_839904461	G/C	3B	839904461	0.44	3.12E-06	0.33	0.07	0.01	5.51
	S3B_851847253	C/A	3B	851847253	0.11	3.77E-06	0.38	0.08	0.01	5.42
	S3B_851847276	G/A	3B	851847276	0.11	3.77E-06	0.38	0.08	0.01	5.42
	S3B_851847281	G/C	3B	851847281	0.11	3.77E-06	0.38	0.08	0.01	5.42

S3B_851847299	T/G	3B	851847299	0.11	3.77E-06	0.38	0.08	0.01	5.42
S3B_851840380	C/G	3B	851840380	0.1	4.21E-06	0.37	0.08	0.01	5.38
S3B_851840409	C/A	3B	851840409	0.1	4.21E-06	0.37	0.08	0.01	5.38
S3B_851840410	C/G	3B	851840410	0.1	4.21E-06	0.37	0.08	0.01	5.38
S2B_6547127	T/C	2B	6547127	0.21	5.60E-06	0.27	0.06	0.01	5.25
S2B_6547137	A/C	2B	6547137	0.21	5.60E-06	0.27	0.06	0.01	5.25
S2B_6547149	C/A	2B	6547149	0.21	5.60E-06	0.27	0.06	0.01	5.25
S3B_848369476	A/T	3B	848369476	0.1	8.61E-06	0.37	0.08	0.01	5.07
S3B_832035476	G/C	3B	832035476	0.05	8.90E-06	0.62	0.14	0.01	5.05
S3B_832035477	G/A	3B	832035477	0.05	8.90E-06	0.62	0.14	0.01	5.05
S3B_832035478	C/G	3B	832035478	0.05	8.90E-06	0.62	0.14	0.01	5.05
S3B_832035497	C/T	3B	832035497	0.05	8.90E-06	0.62	0.14	0.01	5.05
S3B_832035521	A/G	3B	832035521	0.05	8.90E-06	0.62	0.14	0.01	5.05
S3B_832035531	C/T	3B	832035531	0.05	8.90E-06	0.62	0.14	0.01	5.05
S7A_623850376	C/T	7A	623850376	0.1	8.91E-06	0.36	0.08	0.01	5.05
S7A_623850379	G/A	7A	623850379	0.1	8.91E-06	0.36	0.08	0.01	5.05
S7A_623850385	G/C	7A	623850385	0.1	8.91E-06	0.36	0.08	0.01	5.05
S3B_850642585	A/G	3B	850642585	0.11	9.42E-06	0.36	0.08	0.01	5.03
S3B_850642525	A/T	3B	850642525	0.11	1.37E-05	0.36	0.08	0.01	4.86
S3B_850642558	C/A	3B	850642558	0.11	1.37E-05	0.36	0.08	0.01	4.86
S3B_850642559	G/T	3B	850642559	0.11	1.37E-05	0.36	0.08	0.01	4.86
S3B_848375850	G/C	3B	848375850	0.09	1.69E-05	0.36	0.08	0.01	4.77
S3B_850711935	G/A	3B	850711935	0.11	1.80E-05	0.34	0.08	0.01	4.74
S3B_850784464	C/T	3B	850784464	0.1	2.08E-05	0.34	0.08	0.01	4.68
S3B_850784509	G/A	3B	850784509	0.1	2.08E-05	0.34	0.08	0.01	4.68
S2B_7410820	A/C	2B	7410820	0.22	2.11E-05	0.25	0.06	0.01	4.68
S3B_848060000	G/A	3B	848060000	0.07	2.42E-05	0.42	0.10	0.01	4.62
S3B_847834999	T/C	3B	847834999	0.46	2.51E-05	0.28	0.07	0.01	4.60
S3B_847835004	T/C	3B	847835004	0.46	2.51E-05	0.28	0.07	0.01	4.60

	S3A_610265762	G/A	3A	610265762	0.35	2.66E-05	-0.19	0.05	0.01	4.58
	S3B_848383175	C/T	3B	848383175	0.45	3.30E-05	0.26	0.06	0.01	4.48
	S3B_847834966	A/G	3B	847834966	0.46	3.44E-05	0.27	0.07	0.01	4.46
	S3B_847834968	T/G	3B	847834968	0.46	3.44E-05	0.27	0.07	0.01	4.46
	S3B_847834979	A/G	3B	847834979	0.46	3.44E-05	0.27	0.07	0.01	4.46
	S2B_8698719	T/A	2B	8698719	0.18	3.47E-05	0.23	0.06	0.01	4.46
	S2B_5031268	A/G	2B	5031268	0.24	4.12E-05	0.24	0.06	0.01	4.39
	S5D_62660804	G/A	5D	62660804	0.08	4.24E-05	0.31	0.08	0.01	4.37
	S3B_839785283	C/T	3B	839785283	0.15	4.75E-05	0.25	0.06	0.01	4.32
	S2D_2896056	C/G	2D	2896056	0.21	5.27E-05	-0.19	0.05	0.01	4.28
	S2B_7430792	C/G	2B	7430792	0.23	5.45E-05	0.24	0.06	0.01	4.26
	S5D_164599835	T/C	5D	164599835	0.18	5.85E-05	0.23	0.06	0.01	4.23
	S3B_848396643	A/G	3B	848396643	0.46	9.70E-05	0.25	0.06	0.01	4.01
	S3B_839479253	A/T	3B	839479253	0.41	9.91E-05	0.28	0.07	0.01	4.00
	S3B_839479256	T/A	3B	839479256	0.41	9.91E-05	0.28	0.07	0.01	4.00
SOLID	S3B_844887153	T/C	3B	844887153	0.83	1.76E-15	1.74	0.22	0.09	14.75
	S3B_847233049	A/G	3B	847233049	0.82	1.32E-12	1.65	0.24	0.07	11.88
	S3B_844992127	G/A	3B	844992127	0.83	2.35E-12	1.70	0.25	0.07	11.63
	S3B_844846869	G/A	3B	844846869	0.82	8.07E-12	1.52	0.23	0.07	11.09
	S3B_847834966	A/G	3B	847834966	0.16	7.47E-11	-1.54	0.24	0.06	10.13
	S3B_847834968	T/G	3B	847834968	0.16	7.47E-11	-1.54	0.24	0.06	10.13
	S3B_847834979	A/G	3B	847834979	0.16	7.47E-11	-1.54	0.24	0.06	10.13
	S3B_847834999	T/C	3B	847834999	0.16	7.47E-11	-1.54	0.24	0.06	10.13
	S3B_847835004	T/C	3B	847835004	0.16	7.47E-11	-1.54	0.24	0.06	10.13
	S3B_850180824	A/G	3B	850180824	0.17	1.33E-10	-1.44	0.23	0.06	9.88
	S3B_844846941	G/A	3B	844846941	0.83	1.64E-10	1.39	0.22	0.06	9.79
	S3B_848383175	C/T	3B	848383175	0.15	2.12E-10	-1.49	0.24	0.06	9.67
	S3B_847463798	T/A	3B	847463798	0.8	3.09E-10	1.44	0.23	0.06	9.51
	S3B_848393048	A/G	3B	848393048	0.17	5.88E-10	-1.39	0.23	0.06	9.23

S3B_848369446	T/C	3B	848369446	0.17	6.04E-10	-1.40	0.23	0.06	9.22
S3B_847230245	G/A	3B	847230245	0.14	8.87E-10	-1.49	0.25	0.06	9.05
S3B_847925770	G/C	3B	847925770	0.04	3.17E-09	-2.16	0.37	0.05	8.50
S3B_847391191	T/C	3B	847391191	0.79	3.53E-09	1.26	0.22	0.05	8.45
S3B_847656942	G/A	3B	847656942	0.82	3.88E-09	1.36	0.23	0.05	8.41
S3B_847656944	G/A	3B	847656944	0.82	3.88E-09	1.36	0.23	0.05	8.41
S3B_848396643	A/G	3B	848396643	0.17	4.37E-09	-1.33	0.23	0.05	8.36
S3B_848648204	T/C	3B	848648204	0.82	4.61E-09	1.32	0.23	0.05	8.34
S3B_848059135	A/G	3B	848059135	0.14	7.69E-09	-1.39	0.24	0.05	8.11
S3B_847454154	C/T	3B	847454154	0.82	8.29E-09	1.34	0.23	0.05	8.08
S3B_850642529	C/G	3B	850642529	0.18	8.45E-09	-1.28	0.23	0.05	8.07
S3B_850711969	G/A	3B	850711969	0.2	3.19E-08	-1.14	0.21	0.05	7.50
S3B_845037868	T/C	3B	845037868	0.14	4.19E-08	-1.31	0.24	0.04	7.38
S3B_851413871	C/G	3B	851413871	0.17	5.27E-08	-1.22	0.23	0.04	7.28
S3B_851413872	T/A	3B	851413872	0.17	5.27E-08	-1.22	0.23	0.04	7.28
S3B_851413875	T/A	3B	851413875	0.17	5.27E-08	-1.22	0.23	0.04	7.28
S3B_850973066	C/T	3B	850973066	0.03	1.04E-07	-2.16	0.41	0.04	6.98
S3B_850636960	G/T	3B	850636960	0.81	2.12E-07	1.14	0.22	0.04	6.67
S3B_850636975	T/C	3B	850636975	0.81	2.12E-07	1.14	0.22	0.04	6.67
S3B_839479253	A/T	3B	839479253	0.14	2.55E-07	-1.29	0.25	0.04	6.59
S3B_839479256	T/A	3B	839479256	0.14	2.55E-07	-1.29	0.25	0.04	6.59
S3B_851847253	C/A	3B	851847253	0.03	5.86E-07	-1.99	0.40	0.04	6.23
S3B_851847276	G/A	3B	851847276	0.03	5.86E-07	-1.99	0.40	0.04	6.23
S3B_851847281	G/C	3B	851847281	0.03	5.86E-07	-1.99	0.40	0.04	6.23
S3B_851847299	T/G	3B	851847299	0.03	5.86E-07	-1.99	0.40	0.04	6.23
S3B_847705997	C/T	3B	847705997	0.18	9.49E-07	-1.12	0.23	0.04	6.02
S3B_847706024	T/G	3B	847706024	0.18	9.49E-07	-1.12	0.23	0.04	6.02
S3B_847706025	G/T	3B	847706025	0.18	9.49E-07	-1.12	0.23	0.04	6.02
S3B_847706029	T/C	3B	847706029	0.18	9.49E-07	-1.12	0.23	0.04	6.02

S3B_847706036	A/G	3B	847706036	0.18	9.49E-07	-1.12	0.23	0.04	6.02
S3B_847706037	T/A	3B	847706037	0.18	9.49E-07	-1.12	0.23	0.04	6.02
S3B_847706040	C/G	3B	847706040	0.18	9.49E-07	-1.12	0.23	0.04	6.02
S3B_847706042	C/T	3B	847706042	0.18	9.49E-07	-1.12	0.23	0.04	6.02
S3B_847706051	C/T	3B	847706051	0.18	9.49E-07	-1.12	0.23	0.04	6.02
S3B_847706053	C/T	3B	847706053	0.18	9.49E-07	-1.12	0.23	0.04	6.02
S3B_850964363	A/G	3B	850964363	0.03	1.98E-06	-1.93	0.41	0.03	5.70
S3B_848375850	G/C	3B	848375850	0.03	2.01E-06	-1.77	0.38	0.03	5.70
S3B_837057835	A/C	3B	837057835	0.15	2.28E-06	-1.16	0.25	0.03	5.64
S3B_850642585	A/G	3B	850642585	0.03	3.76E-06	-1.84	0.40	0.03	5.42
S3B_839904446	T/C	3B	839904446	0.16	4.11E-06	-1.08	0.24	0.03	5.39
S3B_839904461	G/C	3B	839904461	0.16	4.11E-06	-1.08	0.24	0.03	5.39
S3B_800990989	G/A	3B	800990989	0.13	4.50E-06	0.97	0.21	0.03	5.35
S3B_800991000	A/G	3B	800991000	0.13	4.50E-06	0.97	0.21	0.03	5.35
S3B_851840380	C/G	3B	851840380	0.03	4.88E-06	-1.85	0.41	0.03	5.31
S3B_851840409	C/A	3B	851840409	0.03	4.88E-06	-1.85	0.41	0.03	5.31
S3B_851840410	C/G	3B	851840410	0.03	4.88E-06	-1.85	0.41	0.03	5.31
S3B_848603170	G/A	3B	848603170	0.14	8.24E-06	-1.11	0.25	0.03	5.08
S3B_837389551	A/G	3B	837389551	0.15	9.97E-06	-1.09	0.25	0.03	5.00
S3B_837389556	T/C	3B	837389556	0.15	9.97E-06	-1.09	0.25	0.03	5.00
S3B_837389598	A/T	3B	837389598	0.15	9.97E-06	-1.09	0.25	0.03	5.00
S3B_837389602	T/A	3B	837389602	0.15	9.97E-06	-1.09	0.25	0.03	5.00
S3B_837389603	G/C	3B	837389603	0.15	9.97E-06	-1.09	0.25	0.03	5.00
S3B_837389604	T/A	3B	837389604	0.15	9.97E-06	-1.09	0.25	0.03	5.00
S3B_837389605	A/G	3B	837389605	0.15	9.97E-06	-1.09	0.25	0.03	5.00
S3B_837389607	A/C	3B	837389607	0.15	9.97E-06	-1.09	0.25	0.03	5.00
S3B_837389608	A/C	3B	837389608	0.15	9.97E-06	-1.09	0.25	0.03	5.00
S3B_839317555	G/A	3B	839317555	0.15	1.06E-05	-1.10	0.25	0.03	4.97
S3B_839386837	T/G	3B	839386837	0.15	1.07E-05	-1.04	0.24	0.03	4.97

C1	S3B_839796863	A/G	3B	839796863	0.14	1.31E-05	-1.05	0.24	0.03	4.88
	S3B_839796868	G/A	3B	839796868	0.14	1.31E-05	-1.05	0.24	0.03	4.88
	S3B_850784464	C/T	3B	850784464	0.03	1.44E-05	-1.82	0.42	0.03	4.84
	S3B_850784509	G/A	3B	850784509	0.03	1.44E-05	-1.82	0.42	0.03	4.84
	S3B_848375841	T/A	3B	848375841	0.14	2.05E-05	-1.07	0.25	0.03	4.69
	S3B_850642525	A/T	3B	850642525	0.03	2.06E-05	-1.71	0.40	0.03	4.69
	S3B_850642558	C/A	3B	850642558	0.03	2.06E-05	-1.71	0.40	0.03	4.69
	S3B_850642559	G/T	3B	850642559	0.03	2.06E-05	-1.71	0.40	0.03	4.69
	S3D_614639836	G/T	3D	614639836	0.04	2.58E-05	-1.53	0.37	0.03	4.59
	S3B_848369476	A/T	3B	848369476	0.03	3.37E-05	-1.61	0.39	0.03	4.47
	S3B_848396638	T/G	3B	848396638	0.14	4.14E-05	-0.99	0.24	0.03	4.38
	S3B_850711935	G/A	3B	850711935	0.04	6.65E-05	-1.51	0.38	0.02	4.18
	S3A_745010852	G/C	3A	745010852	0.14	6.81E-05	-1.02	0.26	0.02	4.17
	S3A_745010858	T/C	3A	745010858	0.14	6.81E-05	-1.02	0.26	0.02	4.17
	S3A_745010866	C/G	3A	745010866	0.14	8.79E-05	-1.01	0.26	0.02	4.06
	S3A_745010873	T/C	3A	745010873	0.14	8.79E-05	-1.01	0.26	0.02	4.06
	S3B_848060000	G/A	3B	848060000	0.03	9.18E-05	-1.78	0.46	0.02	4.04
	S3B_841019923	A/G	3B	841019923	0.15	9.63E-05	-0.97	0.25	0.02	4.02
	S2D_478932518	A/C	2D	478932518	0.03	9.67E-05	1.84	0.47	0.02	4.01
	S2B_5645271	T/C	2B	5645271	0.24	5.13E-06	0.25	0.06	0.01	5.29
	S2B_6547127	T/C	2B	6547127	0.21	6.35E-06	0.25	0.06	0.01	5.20
	S2B_6547137	A/C	2B	6547137	0.21	6.35E-06	0.25	0.06	0.01	5.20
	S2B_6547149	C/A	2B	6547149	0.21	6.35E-06	0.25	0.06	0.01	5.20
	S3A_610265762	G/A	3A	610265762	0.35	1.11E-05	-0.19	0.04	0.01	4.95
	S5D_62660804	G/A	5D	62660804	0.08	1.77E-05	0.31	0.07	0.01	4.75
	S2B_7410820	A/C	2B	7410820	0.22	2.31E-05	0.23	0.06	0.01	4.64
	S2B_8698719	T/A	2B	8698719	0.18	2.34E-05	0.23	0.05	0.01	4.63
	S3B_839904446	T/C	3B	839904446	0.44	2.93E-05	0.33	0.08	0.01	4.53
	S3B_839904461	G/C	3B	839904461	0.44	2.93E-05	0.33	0.08	0.01	4.53

	S5D_164599835	T/C	5D	164599835	0.18	3.47E-05	0.23	0.05	0.01	4.46
	S5D_76166207	A/G	5D	76166207	0.42	3.65E-05	0.19	0.05	0.01	4.44
	S2D_2896056	C/G	2D	2896056	0.21	4.03E-05	-0.19	0.05	0.01	4.39
	S5D_79468779	T/C	5D	79468779	0.41	4.93E-05	0.18	0.05	0.01	4.31
	S2B_5031268	A/G	2B	5031268	0.24	5.07E-05	0.23	0.06	0.01	4.30
	S5A_705490182	C/T	5A	705490182	0.2	5.90E-05	-0.24	0.06	0.01	4.23
	S5D_61983925	C/T	5D	61983925	0.08	5.91E-05	0.29	0.07	0.01	4.23
	S2B_7430792	C/G	2B	7430792	0.23	6.79E-05	0.22	0.06	0.01	4.17
	S5D_83461602	A/G	5D	83461602	0.1	9.32E-05	0.25	0.06	0.01	4.03
C2	S2B_5645271	T/C	2B	5645271	0.24	5.55E-06	0.25	0.06	0.01	5.26
	S2B_6547127	T/C	2B	6547127	0.21	9.18E-06	0.25	0.06	0.01	5.04
	S2B_6547137	A/C	2B	6547137	0.21	9.18E-06	0.25	0.06	0.01	5.04
	S2B_6547149	C/A	2B	6547149	0.21	9.18E-06	0.25	0.06	0.01	5.04
	S3A_610265762	G/A	3A	610265762	0.35	1.08E-05	-0.19	0.04	0.01	4.97
	S5D_62660804	G/A	5D	62660804	0.08	1.64E-05	0.32	0.07	0.01	4.78
	S2B_8698719	T/A	2B	8698719	0.18	2.22E-05	0.23	0.05	0.01	4.65
	S5D_164599835	T/C	5D	164599835	0.18	2.45E-05	0.23	0.05	0.01	4.61
	S2B_7410820	A/C	2B	7410820	0.22	2.56E-05	0.23	0.06	0.01	4.59
	S2D_2896056	C/G	2D	2896056	0.21	2.79E-05	-0.20	0.05	0.01	4.55
	S3B_839904446	T/C	3B	839904446	0.44	3.21E-05	0.33	0.08	0.01	4.49
	S3B_839904461	G/C	3B	839904461	0.44	3.21E-05	0.33	0.08	0.01	4.49
	S5D_76166207	A/G	5D	76166207	0.42	3.76E-05	0.19	0.05	0.01	4.42
	S5D_79468779	T/C	5D	79468779	0.41	5.11E-05	0.18	0.05	0.01	4.29
	S2B_5031268	A/G	2B	5031268	0.24	5.85E-05	0.22	0.06	0.01	4.23
	S5A_705490182	C/T	5A	705490182	0.2	5.96E-05	-0.24	0.06	0.01	4.22
	S5D_61983925	C/T	5D	61983925	0.08	5.99E-05	0.29	0.07	0.01	4.22
	S2B_7430792	C/G	2B	7430792	0.23	8.40E-05	0.22	0.06	0.01	4.08
	S5D_83461602	A/G	5D	83461602	0.1	9.72E-05	0.25	0.06	0.01	4.01
C3	S2B_65833439	A/G	2B	65833439	0.27	4.82E-05	-0.29	0.07	0.01	4.32

S1	S2B_65856187	A/C	2B	65856187	0.29	6.18E-05	-0.29	0.07	0.01	4.21
	S3B_850642529	C/G	3B	850642529	0.99	7.15E-05	0.78	0.20	0.01	4.15
	S4B_480704002	C/A	4B	480704002	0.19	2.82E-05	1.08	0.26	0.03	4.55
	S2D_654117623	T/C	2D	654117623	0.3	5.29E-05	0.62	0.16	0.02	4.28
	S3B_800990989	G/A	3B	800990989	0.13	5.72E-05	0.81	0.20	0.02	4.24
S2	S3B_800991000	A/G	3B	800991000	0.13	5.72E-05	0.81	0.20	0.02	4.24
	S4B_480704002	C/A	4B	480704002	0.19	3.65E-05	1.06	0.26	0.03	4.44
	S2D_654117623	T/C	2D	654117623	0.3	3.74E-05	0.63	0.15	0.03	4.43
	S3B_800990989	G/A	3B	800990989	0.13	3.86E-05	0.82	0.20	0.03	4.41
S3	S3B_800991000	A/G	3B	800991000	0.13	3.86E-05	0.82	0.20	0.03	4.41
	S3B_800990989	G/A	3B	800990989	0.14	2.61E-05	0.90	0.22	0.03	4.58
	S3B_800991000	A/G	3B	800991000	0.14	2.61E-05	0.90	0.22	0.03	4.58
	S6A_524789877	A/G	6A	524789877	0.08	4.12E-05	1.38	0.34	0.03	4.38
	S2A_760522014	C/G	2A	760522014	0.1	6.45E-05	1.11	0.28	0.03	4.19
	S4B_480704002	C/A	4B	480704002	0.22	6.72E-05	1.18	0.30	0.03	4.17

Supplementary Table 2.5. Linkage disequilibrium (LD) r^2 calculations between markers that were significant in fixed effect and specific population (hollow stem and solid stem) GWAS (Significant GWAS SNP) and markers that were significant in the original heading date (HD) GWAS. The closer the r^2 is to 1, the more closely linked the two markers are.

Significant GWAS SNP	Significant HD SNP	r^2
S2B_5645271	S2B_60764872	0.0102
S2D_2896056	S2D_33652726 [†]	0.0124
S2D_654117623	S2D_33652726 [†]	0.0066
S3A_610265762	S3A_623350122	0.0039
S3A_745010852	S3A_623350122	0.0011
S7A_623850376	S7A_80297844	0.0096

[†] indicates markers that were not significant in the HD markers of this study, but are associated with the HD locus *Ppd-D1* which is commonly in linkage with WSS resistant MTA

Supplementary Table 2.6. Wald statistic output for significant markers from cutting score (CUT) genome-wide association studies (GWAS) for all initial significant markers and marker interactions where Markers indicates the marker or marker interaction, Df indicates the degrees of freedom in the analysis for that markers, Sum of Squares estimates variation in the model, Wald Statistic is the test statistic for the model where a larger value indicates significance, and P-Value is the probability of observing the test statistic.

Markers	Df	Sum of Squares	Wald Statistic	P-Value
S3A_610265762	2	36.86	22.69	1.18E-05
S2B_5645271	2	52.74	32.46	8.93E-08
S5D_62660804	2	38.53	23.72	7.08E-06
S5D_164599835	2	11.64	7.16	2.78E-02
S2D_2896056	2	23.49	14.46	7.26E-04
S5A_705490182	2	38.58	23.75	6.97E-06
S3B_844887153	2	115.40	71.03	3.33E-16
S3A_610265762:S2B_5645271	4	21.81	13.42	9.39E-03
S3A_610265762:S5D_62660804	4	3.53	2.17	7.04E-01
S2B_5645271:S5D_62660804	4	4.75	2.92	5.71E-01
S3A_610265762:S5D_164599835	4	1.99	1.23	8.74E-01
S2B_5645271:S5D_164599835	4	5.67	3.49	4.79E-01
S5D_62660804:S5D_164599835	3	3.23	1.99	5.74E-01
S3A_610265762:S2D_2896056	4	8.64	5.32	2.56E-01
S2B_5645271:S2D_2896056	4	1.80	1.11	8.93E-01
S5D_62660804:S2D_2896056	4	2.57	1.58	8.12E-01
S5D_164599835:S2D_2896056	4	8.61	5.30	2.58E-01
S3A_610265762:S5A_705490182	4	6.50	4.00	4.06E-01
S2B_5645271:S5A_705490182	4	16.45	10.12	3.84E-02
S5D_62660804:S5A_705490182	4	3.91	2.41	6.61E-01
S5D_164599835:S5A_705490182	4	9.84	6.06	1.95E-01
S2D_2896056:S5A_705490182	4	4.33	2.66	6.16E-01
S3A_610265762:S3B_844887153	4	16.10	9.91	4.20E-02
S2B_5645271:S3B_844887153	4	8.90	5.48	2.42E-01

S5D_62660804:S3B_844887153	4	19.75	12.16	1.62E-02
S5D_164599835:S3B_844887153	4	12.79	7.88	9.63E-02
S2D_2896056:S3B_844887153	4	8.63	5.31	2.57E-01
S5A_705490182:S3B_844887153	4	8.06	4.96	2.92E-01
S3A_610265762:S2B_5645271:S5D_62660804	7	10.00	6.15	5.22E-01
S3A_610265762:S2B_5645271:S5D_164599835	8	10.64	6.55	5.86E-01
S3A_610265762:S5D_62660804:S5D_164599835	4	5.35	3.29	5.10E-01
S2B_5645271:S5D_62660804:S5D_164599835	4	12.04	7.41	1.16E-01
S3A_610265762:S2B_5645271:S2D_2896056	8	19.28	11.87	1.57E-01
S3A_610265762:S5D_62660804:S2D_2896056	5	6.43	3.96	5.56E-01
S2B_5645271:S5D_62660804:S2D_2896056	5	14.16	8.71	1.21E-01
S3A_610265762:S5D_164599835:S2D_2896056	6	10.59	6.52	3.68E-01
S2B_5645271:S5D_164599835:S2D_2896056	5	12.87	7.92	1.61E-01
S5D_62660804:S5D_164599835:S2D_2896056	5	3.44	2.11	8.33E-01
S3A_610265762:S2B_5645271:S5A_705490182	8	13.87	8.54	3.83E-01
S3A_610265762:S5D_62660804:S5A_705490182	3	5.67	3.49	3.22E-01
S2B_5645271:S5D_62660804:S5A_705490182	4	4.65	2.86	5.81E-01
S3A_610265762:S5D_164599835:S5A_705490182	6	12.77	7.86	2.48E-01
S2B_5645271:S5D_164599835:S5A_705490182	7	19.31	11.89	1.04E-01
S5D_62660804:S5D_164599835:S5A_705490182	3	3.11	1.92	5.90E-01
S3A_610265762:S2D_2896056:S5A_705490182	8	3.40	2.09	9.78E-01
S2B_5645271:S2D_2896056:S5A_705490182	8	3.89	2.40	9.66E-01
S5D_62660804:S2D_2896056:S5A_705490182	2	1.27	0.78	6.77E-01
S5D_164599835:S2D_2896056:S5A_705490182	4	5.69	3.50	4.78E-01
S3A_610265762:S2B_5645271:S3B_844887153	8	12.06	7.42	4.92E-01
S3A_610265762:S5D_62660804:S3B_844887153	5	13.73	8.45	1.33E-01
S2B_5645271:S5D_62660804:S3B_844887153	4	14.92	9.18	5.67E-02
S3A_610265762:S5D_164599835:S3B_844887153	6	14.55	8.95	1.76E-01
S2B_5645271:S5D_164599835:S3B_844887153	4	10.66	6.56	1.61E-01

S5D_62660804:S5D_164599835:S3B_844887153	3	6.13	3.77	2.87E-01
S3A_610265762:S2D_2896056:S3B_844887153	8	20.39	12.55	1.28E-01
S2B_5645271:S2D_2896056:S3B_844887153	6	10.10	6.22	3.99E-01
S5D_62660804:S2D_2896056:S3B_844887153	1	6.07	3.74	5.33E-02
S5D_164599835:S2D_2896056:S3B_844887153	2	1.42	0.87	6.47E-01
S3A_610265762:S5A_705490182:S3B_844887153	8	16.31	10.04	2.62E-01
S2B_5645271:S5A_705490182:S3B_844887153	7	12.27	7.55	3.74E-01
S5D_62660804:S5A_705490182:S3B_844887153	3	4.80	2.96	3.99E-01
S5D_164599835:S5A_705490182:S3B_844887153	4	3.88	2.39	6.65E-01
S2D_2896056:S5A_705490182:S3B_844887153	6	4.19	2.58	8.59E-01
S3A_610265762:S2B_5645271:S5D_164599835: S2D_2896056	1	14.91	9.17	2.45E-03
S3A_610265762:S2B_5645271:S5D_164599835: S5A_705490182	2	0.90	0.55	7.58E-01
S3A_610265762:S2B_5645271:S2D_2896056: S5A_705490182	4	6.88	4.23	3.75E-01
S3A_610265762:S5D_62660804:S2D_2896056: S5A_705490182	1	3.79	2.34	1.26E-01
S3A_610265762:S5D_164599835:S2D_2896056: S5A_705490182	1	0.12	0.08	7.82E-01
S2B_5645271:S5D_164599835:S2D_2896056: S5A_705490182	1	0.00	0.00	9.62E-01
S3A_610265762:S2B_5645271:S5D_62660804: S3B_844887153	1	0.23	0.14	7.09E-01
S3A_610265762:S2B_5645271:S5D_164599835: S3B_844887153	3	11.86	7.30	6.29E-02
S3A_610265762:S5D_62660804:S5D_164599835: S3B_844887153	1	1.46	0.90	3.43E-01
S3A_610265762:S2B_5645271:S2D_2896056: S3B_844887153	6	3.19	1.96	9.23E-01
S3A_610265762:S5D_164599835:S2D_2896056: S3B_844887153	1	0.35	0.21	6.43E-01

S2B_5645271:S5D_164599835:S2D_2896056: S3B_844887153	1	1.71	1.05	3.05E-01
S3A_610265762:S2B_5645271:S5A_705490182: S3B_844887153	2	10.36	6.37	4.13E-02
S3A_610265762:S5D_62660804:S5A_705490182: S3B_844887153	1	0.92	0.57	4.51E-01
S3A_610265762:S5D_164599835:S5A_705490182: S3B_844887153	1	2.47	1.52	2.18E-01
S2B_5645271:S5D_164599835:S5A_705490182: S3B_844887153	1	0.28	0.17	6.80E-01
S3A_610265762:S2D_2896056:S5A_705490182: S3B_844887153	3	2.41	1.48	6.86E-01
S2B_5645271:S2D_2896056:S5A_705490182: S3B_844887153	2	2.29	1.41	4.94E-01
S5D_164599835:S2D_2896056:S5A_705490182: S3B_844887153	1	3.02	1.86	1.73E-01
S3A_610265762:S2B_5645271:S5D_164599835: S5A_705490182:S3B_844887153	1	2.72	1.68	1.96E-01
S3A_610265762:S2B_5645271:S2D_2896056: S5A_705490182:S3B_844887153	1	0.00	0.00	9.66E-01

Supplementary Table 2.7. Wald statistic output for significant markers from cutting score (CUT) and solidity (SOLID) genome-wide association studies (GWAS) used for allelic combination (AC) groups where Trait indicates which GWAS models the markers were significant in (CUT or SOLID), Marker indicates the significant marker, Df indicates the degrees of freedom in the analysis for that markers, Sum of Squares estimates variation in the model, Wald Statistic is the test statistic for the model where a larger value indicates significance, and P-Value is the probability of observing the test statistic.

Trait	Marker	Df	Sum of Squares	Wald Statistic	P-Value
CUT	S3A_610265762	2	35.49	21.64	2.00E-05
	S2B_5645271	2	49.42	30.14	2.85E-07
	S5D_62660804	2	36.39	22.19	1.52E-05
	S5A_705490182	2	37.51	22.87	1.08E-05
	S3B_844887153	2	105.67	64.45	1.01E-14
SOLID	S4B_480704002	2	90.68	18.39	1.02E-04
	S3B_844887153	2	426.68	86.52	0.00E+00
	S2D_654117623	2	103.93	21.08	2.65E-05

Supplementary Table 2.8. Allelic combinations created from the cutting score (CUT) significant SNPs where Trait is the trait the ACs are predicting, S3B_844887153, S2B_564527, S3A_610265762, S5A_705490182, and S5D_62660804 are the markers in the AC with a susceptible (S) or resistant (R) allele, where S3B_844887153 represents the *SSII* locus, N is the number of individuals with that AC with measurements for the trait of interest, Trait Mean is the mean of the trait of interest across the individuals in the AC group, Trait SE is the standard error (SE) for the trait in the AC group, AC Frequency is the frequency amongst the population, Predicted Value is the estimated value of the trait for the given AC and trait, and Predicted SE is the standard error for the predicted trait value.

Trait	S3B_844887153	S2B_564527	S3A_610265762	S5A_705490182	S5D_62660804	N	Trait Mean	Trait SE	AC Freq.	Predicted Value	Predicted SE
CUT	S	R	S	S	R	56	1.89	0.18	0.03	2.35	0.21
	R	R	S	S	R	3	2.86	0.93	0	3.7	0.28
	R	R	R	S	R	258	2.83	0.09	0.13	3.14	0.12
	S	R	R	S	R	11	4.82	0.51	0.01	4.39	0.13
	S	S	S	S	R	72	2.87	0.16	0.04	3.08	0.81
	S	R	S	R	R	2	3.62	1.32	0	3.46	0.98
	R	R	S	R	R	348	3.6	0.09	0.18	5.22	0.46
	S	S	R	S	R	12	4.61	0.45	0.01	4.58	0.33
	R	S	S	S	R	32	2.59	0.3	0.02	3.46	0.28
	R	R	R	R	R	1	6.93	NA	0	5.43	0.4
	R	S	S	R	R	37	3.08	0.3	0.02	3.59	0.25
	S	R	S	S	S	49	2.84	0.23	0.02	4.94	0.21
	R	S	R	S	R	56	3.43	0.27	0.03	6.96	1.39
	R	S	R	R	R	2	5.06	0.88	0	6.11	1.4
	S	R	R	R	R	28	4.06	0.3	0.01	5.17	0.66
	S	R	R	S	S	2	3.81	1.94	0	3.18	0.19
	S	S	R	R	R	147	4.67	0.13	0.07	4.33	0.2
	S	R	S	R	S	23	4.9	0.28	0.01	3.78	0.1
	S	S	S	R	R	72	4.59	0.16	0.04	4.62	0.09
	S	S	S	S	S	16	4.62	0.32	0.01	3.74	0.98
	R	R	S	S	S	461	5.04	0.06	0.23	4.2	0.39

	R	R	R	S	S	45	5.86	0.17	0.02	4.92	0.41
	S	S	R	S	S	19	5.43	0.32	0.01	5.27	0.24
	R	R	R	R	S	1	5.99	NA	0	3.79	0.25
	S	R	R	R	S	63	5.23	0.21	0.03	4.63	0.39
	R	S	S	S	S	5	5.45	0.99	0	3.86	0.21
	R	R	S	R	S	15	4.64	0.41	0.01	5.09	0.16
	S	S	R	R	S	134	5.41	0.13	0.07	5.41	1
	R	S	R	R	S	14	6.09	0.39	0.01	5.76	0.4
SOLID	R	R	R	S	R	36	11.72	0.54	0.07	11.09	0.63
	R	R	S	S	R	1	11.24	NA	0	12.71	2.86
	R	R	R	R	R	159	11.7	0.25	0.31	11.53	0.41
	S	R	S	S	R	8	9.93	0.86	0.02	11.19	1.27
	R	S	R	S	R	14	11.02	0.76	0.03	10.6	0.87
	S	R	R	S	R	1	12.16	NA	0	11.33	2.83
	R	S	R	R	R	127	11.11	0.26	0.25	11.45	0.4
	R	S	S	R	R	2	7.09	1.19	0	9.05	1.99
	R	S	S	S	R	21	15.54	0.84	0.04	13.08	0.81
	R	R	S	R	R	23	12.95	0.82	0.04	11.04	0.72
	S	R	R	S	S	20	13.02	0.55	0.04	11.44	0.98
	R	R	R	S	S	19	12.35	0.83	0.04	12.14	0.74
	S	R	S	R	R	3	6.49	0.12	0.01	6.71	2.12
	S	S	S	S	R	22	7.25	0.85	0.04	7.78	0.68
	S	S	R	R	R	9	5.52	0.29	0.02	6.87	1.3
	R	S	S	S	S	6	6.36	0.28	0.01	6.62	1.53
	R	R	S	S	S	34	6.66	0.43	0.07	8	0.58
	S	S	R	S	R	3	7.48	0.59	0.01	6.08	1.73
	R	R	S	R	S	1	5.18	NA	0	3.06	2.92
	R	R	R	R	S	4	7.41	0.52	0.01	6.72	1.53

Supplementary Table 2.9. Wald statistic output for significant markers from solidity (SOLID) genome-wide association studies (GWAS) for all initial significant markers and marker interactions where Markers indicates the marker or marker interaction, Df indicates the degrees of freedom in the analysis for that markers, Sum of Squares estimates variation in the model, Wald Statistic is the test statistic for the model where a larger value indicates significance, and P-Value is the probability of observing the test statistic.

Markers	Df	Sum of Squares	Wald Statistic	P-Value
S4B_480704002	2	89.8364744	18.2097289	0.00011112
S3B_844887153	2	421.984344	85.5356418	0
S2D_654117623	2	103.139107	20.906154	2.89E-05
S4B_480704002:S3B_844887153	4	16.3938595	3.3230126	0.50529439
S4B_480704002:S2D_654117623	4	25.0598385	5.07959454	0.27922767
S3B_844887153:S2D_654117623	4	3.56762219	0.72315207	0.94844341
S4B_480704002:S3B_844887153:S2D_654117623	1	1.0033664	0.20338098	0.65200555

Supplementary Table 2.10. Allelic combinations created from the cutting score (SOLID) significant SNPs where Trait is the trait the ACs are predicting, S3B_844887153, S2D_654117623, and S4B_480704002 are the markers in the AC with susceptible (S) or resistant (R) alleles where S3B_844887153 represents the *SSII* locus, N is the number of individuals with that AC with measurements for the trait of interest, Trait Mean is the mean of the trait of interest across the individuals in the AC group, Trait SE is the standard error (SE) for the trait in the AC group, AC Frequency is the frequency amongst the population, Predicted Value is the estimated value of the trait for the given AC and trait, and Predicted SE is the standard error for the predicted trait value.

Trait	S3B_844887153	S2D_654117623	S4B_480704002	N	Trait Mean	Trait SE	AC Freq.	Predicted Value	Predicted SE
CUT	R	R	R	40	2.27	0.23	0.02	3.16	0.28
	R	R	S	201	2.81	0.11	0.09	3.35	0.13
	R	S	R	120	2.53	0.14	0.06	3.24	0.21
	R	S	S	635	3.42	0.07	0.3	3.67	0.08
	S	R	R	21	4.99	0.46	0.01	4.9	0.34
	S	R	S	143	5.05	0.13	0.07	4.67	0.14
	S	S	R	82	4.98	0.18	0.04	4.71	0.19
	S	S	S	886	5.02	0.05	0.42	4.62	0.08
SOLID	R	R	R	22	15.14	0.96	0.04	14.66	0.78
	R	R	S	109	11.61	0.33	0.2	11.98	0.41
	R	S	R	77	12.9	0.33	0.14	12.91	0.59
	R	S	S	249	11.03	0.19	0.46	10.77	0.32
	S	R	R	1	9.11	NA	0	8.07	2.8
	S	R	S	19	7.16	0.77	0.04	8.19	0.69
	S	S	R	7	8.65	1.19	0.01	8.15	1.11
	S	S	S	52	6.3	0.31	0.1	7.27	0.48

Supplementary Table 2.11. A table of references for other publications that identified QTL on the same chromosomes as this study. Marker and chromosome indicate the locus identified in the study, Trait indicates if the marker was identified in the CUT or SOLID models, and Reference indicates if loci on the same chromosome were identified in previous literature or if it was novel to this study.

Marker	Chromosome	Trait	Reference
S2A_760522014	2A	SOLID	Varella et al. 2015, Nilsen et al. 2017
S2B_5645271	2B	CUT	Ávila et al. 2021, Varella et al. 2019
S2B_65833439	2B	CUT	Ávila et al. 2021, Varella et al. 2019
S2D_2896056	2D	CUT	Sherman et al. 2010, Nilsen et al. 2017, Hai et al. 2005, Dreccer et al. 2022
S2D_654117623	2D	SOLID	Sherman et al. 2010, Nilsen et al. 2017, Hai et al. 2005, Dreccer et al. 2022
S3A_610265762	3A	CUT	Ávila et al. 2021, Varella et al. 2015, Varella et al. 2019, Ren et al. 2025, Dreccer et al. 2022
S4B_480704002	4B	SOLID	Dreccer et al. 2022
S5A_705490182	5A	CUT	Nilsen et al. 2017, Ávila et al. 2021
S5D_62660804	5D	CUT	Varella et al. 2015
S5D_164599835	5D	CUT	Varella et al. 2015
S6A_524789877	6A	SOLID	Ren et al. 2025, Dreccer et al. 2022

CHAPTER III SUPPLEMENTAL

Supplemental Table 3.1. Number of phenotypic observations for each year and location for cutting score (cutting) and solidity (solidity).

Trait	Location	2014	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025
Cutting	Akron	-	-	-	-	-	397	-	649	227	1433	876
	Genoa	-	-	-	-	-	-	-	-	-	-	150
	New Raymer	144	170	296	236	207	195	150	300	395	-	309
	Orchard	86	175	393	379	291	300	246	-	-	150	502
	Roggen	-	-	-	-	-	-	96	-	-	373	350
Solidity	Akron	-	-	-	-	-	-	148	-	-	148	175
	Alliance	-	-	-	-	-	-	-	-	-	100	-
	Fort Collins	-	-	-	21	52	49	147	186	232	249	17
	Grant	-	-	-	-	-	-	-	-	58	50	-
	Mead	-	-	-	-	-	-	-	-	-	100	-
	New Raymer	-	25	132	112	80	-	-	157	-	-	-
	Orchard	-	30	128	110	80	46	-	-	-	-	116
	Sidney	-	-	-	-	-	-	-	-	66	99	-

Supplemental Table 3.2. Summary statistics for cross-validation (CV) and forward-validation (FV) for all models for cutting and solidity.

Trait	Training Population	Model	CV Min	CV Mean	CV Max	FV Min	FV Mean	FV Max
Cutting	All Environments	GBLUP	0.61	0.66	0.72	0.34	0.38	0.42
		GBLUP+SSt1	0.56	0.64	0.71	0.28	0.32	0.35
		BayesA	0.59	0.67	0.73	0.36	0.40	0.43
		BayesLASSO	0.61	0.66	0.74	0.34	0.37	0.40
		Random Forest	0.67	0.67	0.68	0.44	0.45	0.46
	TP_Q1	GBLUP	0.58	0.65	0.72	0.32	0.36	0.38
		GBLUP+SSt1	0.55	0.62	0.69	0.23	0.27	0.31
		BayesA	0.55	0.63	0.70	0.34	0.37	0.43
		bayesLASSO	0.55	0.63	0.70	0.33	0.36	0.39
		Random Forest	0.66	0.67	0.67	0.43	0.43	0.44
	TP_Q2	GBLUP	0.62	0.69	0.78	0.30	0.34	0.37
		GBLUP+SSt1	0.49	0.60	0.70	0.17	0.23	0.27
		BayesA	0.60	0.71	0.76	0.32	0.40	0.44
		bayesLASSO	0.57	0.68	0.76	0.29	0.33	0.38
		Random Forest	0.75	0.76	0.76	0.37	0.38	0.40
	TP_Q3	GBLUP	0.59	0.69	0.78	0.29	0.33	0.38
		GBLUP+SSt1	0.29	0.49	0.65	0.16	0.22	0.29
		BayesA	0.65	0.75	0.82	0.31	0.39	0.45
		bayesLASSO	0.61	0.69	0.81	0.27	0.32	0.37
		Random Forest	0.75	0.77	0.78	0.42	0.43	0.43
Solidity	All Environments	GBLUP	0.53	0.68	0.77	0.58	0.64	0.69
		GBLUP+SSt1	0.44	0.61	0.76	0.50	0.58	0.63
		BayesA	0.52	0.69	0.80	0.55	0.63	0.68
		BayesLASSO	0.51	0.67	0.81	0.56	0.61	0.66
		Random Forest	0.62	0.65	0.67	0.56	0.58	0.61
	TP_Q1	GBLUP	0.49	0.70	0.79	0.56	0.61	0.65

TP_Q2	GBLUP+SSt1	0.48	0.63	0.79	0.49	0.55	0.62
	BayesA	0.60	0.72	0.83	0.54	0.62	0.68
	bayesLASSO	0.59	0.70	0.81	0.56	0.62	0.66
	Random Forest	0.62	0.64	0.66	0.54	0.55	0.56
	GBLUP	0.61	0.77	0.89	0.54	0.58	0.63
TP_Q3	GBLUP+SSt1	0.51	0.71	0.85	0.43	0.51	0.57
	BayesA	0.55	0.79	0.91	0.51	0.59	0.66
	bayesLASSO	0.63	0.77	0.90	0.53	0.59	0.63
	Random Forest	0.70	0.72	0.75	0.55	0.56	0.57
	GBLUP	0.70	0.85	0.95	0.50	0.54	0.61
	GBLUP+SSt1	0.56	0.77	0.92	0.40	0.46	0.54
	BayesA	0.73	0.86	0.94	0.47	0.55	0.62
	bayesLASSO	0.65	0.84	0.96	0.43	0.53	0.60
	Random Forest	0.81	0.85	0.87	0.58	0.59	0.60

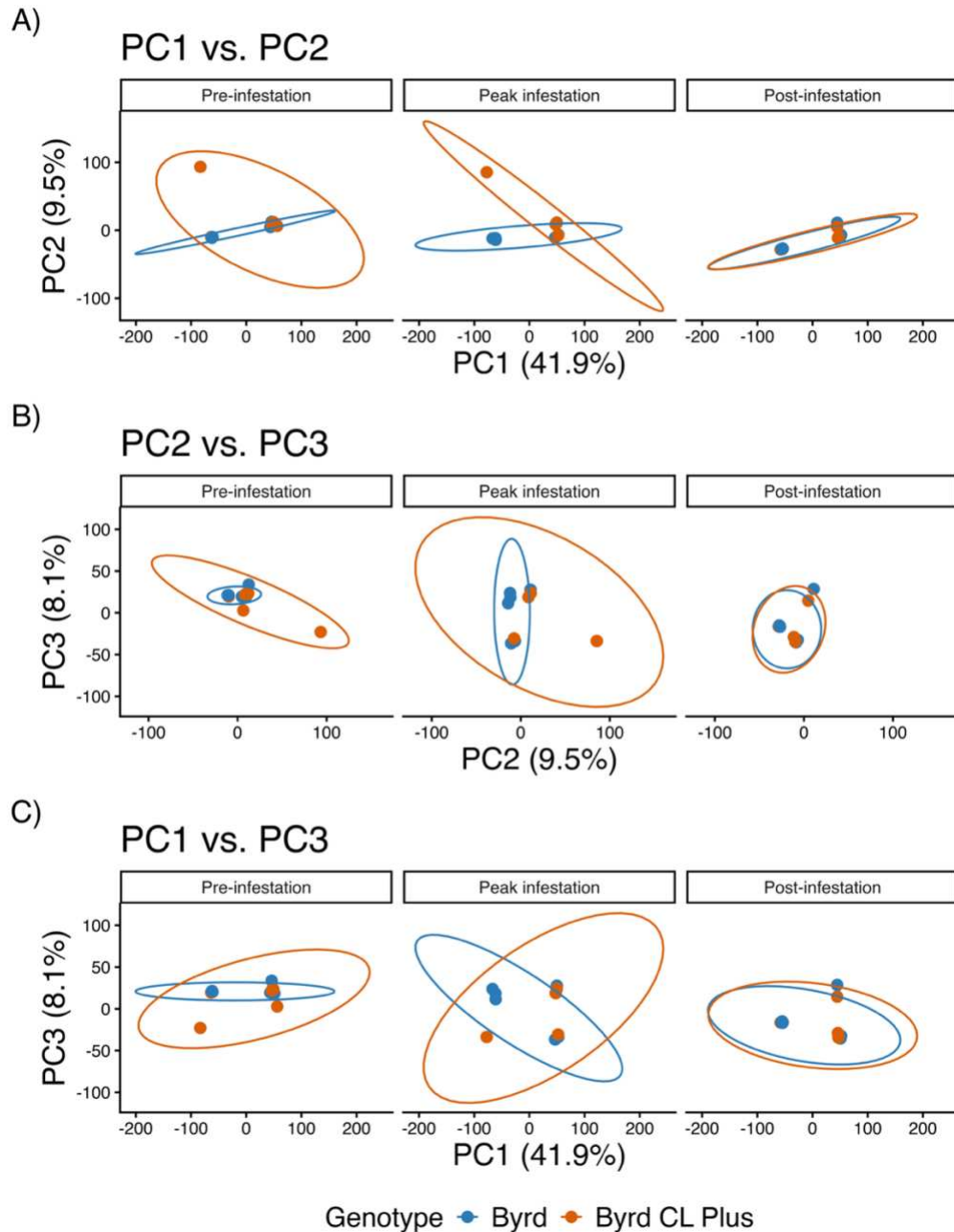
Supplemental Table 3.3. Information for training populations (TP) for all three heritability quantiles for cutting and solidity including the heritability cutoff for a TP group (h_{GBLUP}^2), the name of the TP, and the number of environments, BLUEs, and unique genotypes in each TP.

Trait	h_{GBLUP}^2	TP Name	N Environments	N BLUEs	N Genotypes
Cutting	0.38	TP_Q1 cutting	14	1468	820
	0.51	TP_Q2 cutting	10	1153	785
	0.63	TP_Q3 cutting	5	500	448
Solidity	0.27	TP_Q1 solidity	8	913	435
	0.63	TP_Q2 solidity	5	441	149
	0.76	TP_Q3 solidity	3	273	137

Supplemental Table 3.4. Statistical differences between models based on the Nemenyi test cannot be appropriately formatted as a table that fits in a word document, so they will be included in the final submission as an excel file.

CHAPTER IV SUPPLEMENTAL

Per-Timepoint PCAs of Metabolite Expression



Supplemental Figure 4.1. Principal component (PC) plots for metabolite expression separated by time point where (A) shows variance explained by PC1 and PC2, (B) shows PC2 and PC3, and (C) shows PC1 and PC3. Blue ellipses and datapoints indicate Byrd and orange ellipses and datapoints indicate Byrd CL Plus. Ellipses indicate the spread of the genotypes at different timepoints.

Supplemental Table 4.1. Fisher's test results for protein accessions across all time points including number of present or absent proteins for each genotype at each time point and the p-value of the Fisher's exact test.

Time	Accession	Year	N Present Hollow	N Present Solid	N Present Hollow	N Present Solid	p-value
t1	A0A1D5UJR8	2024	1	1	2	2	1.0
	A0A3B5ZV89	2024	2	1	1	2	1.0
	A0A3B6IRH0	2024	1	2	2	1	1.0
	A0A3B6JCT3	2024	1	3	2	0	0.4
	A0A3B6RN99	2024	2	1	1	2	1.0
	A0A3B6RPM1	2024	2	1	1	2	1.0
	A0A3B6TM69	2024	1	1	2	2	1.0
	A0A3B6U7G0	2024	3	2	0	1	1.0
	ALBU_HUMAN	2024	2	0	1	3	0.4
	A0A1D5UJR8	2025	2	1	1	2	1.0
	A0A3B5ZV89	2025	0	2	3	1	0.4
	A0A3B6IRH0	2025	0	1	3	2	1.0
	A0A3B6JCT3	2025	1	1	2	2	1.0
	A0A3B6RN99	2025	3	0	0	3	0.1
	A0A3B6RPM1	2025	3	0	0	3	0.1
	A0A3B6TM69	2025	1	3	2	0	0.4
	A0A3B6U7G0	2025	2	0	1	3	0.4
	ALBU_HUMAN	2025	1	1	2	2	1.0
t2	A0A1D5UJR8	2024	1	0	2	3	1.0
	A0A3B5ZV89	2024	2	2	1	1	1.0
	A0A3B6IRH0	2024	0	1	3	2	1.0
	A0A3B6JCT3	2024	1	1	2	2	1.0
	A0A3B6RN99	2024	0	0	3	3	1.0
	A0A3B6RPM1	2024	0	0	3	3	1.0
	A0A3B6TM69	2024	2	2	1	1	1.0

	A0A3B6U7G0	2024	2	0	1	3	0.4
	ALBU_HUMAN	2024	3	2	0	1	1.0
	A0A1D5UJR8	2025	2	1	1	2	1.0
	A0A3B5ZV89	2025	0	1	3	2	1.0
	A0A3B6IRH0	2025	1	2	2	1	1.0
	A0A3B6JCT3	2025	1	3	2	0	0.4
	A0A3B6RN99	2025	3	2	0	1	1.0
	A0A3B6RPM1	2025	3	2	0	1	1.0
	A0A3B6TM69	2025	0	1	3	2	1.0
	A0A3B6U7G0	2025	1	1	2	2	1.0
	ALBU_HUMAN	2025	1	0	2	3	1.0
t3	A0A1D5UJR8	2024	2	0	1	3	0.4
	A0A3B5ZV89	2024	0	3	3	0	0.1
	A0A3B6IRH0	2024	0	0	3	3	1.0
	A0A3B6JCT3	2024	2	1	1	2	1.0
	A0A3B6RN99	2024	0	0	3	3	1.0
	A0A3B6RPM1	2024	2	1	1	2	1.0
	A0A3B6TM69	2024	0	3	3	0	0.1
	A0A3B6U7G0	2024	3	1	0	2	0.4
	ALBU_HUMAN	2024	3	1	0	2	0.4
	A0A1D5UJR8	2025	3	0	0	3	0.1
	A0A3B5ZV89	2025	1	0	2	3	1.0
	A0A3B6IRH0	2025	2	3	1	0	1.0
	A0A3B6JCT3	2025	0	2	3	1	0.4
	A0A3B6RN99	2025	3	2	0	1	1.0
	A0A3B6RPM1	2025	2	1	1	2	1.0
	A0A3B6TM69	2025	2	3	1	0	1.0
	A0A3B6U7G0	2025	3	1	0	2	0.4

ALBU HUMAN	2025	0	0	3	3	1.0
------------	------	---	---	---	---	-----

Supplemental Table 4.2. Wald table results for effects of genotype (ID), year (YR) and genotype by year (ID:YR) on heading date in Byrd and Byrd CL Plus.

Variable	Df	Sum of Sq	Wald statistic	P-value
(Intercept)	1	207069.51	338861.72	0.00E+00
ID	1	4.90	8.02	4.63E-03
YR	1	52.27	85.53	0.00E+00
ID:YR	1	0.07	0.11	7.41E-01
residual				
(MS)	NA	0.61	NA	NA

Supplemental Table 4.3. Positive and negative correlations for significant unannotated metabolites ($\text{Log2FC} \geq |5|$, $\text{p-value}_{\text{unadjusted}} < 0.05$) identified at all timepoints. Correlations (r) indicates the correlation between the unannotated and annotated metabolites, $\text{p-value}_{\text{FDR}}$ indicates the adjusted significance of the correlation p-value, and the pathways, superclasses, and classes associated with the annotated metabolites.

Unannotated Metabolite	Time	r	p-value _{FDR}	Correlated Annotated Pathway	Correlated Annotated Superclass	Correlated Annotated Class
C4755	1	0.75	1.22E-04	Alkaloids	Pseudoalkaloids (transamidation)	Capsaicins and Capsaicinoids
		-0.70	4.35E-04	Alkaloids	Pseudoalkaloids (transamidation)	Open-chain polyketides
		0.74	1.57E-04	Alkaloids	Pseudoalkaloids (transamidation)	Phenylalanine-derived alkaloids
		-0.73	2.03E-04	Amino acids and Peptides	Oligopeptides	Cyclic peptides
		0.76	7.88E-05	Amino acids and Peptides	Oligopeptides	Linear peptides
		0.74	1.86E-04	Amino acids and Peptides	Oligopeptides	Microcolins and mirabimids
		-0.75	1.22E-04	Amino acids and Peptides	Small peptides	Dipeptides
		0.77	6.00E-05	Amino acids and Peptides	Small peptides	Pyridine alkaloids
		-0.83	5.54E-06	Amino acids and Peptides	Small peptides	Tripeptides
		-0.77	6.86E-05	Carbohydrates	Nucleosides	Purine nucleosides
		-0.74	1.57E-04	Fatty acids	Fatty esters	Wax diesters
		0.71	3.99E-04	Fatty acids	Glycerolipids	Triacylglycerols
		-0.71	4.17E-04	Fatty acids	Linear polyketides	Open-chain polyketides
		0.84	3.56E-06	Polyketides	Macrolides	Macrolide lactones
		-0.75	1.25E-04	Polyketides	Polycyclic aromatic polyketides	Bisnaphthalenes
		0.71	3.53E-04	Shikimates and Phenylpropanoids	Flavonoids	Flavones
		-0.75	1.18E-04	Shikimates and Phenylpropanoids	Flavonoids	Flavones

		0.72	2.70E-04	Shikimates and Phenylpropanoids	Flavonoids	Flavonols
		0.76	9.13E-05	Shikimates and Phenylpropanoids	Flavonoids	Flavonols
		-0.74	1.86E-04	Shikimates and Phenylpropanoids	Lignans	Furofuranoid lignans
		0.81	1.56E-05	Shikimates and Phenylpropanoids	Phenolic acids (C6-C1)	Simple phenolic acids
		0.75	1.25E-04	Terpenoids	Carotenoids (C40)	Carotenoids (C40, α - β -carotene)
		0.76	8.92E-05	Terpenoids	Carotenoids (C40)	Carotenoids (C40, α - β -carotene)
		-0.73	2.49E-04	Terpenoids	Monoterpenoids	Secoiridoid monoterpenoids
		-0.76	7.88E-05	Terpenoids	Triterpenoids	Dammarane and Protostane triterpenoids
C5113	1	-0.70	4.86E-04	Amino acids and Peptides	Small peptides	Aminoacids
		-0.71	3.23E-04	Carbohydrates	Nucleosides	Purine nucleosides
		-0.70	4.73E-04	Shikimates and Phenylpropanoids	Lignans	Neolignans
C5600	1	-0.85	3.56E-06	Shikimates and Phenylpropanoids	Tropolones	Tropolones and derivatives (Shikimate)
C6396	1	-0.71	3.57E-04	Carbohydrates	Nucleosides	Purine alkaloids
		-0.72	3.11E-04	Fatty acids	Linear polyketides	Open-chain polyketides
C6882	1	-0.77	6.86E-05	Shikimates and Phenylpropanoids	Coumarins	Triketide meroterpenoids
C7125	1	-0.74	1.61E-04	Carbohydrates	Nucleosides	Purine alkaloids
		0.72	3.04E-04	Fatty acids	Glycerolipids	Triacylglycerols
		-0.73	2.51E-04	Polyketides	Naphthalenes	Flavones

		0.71	4.17E-04	Shikimates and Phenylpropanoids	Flavonoids	Flavonols
		-0.70	4.67E-04	Alkaloids	Ornithine alkaloids	Polyamines
		-0.75	1.15E-04	Alkaloids	Pseudoalkaloids (transamidation)	Open-chain polyketides
		0.72	3.05E-04	Alkaloids	Pseudoalkaloids (transamidation)	Phenylalanine- derived alkaloids
		-0.78	4.44E-05	Amino acids and Peptides	Oligopeptides	Cyclic peptides
		0.76	7.88E-05	Amino acids and Peptides	Oligopeptides	Linear peptides
		-0.78	4.10E-05	Carbohydrates	Nucleosides	Purine nucleosides
C7261	1	-0.76	9.52E-05	Fatty acids	Fatty acyls	Fatty alcohols
		-0.75	1.15E-04	Fatty acids	Fatty esters	Wax diesters
		-0.73	2.12E-04	Fatty acids	Fatty esters	Wax monoesters
		-0.80	2.37E-05	Fatty acids	Sesquiterpenoids	Dicarboxylic acids
		0.72	3.05E-04	Polyketides	Macrolides	Macrolide lactones
		0.71	3.67E-04	Shikimates and Phenylpropanoids	Flavonoids	Flavonols
		-0.71	4.16E-04	Shikimates and Phenylpropanoids	Lignans	Arylnaphthalene and aryltetralin lignans
		-0.71	4.01E-04	Terpenoids	Triterpenoids	Quassinoids
C7524	1	-0.74	1.38E-04	Carbohydrates	Nucleosides	Purine alkaloids
C0643	2	0.87	1.73E-06	Polyketides	Naphthalenes	Bisnaphthalenes
C0671	2	-0.72	3.11E-04	Terpenoids	Sesquiterpenoids	Hydrocarbons
		-0.75	1.25E-04	Terpenoids	Triterpenoids	Quassinoids
C1576	2	0.86	1.88E-06	Polyketides	Naphthalenes	Bisnaphthalenes
C3110	2	0.72	3.05E-04	Shikimates and Phenylpropanoids	Flavonoids	Flavones
		-0.70	4.39E-04	Alkaloids	Ornithine alkaloids	Polyamines
C3899	2	-0.72	2.64E-04	Shikimates and Phenylpropanoids	Flavonoids	Flavanones

		-0.72	3.17E-04	Shikimates and Phenylpropanoids	Lignans	Furofuranoid lignans
C5031	2	-0.74	1.72E-04	Shikimates and Phenylpropanoids	Lignans	Arylnaphthalene and aryltetralin lignans
C5906	2	-0.70	4.51E-04	Polyketides	Macrolides	Open-chain polyketides
		0.70	4.55E-04	Shikimates and Phenylpropanoids	Flavonoids	Flavonols
C5970	2	-0.73	2.35E-04	Alkaloids	Nicotinic acid alkaloids	Pyridine alkaloids
		0.73	2.02E-04	Amino acids and Peptides	Oligopeptides	Linear peptides
C6099	2	0.72	3.16E-04	Amino acids and Peptides	Small peptides	Aminoacids
C6153	2	0.71	3.35E-04	Polyketides	Naphthalenes	Bisnaphthalenes
C6198	2	-0.72	2.96E-04	Alkaloids	Pseudoalkaloids (transamidation)	Purine alkaloids
		0.78	3.83E-05	Amino acids and Peptides	Small peptides	Pyridine alkaloids
		-0.73	2.02E-04	Amino acids and Peptides	Small peptides	Tripeptides
		-0.72	2.91E-04	Shikimates and Phenylpropanoids	Flavonoids	Flavones
		-0.71	3.32E-04	Shikimates and Phenylpropanoids	Lignans	Furofuranoid lignans
		-0.74	1.57E-04	Terpenoids	Monoterpenoids	Secoiridoid monoterpenoids
C6728	2	-0.73	2.38E-04	Polyketides	Macrolides	Open-chain polyketides
C6932	2	-0.72	2.68E-04	Alkaloids	Pseudoalkaloids (transamidation)	Open-chain polyketides
		-0.71	3.58E-04	Amino acids and Peptides	Small peptides	Dipeptides
		-0.73	1.86E-04	Amino acids and Peptides	Small peptides	Tripeptides
		-0.72	3.05E-04	Fatty acids	Fatty esters	Wax diesters
		0.73	2.33E-04	Fatty acids	Glycerolipids	Triacylglycerols

		-0.75	1.22E-04	Fatty acids	Sesquiterpenoids	Dicarboxylic acids
		0.72	3.14E-04	Polyketides	Macrolides	Macrolide lactones
		-0.75	1.18E-04	Polyketides	Macrolides	Open-chain polyketides
		-0.77	6.65E-05	Shikimates and Phenylpropanoids	Coumarins	Triketide meroterpenoids
		0.72	3.05E-04	Shikimates and Phenylpropanoids	Flavonoids	Flavones
		0.83	7.31E-06	Shikimates and Phenylpropanoids	Flavonoids	Flavonols
		-0.76	7.88E-05	Shikimates and Phenylpropanoids	Lignans	Arylnaphthalene and aryltetralin lignans
		0.72	3.05E-04	Terpenoids	Carotenoids (C40)	Carotenoids (C40, α - β - γ - δ - ϵ - ζ - η - θ - ι - κ - λ - μ - ν - ξ - $\omicron$ - π - ρ - σ - τ - υ - ϕ - χ - ψ - ω)
		-0.72	2.94E-04	Terpenoids	Triterpenoids	Dammarane and Protostane triterpenoids
C7234	2	0.71	3.26E-04	Amino acids and Peptides	Oligopeptides	Tripeptides
		-0.81	1.39E-05	Alkaloids	Nicotinic acid alkaloids	Pyridine alkaloids
		-0.77	6.86E-05	Alkaloids	Ornithine alkaloids	Phenylethylamines
		-0.82	1.03E-05	Alkaloids	Pseudoalkaloids (transamidation)	Dipeptides
		-0.80	1.93E-05	Alkaloids	Pseudoalkaloids (transamidation)	Open-chain polyketides
C7335	2	0.72	3.17E-04	Amino acids and Peptides	Oligopeptides	Linear peptides
		-0.79	2.37E-05	Amino acids and Peptides	Small peptides	Aminoacids
		-0.78	5.48E-05	Fatty acids	Fatty acyls	Fatty alcohols
		-0.75	1.25E-04	Fatty acids	Fatty esters	Wax diesters
		-0.75	1.10E-04	Fatty acids	Fatty esters	Wax monoesters
		0.71	3.57E-04	Fatty acids	Glycerolipids	Triacylglycerols

0.72	3.05E-04	Fatty acids	Glycerolipids	Triacylglycerols
-0.73	2.02E-04	Fatty acids	Glycerophospholipids	Glycerophosphocholines
-0.79	3.43E-05	Fatty acids	Linear polyketides	Open-chain polyketides
-0.77	6.00E-05	Fatty acids	Sesquiterpenoids	Dicarboxylic acids
0.71	4.08E-04	Polyketides	Macrolides	Macrolide lactones
-0.75	1.22E-04	Polyketides	Naphthalenes	Flavones
-0.76	7.88E-05	Polyketides	Polycyclic aromatic polyketides	Bisnaphthalenes
-0.77	7.88E-05	Polyketides	Triterpenoids	Limonoids
-0.75	1.25E-04	Shikimates and Phenylpropanoids	Coumarins	Triketide meroterpenoids
-0.71	4.16E-04	Shikimates and Phenylpropanoids	Flavonoids	Flavanones
0.73	2.25E-04	Shikimates and Phenylpropanoids	Flavonoids	Flavones
0.83	7.31E-06	Shikimates and Phenylpropanoids	Flavonoids	Flavonols
-0.80	1.67E-05	Shikimates and Phenylpropanoids	Lignans	Arylnaphthalene and aryltetralin lignans
-0.81	1.37E-05	Shikimates and Phenylpropanoids	Lignans	Bisnaphthalenes
-0.80	1.88E-05	Shikimates and Phenylpropanoids	Naphthalenes	Bisnaphthalenes
0.75	1.22E-04	Shikimates and Phenylpropanoids	Phenolic acids (C6-C1)	Simple phenolic acids
0.72	3.05E-04	Terpenoids	Glycerolipids	Diacylglycerols
-0.80	1.88E-05	Terpenoids	Triterpenoids	Cucurbitane triterpenoids

		-0.83	5.54E-06	Terpenoids	Triterpenoids	Dammarane and Protostane triterpenoids
		-0.73	2.02E-04	Terpenoids	Triterpenoids	Quassinoids
		-0.71	3.58E-04	Alkaloids	Pseudoalkaloids (transamidation)	Open-chain polyketides
		0.72	2.99E-04	Alkaloids	Pseudoalkaloids (transamidation)	Phenylalanine- derived alkaloids
		0.71	4.04E-04	Amino acids and Peptides	Oligopeptides	Linear peptides
		-0.78	4.81E-05	Amino acids and Peptides	Small peptides	Tripeptides
		-0.73	2.40E-04	Fatty acids	Fatty esters	Wax diesters
		0.77	5.94E-05	Polyketides	Macrolides	Macrolide lactones
C7423	2	-0.70	4.78E-04	Polyketides	Macrolides	Open-chain polyketides
		0.73	1.95E-04	Shikimates and Phenylpropanoids	Flavonoids	Flavones
		-0.72	2.72E-04	Shikimates and Phenylpropanoids	Lignans	Arylnaphthalene and aryltetralin lignans
		-0.73	1.86E-04	Shikimates and Phenylpropanoids	Lignans	Furofuranoid lignans
		0.74	1.72E-04	Terpenoids	Carotenoids (C40)	Carotenoids (C40, C ₄₀ -C ₅₀)
C3752	3	-0.75	1.22E-04	Fatty acids	Fatty acyls	Fatty alcohols
		0.75	1.26E-04	Amino acids and Peptides	Small peptides	Aminoacids
		0.73	2.37E-04	Carbohydrates	Aminosugars and aminoglycosides	Aminosugars
C4591	3	0.70	4.71E-04	Fatty acids	Fatty acyls	Fatty alcohols
		-0.75	1.18E-04	Shikimates and Phenylpropanoids	Flavonoids	Flavones
C5644	3	0.71	3.20E-04	Carbohydrates	Nucleosides	Purine nucleosides

C6182	3	-0.76	9.12E-05	Alkaloids	Nicotinic acid alkaloids	Pyridine alkaloids
		-0.74	1.71E-04	Alkaloids	Ornithine alkaloids	Phenylethylamines
		-0.80	1.88E-05	Alkaloids	Ornithine alkaloids	Polyamines
		-0.75	1.31E-04	Alkaloids	Pseudoalkaloids (transamidation)	Dipeptides
		-0.73	2.31E-04	Alkaloids	Pseudoalkaloids (transamidation)	Open-chain polyketides
		-0.74	1.35E-04	Amino acids and Peptides	Oligopeptides	Cyclic peptides
		0.82	1.23E-05	Amino acids and Peptides	Oligopeptides	Linear peptides
		-0.70	4.73E-04	Amino acids and Peptides	Small peptides	Aminoacids
		-0.71	4.01E-04	Amino acids and Peptides	Small peptides	Aminoacids
		-0.82	1.12E-05	Amino acids and Peptides	Small peptides	Aminoacids
		0.73	2.02E-04	Amino acids and Peptides	Small peptides	Pyridine alkaloids
		-0.72	3.04E-04	Amino acids and Peptides	Small peptides	Tripeptides
		-0.80	1.88E-05	Carbohydrates	Nucleosides	Purine nucleosides
		-0.76	7.88E-05	Fatty acids	Fatty acyls	Fatty alcohols
		-0.77	6.99E-05	Fatty acids	Fatty esters	Wax diesters
		-0.81	1.37E-05	Fatty acids	Fatty esters	Wax monoesters
		0.70	4.86E-04	Fatty acids	Glycerolipids	Triacylglycerols
		0.76	8.73E-05	Fatty acids	Glycerolipids	Triacylglycerols
		-0.72	2.68E-04	Fatty acids	Glycerophospholipids	Glycerophosphocholines
		-0.75	1.08E-04	Fatty acids	Linear polyketides	Open-chain polyketides
		-0.79	3.63E-05	Fatty acids	Sesquiterpenoids	Dicarboxylic acids
		0.72	2.53E-04	Polyketides	Macrolides	Macrolide lactones
		-0.78	4.10E-05	Polyketides	Naphthalenes	Flavones
		-0.74	1.72E-04	Polyketides	Polycyclic aromatic polyketides	Bisnaphthalenes

-0.81	1.37E-05	Polyketides	Triterpenoids	Limonoids
-0.71	3.99E-04	Shikimates and Phenylpropanoids	Flavonoids	Flavanones
-0.71	3.32E-04	Shikimates and Phenylpropanoids	Flavonoids	Flavones
0.78	5.48E-05	Shikimates and Phenylpropanoids	Flavonoids	Flavones
-0.76	7.88E-05	Shikimates and Phenylpropanoids	Lignans	Arylnaphthalene and aryltetralin lignans
-0.71	3.44E-04	Shikimates and Phenylpropanoids	Lignans	Bisnaphthalenes
-0.73	2.49E-04	Shikimates and Phenylpropanoids	Lignans	Furofuranoid lignans
-0.72	2.91E-04	Shikimates and Phenylpropanoids	Naphthalenes	Bisnaphthalenes
0.75	1.22E-04	Shikimates and Phenylpropanoids	Phenolic acids (C6- C1)	Simple phenolic acids
0.79	3.43E-05	Terpenoids	Carotenoids (C40)	Carotenoids (C40, C ₄₀ -C ₄₀)
-0.71	3.57E-04	Terpenoids	Triterpenoids	Cucurbitane triterpenoids
-0.78	5.33E-05	Terpenoids	Triterpenoids	Quassinoids

Supplemental Table 4.4. Unannotated metabolites from each timepoint that were correlated with annotated metabolites. The correlated associated pathways, correlated associated superclass, and correlated associated class indicate the pathways, superclasses, and classes associated with the annotated metabolites that the unannotated metabolites were correlated with. Time indicates which timepoint the unannotated metabolite was significant ($\text{Log}_2\text{FC} \geq |5|$, $p\text{-value}_{\text{unadjusted}} < 0.05$).

Unannotated Metabolite	Time	Correlated Associated Pathway	Correlated Associated Superclass	Correlated Associated Class
C4755	1	Fatty acids, Shikimates and Phenylpropanoids, Amino acids and Peptides, Alkaloids, Terpenoids, Polyketides	Glycerolipids, Flavonoids, Oligopeptides, Carotenoids, Pseudoalkaloids, Oligopeptides, Small Peptides, Phenolic acids, Macrolides	Triacylglycerols, Flavones, Flavonols, Microlins and Mirabimids, Phenylalanine-derived alkaloids, Carotenoids, Capsaicins and Capsaicinoids, Linear Peptides, Pyridine Alkaloids, Simple Phenolic Acids, Macrolide Lactones
C7125	1	Shikimates and Phenylpropanoids, Fatty Acids	Flavonoids, Glycerolipids	Flavonols, Triacylglycerols
C7261	1	Shikimates and Phenylpropanoids, Alkaloids, Polyketides, Amino acids and Peptides	Flavonoids, Pseudoalkaloids, Macrolides, Oligopeptides	Flavonols, Phenylalanine-derived Alkaloids, Macrolide Lactones, Linear Peptides
C0643	2	Polyketides	Naphthalenes	Bisnaphthalenes
C1576	2	Polyketides	Naphthalenes	Bisnaphthalenes
C3110	2	Shikimates and Phenylpropanoids	Flavonoids	Flavones
C5906	2	Shikimates and Phenylpropanoids	Flavonoids	Flavonols
C5970	2	Amino acids and Peptides	Oligopeptides	Linear peptides
C6099	2	Amino acids and Peptides	Small peptides	Aminoacids
C6153	2	Polyketides	Naphthalenes	Bisnaphthalenes
C6198	2	Amino acids and Peptides	Small peptides	Pyridine alkaloids

C6932	2	Polyketides, Shikimates and Phenylpropanoids, Terpenoids, Fatty Acids	Macrolides, Flavonoids, Carotenoids, Glycerolipids	Macrolide Lactones, Flavones, Carotenoids, Triacylglycerols, Flavonols
C7234	2	Amino acids and Peptides	Oligopeptides	Tripeptides
C7335	2	Polyketides, Fatty Acids, Amino Acids and Peptides, Terpenoids	Macrolides, Glycerolipids, Oligopeptides, Flavonoids, Phenolic Acids	Macrolide lactones, Triacylglycerols, Linear Peptides, Diacylglycerols, Flavones, Simple Phenolic Acids, Flavonols
C7423	2	Amino acids and Peptides, Alkaloids, Shikimates and Phenylpropanoids, Terpenoids, Polyketides	Oligopeptides, Pseudoalkaloids, Flavonoids, Carotenoids, Macrolides	Linear Peptides, Phenylalanine-derived Alkaloids, Flavones, Carotenoids, Macrolide Lactones
C4591	3	Fatty Acids, Carbohydrates, Amino Acids and Peptides	Fatty Acyls, Aminosugars and Aminoglycosides, Small Peptides	Fatty Alcohols, Aminosugars, Aminoacids
C5644	3	Carbohydrates	Nucleosides	Purine nucleosides
C6182	3	Fatty acids, Polyketides, Amino acids and Peptides, Shikimates and Phenylpropanoids, Terpenoids	Glycerolipids, Macrolides, Small Peptides, Phenolic Acids, Flavonoids, Carotenoids, Oligopeptides	Triacylglycerols, Macrolide Lactones, Pyridine Alkaloids, Simple Phenolic Acids, Flavones, Carotenoids, Linear Peptides

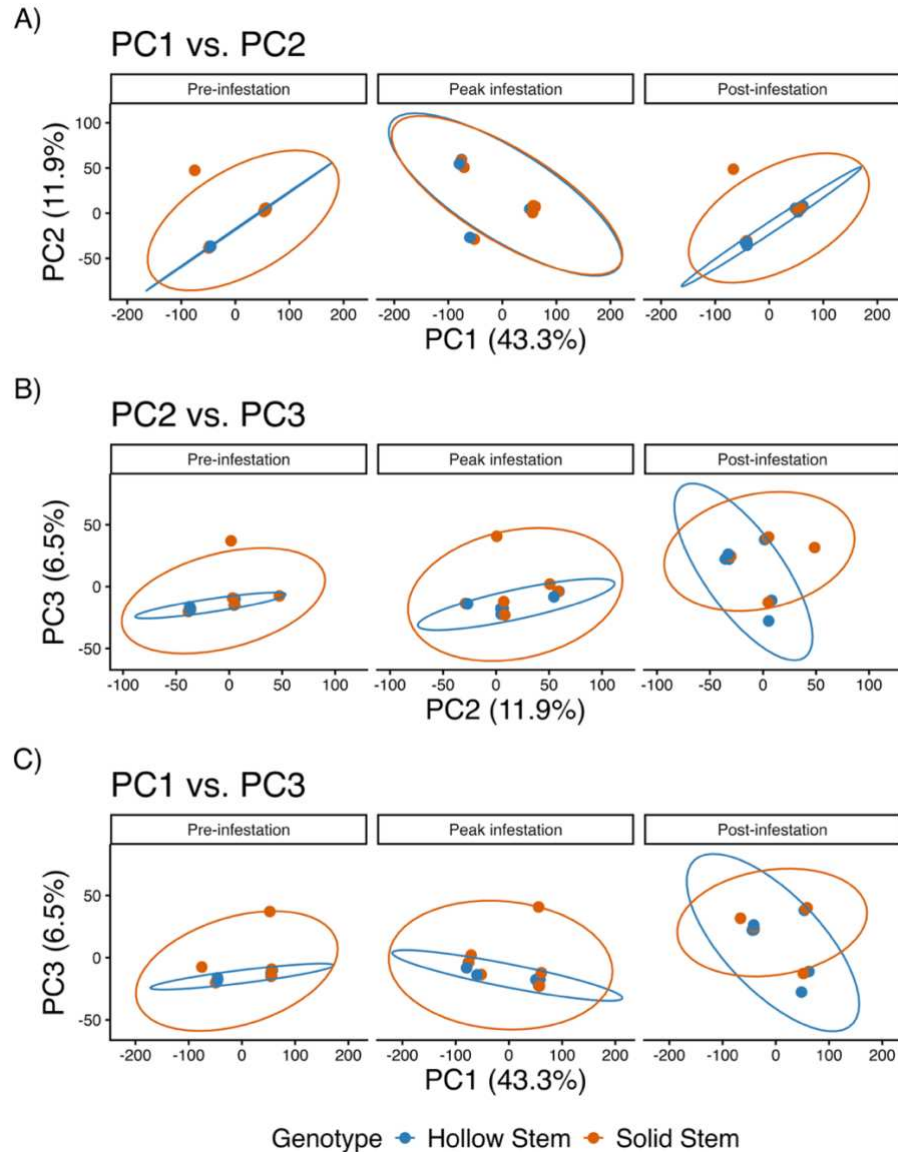
Supplemental Table 4.5. Correlations for unannotated metabolites (UA met), phytohormones (PH), annotated metabolites (A Met), pathways and classes associated with the annotated metabolites. The correlations are labeled as follows: correlation for unannotated metabolite and phytohormone (UA Met and PH r), correlation for unannotated metabolite and annotated metabolite (UA Met and A Met r), and correlation for annotated metabolite and phytohormone (A Met and PH r).

UA Met	t	PH	A Met	Pathway	Class	UA Met and PH r	UA Met and A Met r	A Met and PH r
C4755	1	ABA	C0510	Fatty acids	Open-chain polyketides	0.81	-0.71	-0.84
			C1482	Fatty acids	Triacylglycerols	0.81	0.71	0.81
			C3137	Shikimates and Phenylpropanoids	Flavonols	0.81	0.72	0.82
			C0502	Fatty acids	Wax diesters	0.81	-0.74	-0.88
			C1336	Polyketides	Bisnaphthalenes	0.81	-0.75	-0.82
			C4690	Terpenoids	Carotenoids (C40, $\text{C}_{40}^{\leq}-\text{C}_{40}^{\leq}$)	0.81	0.75	0.82
			C6226	Polyketides	Macrolide lactones	0.81	0.84	0.82
C7423	2	ABA	C0502	Fatty acids	Wax diesters	0.93	-0.73	-0.82
			C2130	Shikimates and Phenylpropanoids	Flavones	0.93	0.73	0.84
C7423	2	JA-Ile	C2130	Shikimates and Phenylpropanoids	Flavones	0.83	0.73	0.82
			C2663	Terpenoids	Carotenoids (C40, $\text{C}_{40}^{\leq}-\text{C}_{40}^{\leq}$)	0.83	0.74	0.89
C6932	2	IAA	C2663	Terpenoids	Carotenoids (C40, $\text{C}_{40}^{\leq}-\text{C}_{40}^{\leq}$)	-0.84	0.72	-0.95
			C0850	Terpenoids	Dammarane and Protostane triterpenoids	-0.84	-0.72	0.83
			C0572	Amino acids and Peptides	Tripeptides	-0.84	-0.73	0.90
			C3137	Shikimates and Phenylpropanoids	Flavonols	-0.84	0.83	-0.87
C7423	2	IAA	C2130	Shikimates and Phenylpropanoids	Flavones	-0.84	0.73	-0.82
			C1779	Shikimates and Phenylpropanoids	Furofuranoid lignans	-0.84	-0.73	0.85
			C2663	Terpenoids	Carotenoids (C40, $\text{C}_{40}^{\leq}-\text{C}_{40}^{\leq}$)	-0.84	0.74	-0.95

C0572	Amino acids and Peptides	Tripeptides	-0.84	-0.78	0.90
-------	-----------------------------	-------------	-------	-------	------

CHAPTER V SUPPLEMENTAL

Per-Timepoint PCAs of Metabolite Expression



Supplemental Figure 5.1. Principal component (PC) plots for metabolite expression separated by time point where (A) shows variance explained by PC1 and PC2, (B) shows PC2 and PC3, and (C) shows PC1 and PC3. Blue ellipses and datapoints indicate the hollow stem line (CO20SFD047R.4) and orange ellipses and datapoints indicate the solid stem line (CO20SFD047R.1). Ellipses indicate the spread of the genotypes at different timepoints.

Supplemental Table 5.1: Fisher's test results for protein accessions across all time points including number of present or absent proteins for each genotype at each time point and the p-value of the Fisher's exact test.

Time	Accession	Year	N Present Byrd	N Present Byrd CL Plus	N Absent Byrd	N Absent Byrd CL Plus	p-value
tl	A0A1D5V002	2024	2	2	1	0	1.0
	A0A3B6B2N9	2024	3	2	0	0	1.0
	A0A3B6E8H5	2024	2	2	1	0	1.0
	A0A3B6EPG4	2024	3	2	0	0	1.0
	A0A3B6FIX9	2024	3	2	0	0	1.0
	A0A3B6GXA9	2024	3	2	0	0	1.0
	A0A3B6IPK1	2024	1	0	2	2	1.0
	A0A3B6KEC6	2024	3	2	0	0	1.0
	A0A3B6KEI2	2024	3	2	0	0	1.0
	A0A3B6MNS6	2024	3	2	0	0	1.0
	A0A3B6N1C6	2024	2	2	1	0	1.0
	A0A3B6NNV4	2024	3	1	0	1	0.4
	A0A3B6RD10	2024	0	0	3	2	1.0
	A0A3B6TVA1	2024	1	1	2	1	1.0
	A0A3B6U7G0	2024	1	1	2	1	1.0
	A0A1D5V002	2025	0	1	3	2	1.0
	A0A3B6B2N9	2025	1	0	2	3	1.0
	A0A3B6E8H5	2025	1	0	2	3	1.0
	A0A3B6EPG4	2025	1	0	2	3	1.0
	A0A3B6FIX9	2025	1	0	2	3	1.0
	A0A3B6GXA9	2025	1	0	2	3	1.0
	A0A3B6IPK1	2025	0	3	3	0	0.1
	A0A3B6KEC6	2025	0	0	3	3	1.0
	A0A3B6KEI2	2025	1	0	2	3	1.0
	A0A3B6MNS6	2025	0	0	3	3	1.0

	A0A3B6N1C6	2025	0	2	3	1	0.4
	A0A3B6NNV4	2025	3	1	0	2	0.4
	A0A3B6RD10	2025	3	2	0	1	1.0
	A0A3B6TVA1	2025	3	0	0	3	0.1
	A0A3B6U7G0	2025	2	1	1	2	1.0
t2	A0A1D5V002	2024	1	2	2	0	0.4
	A0A3B6B2N9	2024	1	2	2	0	0.4
	A0A3B6E8H5	2024	1	0	2	2	1.0
	A0A3B6EPG4	2024	2	2	1	0	1.0
	A0A3B6FIX9	2024	2	0	1	2	0.4
	A0A3B6GXA9	2024	3	1	0	1	0.4
	A0A3B6IPK1	2024	0	2	3	0	0.1
	A0A3B6KEC6	2024	3	2	0	0	1.0
	A0A3B6KEI2	2024	3	1	0	1	0.4
	A0A3B6MNS6	2024	1	1	2	1	1.0
	A0A3B6N1C6	2024	0	2	3	0	0.1
	A0A3B6NNV4	2024	1	0	2	2	1.0
	A0A3B6RD10	2024	0	0	3	2	1.0
	A0A3B6TVA1	2024	2	0	1	2	0.4
	A0A3B6U7G0	2024	2	0	1	2	0.4
	A0A1D5V002	2025	1	1	2	2	1.0
	A0A3B6B2N9	2025	3	0	0	3	0.1
	A0A3B6E8H5	2025	2	0	1	3	0.4
	A0A3B6EPG4	2025	1	0	2	3	1.0
	A0A3B6FIX9	2025	3	0	0	3	0.1
	A0A3B6GXA9	2025	2	0	1	3	0.4
	A0A3B6IPK1	2025	2	1	1	2	1.0
	A0A3B6KEC6	2025	0	0	3	3	1.0

	A0A3B6KEI2	2025	0	0	3	3	1.0
	A0A3B6MNS6	2025	2	0	1	3	0.4
	A0A3B6N1C6	2025	1	1	2	2	1.0
	A0A3B6NNV4	2025	2	1	1	2	1.0
	A0A3B6RD10	2025	3	2	0	1	1.0
	A0A3B6TVA1	2025	1	2	2	1	1.0
	A0A3B6U7G0	2025	2	2	1	1	1.0
t3	A0A1D5V002	2024	1	2	2	0	0.4
	A0A3B6B2N9	2024	3	1	0	1	0.4
	A0A3B6E8H5	2024	2	1	1	1	1.0
	A0A3B6EPG4	2024	2	2	1	0	1.0
	A0A3B6FIX9	2024	3	0	0	2	0.1
	A0A3B6GXA9	2024	2	0	1	2	0.4
	A0A3B6IPK1	2024	0	2	3	0	0.1
	A0A3B6KEC6	2024	2	2	1	0	1.0
	A0A3B6KEI2	2024	2	2	1	0	1.0
	A0A3B6MNS6	2024	2	2	1	0	1.0
	A0A3B6N1C6	2024	2	1	1	1	1.0
	A0A3B6NNV4	2024	1	1	2	1	1.0
	A0A3B6RD10	2024	1	0	2	2	1.0
	A0A3B6TVA1	2024	3	1	0	1	0.4
	A0A3B6U7G0	2024	1	0	2	2	1.0
	A0A1D5V002	2025	0	1	3	2	1.0
	A0A3B6B2N9	2025	3	1	0	2	0.4
	A0A3B6E8H5	2025	1	0	2	3	1.0
	A0A3B6EPG4	2025	1	0	2	3	1.0
	A0A3B6FIX9	2025	2	0	1	3	0.4
	A0A3B6GXA9	2025	3	0	0	3	0.1

A0A3B6IPK1	2025	2	3	1	0	1.0
A0A3B6KEC6	2025	2	0	1	3	0.4
A0A3B6KEI2	2025	1	0	2	3	1.0
A0A3B6MNS6	2025	1	0	2	3	1.0
A0A3B6N1C6	2025	0	2	3	1	0.4
A0A3B6NNV4	2025	3	0	0	3	0.1
A0A3B6RD10	2025	3	1	0	2	0.4
A0A3B6TVA1	2025	2	1	1	2	1.0
A0A3B6U7G0	2025	2	2	1	1	1.0

Supplemental Table 5.2. Positive and negative correlations for significant unannotated metabolites ($\text{Log2FC} \geq |5|$, $\text{p-value}_{\text{unadjusted}} < 0.05$) identified at all timepoints. Correlations (r) indicates the correlation between the unannotated and annotated metabolites, $\text{p-value}_{\text{FDR}}$ indicates the adjusted significance of the correlation p-value, and the pathways, superclasses, and classes associated with the unannotated metabolites.

Unannotated Metabolite	time	r	$\text{p-value}_{\text{FDR}}$	Correlated Associated Pathway	Correlated Associated Superclass	Correlated Associated Class
C4755	1	-0.73	4.81E-05	Alkaloids	Pseudoalkaloids (transamidation)	Open-chain polyketides
		-0.77	7.98E-06	Alkaloids	Ornithine alkaloids	Polyamines
		-0.83	8.21E-07	Alkaloids	Pseudoalkaloids (transamidation)	Purine alkaloids
		-0.72	5.74E-05	Alkaloids	Nicotinic acid alkaloids	Pyridine alkaloids
		-0.73	4.02E-05	Amino acids and Peptides	Small peptides	Aminoacids
		-0.72	6.92E-05	Amino acids and Peptides	Small peptides	Dipeptides
		0.84	5.53E-07	Amino acids and Peptides	Oligopeptides	Linear peptides
		0.80	2.48E-06	Amino acids and Peptides	Small peptides	Pyridine alkaloids
		-0.77	9.29E-06	Amino acids and Peptides	Small peptides	Tripeptides
		0.78	6.70E-06	Amino acids and Peptides	Small peptides	Tripeptides
		-0.75	2.11E-05	Carbohydrates	Nucleosides	Purine alkaloids
		-0.80	3.03E-06	Carbohydrates	Nucleosides	Purine nucleosides
		-0.80	2.62E-06	Carbohydrates	Nucleosides	Purine nucleosides
		-0.70	1.04E-04	Fatty acids	Sesquiterpenoids	Dicarboxylic acids
		-0.75	2.15E-05	Fatty acids	Fatty acyls	Fatty alcohols
		-0.73	4.20E-05	Fatty acids	Linear polyketides	Open-chain polyketides
		-0.76	1.52E-05	Fatty acids	Fatty esters	Wax diesters
		-0.77	1.01E-05	Fatty acids	Fatty esters	Wax monoesters
		-0.73	4.63E-05	Polyketides	Polycyclic aromatic polyketides	Bisnaphthalenes
		-0.71	9.02E-05	Polyketides	Naphthalenes	Flavones
		-0.71	7.92E-05	Polyketides	Triterpenoids	Limonoids

		-0.78	6.17E-06	Polyketides	Macrolides	Open-chain polyketides
		-0.70	1.11E-04	Shikimates and Phenylpropanoids	Naphthalenes	Bisnaphthalenes
		-0.77	7.88E-06	Shikimates and Phenylpropanoids	Flavonoids	Flavanones
		0.74	2.50E-05	Shikimates and Phenylpropanoids	Flavonoids	Flavones
		0.78	5.14E-06	Shikimates and Phenylpropanoids	Flavonoids	Flavones
		-0.80	2.45E-06	Shikimates and Phenylpropanoids	Flavonoids	Flavones
		-0.80	2.52E-06	Shikimates and Phenylpropanoids	Lignans	Furofuranoid lignans
		-0.82	1.12E-06	Shikimates and Phenylpropanoids	Monoterpenoids	Secoiridoid monoterpenoids
		0.71	7.27E-05	Shikimates and Phenylpropanoids	Phenolic acids (C6-C1)	Simple phenolic acids
		-0.74	2.81E-05	Terpenoids	Triterpenoids	Cucurbitane triterpenoids
		0.80	2.52E-06	Terpenoids	Glycerolipids	Diacylglycerols
		-0.72	5.74E-05	Terpenoids	Sesquiterpenoids	Hydrocarbons
C5443	1	0.71	9.79E-05	Carbohydrates	Aminosugars and aminoglycosides	Aminosugars
		-0.72	5.54E-05	Alkaloids	Ornithine alkaloids	Phenylethylamines
		-0.71	7.64E-05	Amino acids and Peptides	Small peptides	Aminoacids
C5600	1	-0.72	5.69E-05	Fatty acids	Linear polyketides	Open-chain polyketides
		0.71	8.90E-05	Shikimates and Phenylpropanoids	Flavonoids	Flavonols
C6882	1	-0.71	7.17E-05	Polyketides	Polycyclic aromatic polyketides	Bisnaphthalenes

		-0.71	8.55E-05	Shikimates and Phenylpropanoids	Lignans	Arylnaphthalene and aryltetralin lignans
		-0.74	3.40E-05	Shikimates and Phenylpropanoids	Coumarins	Triketide
		-0.71	9.02E-05	Terpenoids	Triterpenoids	meroterpenoids Quassinoids
C7125	1	0.73	3.57E-05	Terpenoids	Carotenoids (C40)	Carotenoids (C40, C ₄₀ -C ₅₀)
		0.74	3.29E-05	Alkaloids	Pseudoalkaloids (transamidation)	Capsaicins and Capsaicinoids
		-0.77	8.56E-06	Alkaloids	Pseudoalkaloids (transamidation)	Open-chain polyketides
		0.77	1.03E-05	Alkaloids	Pseudoalkaloids (transamidation)	Phenylalanine-derived alkaloids
		-0.71	8.13E-05	Alkaloids	Ornithine alkaloids	Phenylethylamines
		-0.78	6.46E-06	Alkaloids	Ornithine alkaloids	Polyamines
		-0.78	5.40E-06	Alkaloids	Pseudoalkaloids (transamidation)	Purine alkaloids
		-0.74	3.37E-05	Alkaloids	Nicotinic acid alkaloids	Pyridine alkaloids
		-0.77	9.43E-06	Amino acids and Peptides	Small peptides	Aminoacids
C7261	1	-0.70	9.96E-05	Amino acids and Peptides	Oligopeptides	Cyclic peptides
		-0.72	5.16E-05	Amino acids and Peptides	Small peptides	Dipeptides
		0.76	1.34E-05	Amino acids and Peptides	Oligopeptides	Linear peptides
		0.74	2.53E-05	Amino acids and Peptides	Oligopeptides	Microcolins and mirabimids
		0.76	1.47E-05	Amino acids and Peptides	Small peptides	Pyridine alkaloids
		-0.78	5.10E-06	Amino acids and Peptides	Small peptides	Tripeptides
		0.81	1.74E-06	Amino acids and Peptides	Small peptides	Tripeptides
		-0.77	8.78E-06	Carbohydrates	Nucleosides	Purine alkaloids
		-0.74	2.50E-05	Carbohydrates	Nucleosides	Purine nucleosides
		-0.77	9.12E-06	Carbohydrates	Nucleosides	Purine nucleosides
		-0.78	5.90E-06	Fatty acids	Sesquiterpenoids	Dicarboxylic acids

-0.78	4.99E-06	Fatty acids	Fatty acyls	Fatty alcohols
-0.70	1.04E-04	Fatty acids	Linear polyketides	Open-chain polyketides
-0.78	6.10E-06	Fatty acids	Fatty esters	Wax diesters
-0.75	1.77E-05	Fatty acids	Fatty esters	Wax monoesters
-0.74	2.42E-05	Polyketides	Polycyclic aromatic polyketides	Bisnaphthalenes
-0.83	8.21E-07	Polyketides	Naphthalenes	Flavones
-0.73	4.16E-05	Polyketides	Triterpenoids	Limonoids
-0.70	1.02E-04	Polyketides	Macrolides	Open-chain polyketides
-0.76	1.32E-05	Shikimates and Phenylpropanoids	Lignans	Arylnaphthalene and aryltetralin lignans
-0.72	6.24E-05	Shikimates and Phenylpropanoids	Naphthalenes	Bisnaphthalenes
-0.73	3.64E-05	Shikimates and Phenylpropanoids	Flavonoids	Flavanones
0.72	6.85E-05	Shikimates and Phenylpropanoids	Flavonoids	Flavones
-0.77	7.14E-06	Shikimates and Phenylpropanoids	Flavonoids	Flavones
0.72	6.73E-05	Shikimates and Phenylpropanoids	Flavonoids	Flavonols
-0.71	7.17E-05	Shikimates and Phenylpropanoids	Lignans	Furofuranoid lignans
0.77	1.03E-05	Shikimates and Phenylpropanoids	Phenolic acids (C6-C1)	Simple phenolic acids
0.70	1.07E-04	Terpenoids	Carotenoids (C40)	Carotenoids (C40, α - β - γ - δ - ϵ - ζ - η - θ - ι - κ - λ - μ - ν - ξ - $\omicron$ - π - ρ - σ - τ - υ - ϕ - χ - ψ - ω)
-0.86	3.01E-07	Terpenoids	Triterpenoids	Cucurbitane triterpenoids

		-0.72	5.34E-05	Terpenoids	Triterpenoids	Dammarane and Protostane triterpenoids
		0.74	2.87E-05	Terpenoids	Glycerolipids	Diacylglycerols
		-0.79	3.85E-06	Terpenoids	Sesquiterpenoids	Hydrocarbons
		-0.75	1.55E-05	Terpenoids	Triterpenoids	Quassinoids
C7437	1	0.73	4.26E-05	Amino acids and Peptides	Small peptides	Tripeptides
C3110	2	0.78	6.03E-06	Alkaloids	Ornithine alkaloids	Dipeptides
		0.72	5.54E-05	Polyketides	Polycyclic aromatic polyketides	Anthraquinones and anthrones
C3899	2	-0.70	1.11E-04	Alkaloids	Ornithine alkaloids	Phenylethylamines
		-0.72	5.16E-05	Amino acids and Peptides	Small peptides	Aminoacids
		-0.71	8.36E-05	Fatty acids	Linear polyketides	Open-chain polyketides
		0.73	3.57E-05	Shikimates and Phenylpropanoids	Flavonoids	Flavonols
C5031	2	-0.79	3.70E-06	Alkaloids	Pseudoalkaloids (transamidation)	Dipeptides
		-0.71	9.63E-05	Alkaloids	Pseudoalkaloids (transamidation)	Open-chain polyketides
		-0.72	5.88E-05	Alkaloids	Ornithine alkaloids	Phenylethylamines
		-0.75	1.57E-05	Alkaloids	Ornithine alkaloids	Polyamines
		-0.73	3.50E-05	Amino acids and Peptides	Small peptides	Aminoacids
		-0.76	1.20E-05	Amino acids and Peptides	Small peptides	Aminoacids
		0.70	1.08E-04	Amino acids and Peptides	Oligopeptides	Microcolins and mirabimids
		0.75	2.26E-05	Amino acids and Peptides	Oligopeptides	Tripeptides
		-0.75	1.61E-05	Carbohydrates	Aminosugars and aminoglycosides	Aminosugars
		0.75	2.20E-05	Carbohydrates	Nucleosides	Pyrimidine nucleosides

		-0.79	4.77E-06	Fatty acids	Sesquiterpenoids	Dicarboxylic acids
		-0.76	1.44E-05	Fatty acids	Fatty acyls	Fatty alcohols
		-0.71	8.90E-05	Fatty acids	Linear polyketides	Open-chain polyketides
		-0.72	5.20E-05	Fatty acids	Fatty esters	Wax diesters
		-0.76	1.45E-05	Polyketides	Polycyclic aromatic polyketides	Bisnaphthalenes
		-0.71	7.51E-05	Polyketides	Macrolides	Open-chain polyketides
		-0.72	6.08E-05	Shikimates and Phenylpropanoids	Lignans	Arylnaphthalene and aryltetralin lignans
		-0.71	7.27E-05	Shikimates and Phenylpropanoids	Lignans	Bisnaphthalenes
		0.70	1.05E-04	Shikimates and Phenylpropanoids	Flavonoids	Flavones
		0.74	2.36E-05	Shikimates and Phenylpropanoids	Flavonoids	Flavones
		0.72	6.24E-05	Shikimates and Phenylpropanoids	Flavonoids	Flavonols
		0.77	8.55E-06	Shikimates and Phenylpropanoids	Flavonoids	Flavonols
		-0.77	1.03E-05	Shikimates and Phenylpropanoids	Coumarins	Triketide meroterpenoids
		-0.75	2.20E-05	Terpenoids	Triterpenoids	Cucurbitane triterpenoids
		0.70	1.07E-04	Terpenoids	Glycerolipids	Diacylglycerols
C5520	2	-0.71	7.68E-05	Carbohydrates	Nucleosides	Purine alkaloids
		-0.71	8.90E-05	Carbohydrates	Nucleosides	Purine nucleosides
		-0.71	9.63E-05	Fatty acids	Fatty esters	Wax monoesters
C5654	2	-0.75	2.09E-05	Alkaloids	Pseudoalkaloids (transamidation)	Dipeptides
		-0.75	1.68E-05	Alkaloids	Ornithine alkaloids	Phenylethylamines

		-0.78	5.14E-06	Amino acids and Peptides	Small peptides	Aminoacids
		0.78	6.62E-06	Amino acids and Peptides	Oligopeptides	Tripeptides
		-0.71	9.72E-05	Fatty acids	Linear polyketides	Open-chain polyketides
		-0.70	1.14E-04	Shikimates and Phenylpropanoids	Naphthalenes	Bisnaphthalenes
C5975	2	-0.71	7.68E-05	Amino acids and Peptides	Small peptides	Aminoacids
C6099	2	0.72	5.16E-05	Amino acids and Peptides	Small peptides	Dipeptides
		0.79	4.68E-06	Amino acids and Peptides	Small peptides	Tripeptides
		0.71	8.84E-05	Polyketides	Triterpenoids	Limonoids
C6153	2	0.73	4.02E-05	Shikimates and Phenylpropanoids	Flavonoids	Flavanones
		-0.72	5.96E-05	Shikimates and Phenylpropanoids	Flavonoids	Flavones
		-0.74	2.87E-05	Alkaloids	Pseudoalkaloids (transamidation)	Purine alkaloids
		0.70	9.96E-05	Amino acids and Peptides	Small peptides	Pyridine alkaloids
		-0.71	8.55E-05	Amino acids and Peptides	Small peptides	Tripeptides
		-0.71	7.30E-05	Fatty acids	Linear polyketides	Open-chain polyketides
C6198	2	-0.71	8.71E-05	Shikimates and Phenylpropanoids	Flavonoids	Flavanones
		0.73	4.26E-05	Shikimates and Phenylpropanoids	Flavonoids	Flavones
		0.73	4.21E-05	Shikimates and Phenylpropanoids	Flavonoids	Flavones
		-0.71	7.22E-05	Shikimates and Phenylpropanoids	Lignans	Furofuranoid lignans
C6710	2	0.80	2.52E-06	Carbohydrates	Nucleosides	Pyrimidine nucleosides
C6728	2	-0.79	3.56E-06	Alkaloids	Pseudoalkaloids (transamidation)	Dipeptides

		-0.73	4.53E-05	Alkaloids	Pseudoalkaloids (transamidation)	Open-chain polyketides
		-0.80	2.38E-06	Alkaloids	Ornithine alkaloids	Phenylethylamines
		-0.73	4.05E-05	Amino acids and Peptides	Small peptides	Aminoacids
		-0.76	1.12E-05	Amino acids and Peptides	Small peptides	Aminoacids
		-0.76	1.49E-05	Carbohydrates	Aminosugars and aminoglycosides	Aminosugars
		-0.79	3.78E-06	Carbohydrates	Nucleosides	Purine alkaloids
		-0.73	4.09E-05	Carbohydrates	Nucleosides	Purine nucleosides
		0.71	7.57E-05	Fatty acids	Glycerolipids	Triacylglycerols
		-0.75	1.93E-05	Polyketides	Naphthalenes	Flavones
		-0.71	7.22E-05	Shikimates and Phenylpropanoids	Lignans	Arylnaphthalene and aryltetralin lignans
		-0.71	8.13E-05	Shikimates and Phenylpropanoids	Lignans	Bisnaphthalenes
		0.73	4.05E-05	Shikimates and Phenylpropanoids	Flavonoids	Flavones
		0.73	4.29E-05	Shikimates and Phenylpropanoids	Flavonoids	Flavonols
		0.80	2.55E-06	Shikimates and Phenylpropanoids	Flavonoids	Flavonols
		-0.73	4.26E-05	Terpenoids	Triterpenoids	Dammarane and Protostane triterpenoids
		-0.73	4.70E-05	Terpenoids	Triterpenoids	Quassinoids
C7335	2	-0.78	6.17E-06	Alkaloids	Pseudoalkaloids (transamidation)	Dipeptides
		-0.74	3.04E-05	Alkaloids	Ornithine alkaloids	Phenylethylamines
		-0.71	8.64E-05	Amino acids and Peptides	Small peptides	Aminoacids
		-0.73	4.36E-05	Amino acids and Peptides	Small peptides	Aminoacids
		0.72	5.30E-05	Amino acids and Peptides	Oligopeptides	Tripeptides

		-0.71	7.64E-05	Carbohydrates	Nucleosides	Purine nucleosides
		0.75	2.20E-05	Fatty acids	Glycerolipids	Triacylglycerols
		-0.75	2.18E-05	Polyketides	Macrolides	Open-chain polyketides
		0.82	1.01E-06	Shikimates and Phenylpropanoids	Flavonoids	Flavonols
		-0.75	1.66E-05	Shikimates and Phenylpropanoids	Coumarins	Triketide meroterpenoids
		-0.72	5.45E-05	Terpenoids	Triterpenoids	Cucurbitane triterpenoids
		-0.71	7.98E-05	Terpenoids	Triterpenoids	Dammarane and Protostane triterpenoids
C7423	2	-0.71	8.06E-05	Alkaloids	Pseudoalkaloids (transamidation)	Open-chain polyketides
		-0.75	1.81E-05	Alkaloids	Ornithine alkaloids	Polyamines
		0.74	2.32E-05	Amino acids and Peptides	Oligopeptides	Linear peptides
		-0.71	8.79E-05	Carbohydrates	Aminosugars and aminoglycosides	Aminosugars
		-0.76	1.15E-05	Fatty acids	Sesquiterpenoids	Dicarboxylic acids
		-0.75	1.70E-05	Fatty acids	Fatty acyls	Fatty alcohols
		-0.82	1.18E-06	Fatty acids	Fatty esters	Wax diesters
		-0.71	7.27E-05	Polyketides	Macrolides	Open-chain polyketides
		-0.82	1.28E-06	Terpenoids	Triterpenoids	Cucurbitane triterpenoids
C7466	2	-0.71	9.34E-05	Amino acids and Peptides	Oligopeptides	Cyclic peptides
		0.71	9.25E-05	Amino acids and Peptides	Small peptides	Dipeptides
C3752	3	-0.75	1.81E-05	Alkaloids	Pseudoalkaloids (transamidation)	Dipeptides
		-0.70	1.04E-04	Alkaloids	Pseudoalkaloids (transamidation)	Open-chain polyketides

		-0.83	8.21E-07	Alkaloids	Ornithine alkaloids	Polyamines
		-0.76	1.50E-05	Amino acids and Peptides	Small peptides	Aminoacids
		-0.70	1.10E-04	Carbohydrates	Saccharides	Monosaccharides
		-0.80	3.20E-06	Fatty acids	Sesquiterpenoids	Dicarboxylic acids
		-0.78	5.74E-06	Fatty acids	Fatty acyls	Fatty alcohols
		-0.79	3.53E-06	Fatty acids	Fatty esters	Wax diesters
		-0.76	1.19E-05	Fatty acids	Fatty esters	Wax monoesters
		-0.71	8.84E-05	Polyketides	Triterpenoids	Limonoids
		-0.72	6.80E-05	Polyketides	Macrolides	Open-chain polyketides
		-0.73	3.84E-05	Shikimates and Phenylpropanoids	Lignans	Arylnaphthalene and aryltetralin lignans
		-0.73	3.56E-05	Shikimates and Phenylpropanoids	Coumarins	Triketide meroterpenoids
		0.72	7.06E-05	Terpenoids	Carotenoids (C40)	Carotenoids (C40, ≤ -0.7)
		-0.74	2.33E-05	Terpenoids	Triterpenoids	Cucurbitane triterpenoids
C6182	3	0.73	4.31E-05	Alkaloids	Pseudoalkaloids (transamidation)	Capsaicins and Capsaicinoids
		0.74	3.24E-05	Alkaloids	Pseudoalkaloids (transamidation)	Phenylalanine-derived alkaloids
		-0.73	3.79E-05	Alkaloids	Ornithine alkaloids	Polyamines
		-0.77	8.57E-06	Alkaloids	Pseudoalkaloids (transamidation)	Purine alkaloids
		-0.78	6.28E-06	Alkaloids	Nicotinic acid alkaloids	Pyridine alkaloids
		-0.73	4.76E-05	Amino acids and Peptides	Small peptides	Aminoacids
		-0.73	4.05E-05	Amino acids and Peptides	Small peptides	Aminoacids
		-0.74	2.81E-05	Amino acids and Peptides	Small peptides	Aminoacids
		-0.71	9.79E-05	Amino acids and Peptides	Oligopeptides	Cyclic peptides
		-0.75	2.20E-05	Amino acids and Peptides	Small peptides	Dipeptides

0.78	5.40E-06	Amino acids and Peptides	Oligopeptides	Linear peptides
0.74	2.32E-05	Amino acids and Peptides	Small peptides	Pyridine alkaloids
0.73	3.56E-05	Amino acids and Peptides	Small peptides	Tripeptides
-0.78	5.56E-06	Amino acids and Peptides	Small peptides	Tripeptides
-0.77	1.03E-05	Carbohydrates	Nucleosides	Purine alkaloids
-0.72	6.73E-05	Carbohydrates	Nucleosides	Purine nucleosides
-0.81	1.62E-06	Carbohydrates	Nucleosides	Purine nucleosides
-0.74	2.33E-05	Fatty acids	Sesquiterpenoids	Dicarboxylic acids
-0.74	3.40E-05	Fatty acids	Fatty acyls	Fatty alcohols
-0.80	2.52E-06	Fatty acids	Linear polyketides	Open-chain polyketides
-0.79	4.29E-06	Fatty acids	Fatty esters	Wax diesters
-0.82	1.01E-06	Fatty acids	Fatty esters	Wax monoesters
-0.70	1.04E-04	Polyketides	Naphthalenes	Flavones
-0.79	3.38E-06	Polyketides	Triterpenoids	Limonoids
0.76	1.37E-05	Polyketides	Macrolides	Macrolide lactones
-0.78	6.11E-06	Shikimates and Phenylpropanoids	Naphthalenes	Bisnaphthalenes
-0.77	8.49E-06	Shikimates and Phenylpropanoids	Flavonoids	Flavanones
-0.76	1.17E-05	Shikimates and Phenylpropanoids	Flavonoids	Flavones
0.77	7.06E-06	Shikimates and Phenylpropanoids	Flavonoids	Flavones
0.79	4.73E-06	Shikimates and Phenylpropanoids	Flavonoids	Flavones
0.72	5.30E-05	Shikimates and Phenylpropanoids	Flavonoids	Flavonols
-0.77	8.95E-06	Shikimates and Phenylpropanoids	Lignans	Furofuranoid lignans

		-0.73	4.45E-05	Shikimates and Phenylpropanoids	Monoterpenoids	Secoiridoid monoterpenoids
		0.76	1.13E-05	Shikimates and Phenylpropanoids	Phenolic acids (C6-C1)	Simple phenolic acids
		-0.78	4.89E-06	Terpenoids	Sesquiterpenoids	Hydrocarbons
		-0.70	1.08E-04	Terpenoids	Triterpenoids	Quassinoids
		-0.71	7.12E-05	Alkaloids	Ornithine alkaloids	Phenylethylamines
		-0.72	7.06E-05	Amino acids and Peptides	Small peptides	Aminoacids
		-0.70	9.96E-05	Fatty acids	Linear polyketides	Open-chain polyketides
C6612	3	0.83	8.21E-07	Shikimates and Phenylpropanoids	Flavonoids	Flavones
		0.71	8.79E-05	Shikimates and Phenylpropanoids	Flavonoids	Flavonols
		0.72	5.94E-05	Shikimates and Phenylpropanoids	Flavonoids	Flavonols

Supplemental Table 5.3. All correlations for unannotated metabolites (UA met), phytohormones (PH), annotated metabolites (A Met), pathways and classes associated with the annotated metabolites. The correlations are labeled as follows: correlation for unannotated metabolite and phytohormone (UA Met and PH *r*), correlation for unannotated metabolite and annotated metabolite (UA Met and A Met *r*), and correlation for annotated metabolite and phytohormone (A Met and PH *r*).

UA Met	t	PH	A Met	Pathway	Class	UA Met and PH <i>r</i>	UA Met and A Met <i>r</i>	A Met and PH <i>r</i>
C6882	1	JA	C0769	Shikimates and Phenylpropanoids	Triketide meroterpenoids	0.71	-0.74	-0.78
			C1841	Fatty acids	Dicarboxylic acids	0.79	-0.70	-0.87
			C1869	Polyketides	Flavones	0.79	-0.71	-0.92
			C0611	Polyketides	Limonoids	0.79	-0.71	-0.72
			C7412	Terpenoids	Hydrocarbons	0.79	-0.72	-0.72
			C1074	Alkaloids	Open-chain polyketides	0.79	-0.73	-0.78
			C1336	Polyketides	Bisnaphthalenes	0.79	-0.73	-0.83
			C0510	Fatty acids	Open-chain polyketides	0.79	-0.73	-0.88
			C0745	Amino acids and Peptides	Aminoacids	0.79	-0.73	-0.89
			C6731	Shikimates and Phenylpropanoids	Flavones	0.79	0.74	0.90
C4755	1	ABA	C0989	Fatty acids	Fatty alcohols	0.79	-0.75	-0.80
			C0767	Carbohydrates	Purine alkaloids	0.79	-0.75	-0.74
			C0502	Fatty acids	Wax diesters	0.79	-0.76	-0.80
			C0210	Fatty acids	Wax monoesters	0.79	-0.77	-0.78
			C0572	Amino acids and Peptides	Tripeptides	0.79	-0.77	-0.74
			C2251	Shikimates and Phenylpropanoids	Flavanones	0.79	-0.77	-0.85
			C0644	Alkaloids	Polyamines	0.79	-0.77	-0.86
			C3916	Amino acids and Peptides	Tripeptides	0.79	0.78	0.74
			C2059	Polyketides	Open-chain polyketides	0.79	-0.78	-0.71

			C2130	Shikimates and Phenylpropanoids	Flavones	0.79	0.78	0.74
			C0744	Carbohydrates	Purine nucleosides	0.79	-0.80	-0.83
			C1322	Amino acids and Peptides	Pyridine alkaloids	0.79	0.80	0.84
			C1779	Shikimates and Phenylpropanoids	Furofuranoid lignans	0.79	-0.80	-0.78
			C0065	Shikimates and Phenylpropanoids	Flavones	0.79	-0.80	-0.81
			C3062	Terpenoids	Diacylglycerols	0.79	0.80	0.73
C5600	1	ABA	C5075	Shikimates and Phenylpropanoids	Flavonols	0.78	0.71	0.78
			C0745	Amino acids and Peptides	Aminoacids	0.78	-0.71	-0.89
			C0510	Fatty acids	Open-chain polyketides	0.78	-0.72	-0.88
C7261	1	ABA	C0510	Fatty acids	Open-chain polyketides	0.90	-0.70	-0.88
			C2059	Polyketides	Open-chain polyketides	0.90	-0.70	-0.71
			C1458	Amino acids and Peptides	Cyclic peptides	0.90	-0.70	-0.85
			C1779	Shikimates and Phenylpropanoids	Furofuranoid lignans	0.90	-0.71	-0.78
			C6731	Shikimates and Phenylpropanoids	Flavones	0.90	0.72	0.90
			C5075	Shikimates and Phenylpropanoids	Flavonols	0.90	0.72	0.78
			C0850	Terpenoids	Dammarane and Protostane triterpenoids	0.90	-0.72	-0.71
			C0611	Polyketides	Limonoids	0.90	-0.73	-0.72
			C2251	Shikimates and Phenylpropanoids	Flavanones	0.90	-0.73	-0.85
			C6121	Alkaloids	Capsaicins and Capsaicinoids	0.90	0.74	0.78

			C3062	Terpenoids	Diacylglycerols	0.90	0.74	0.73
			C1291	Amino acids and Peptides	Microcolins and mirabimids	0.90	0.74	0.72
			C0744	Carbohydrates	Purine nucleosides	0.90	-0.74	-0.83
			C1336	Polyketides	Bisnaphthalenes	0.90	-0.74	-0.83
			C0210	Fatty acids	Wax monoesters	0.90	-0.75	-0.78
			C0469	Terpenoids	Quassinoids	0.90	-0.75	-0.72
			C1322	Amino acids and Peptides	Pyridine alkaloids	0.90	0.76	0.84
			C0745	Amino acids and Peptides	Aminoacids	0.90	-0.77	-0.89
			C0767	Carbohydrates	Purine alkaloids	0.90	-0.77	-0.74
			C1074	Alkaloids	Open-chain polyketides	0.90	-0.77	-0.78
			C0065	Shikimates and Phenylpropanoids	Flavones	0.90	-0.77	-0.81
			C0644	Alkaloids	Polyamines	0.90	-0.78	-0.86
			C0502	Fatty acids	Wax diesters	0.90	-0.78	-0.80
			C1841	Fatty acids	Dicarboxylic acids	0.90	-0.78	-0.87
			C0572	Amino acids and Peptides	Tripeptides	0.90	-0.78	-0.74
			C0989	Fatty acids	Fatty alcohols	0.90	-0.78	-0.80
			C7412	Terpenoids	Hydrocarbons	0.90	-0.79	-0.72
			C3916	Amino acids and Peptides	Tripeptides	0.90	0.81	0.74
			C1869	Polyketides	Flavones	0.90	-0.83	-0.92
C6882	1	JA-Ile	C1368	Shikimates and Phenylpropanoids	Arylnaphthalene and aryltetralin lignans	0.87	-0.71	-0.71
C7261	1	IAA	C4690	Terpenoids	Carotenoids (C40, $\leq$ - $\leq$)	-0.87	0.70	-0.76
			C2059	Polyketides	Open-chain polyketides	-0.87	-0.70	0.74

C1458	Amino acids and Peptides	Cyclic peptides	-0.87	-0.70	0.79
C3665	Alkaloids	Phenylethylamines	-0.87	-0.71	0.71
C1779	Shikimates and Phenylpropanoids	Furofuranoid lignans	-0.87	-0.71	0.80
C5075	Shikimates and Phenylpropanoids	Flavonols	-0.87	0.72	-0.76
C0993	Shikimates and Phenylpropanoids	Bisnaphthalenes	-0.87	-0.72	0.85
C1006	Amino acids and Peptides	Dipeptides	-0.87	-0.72	0.77
C0611	Polyketides	Limonoids	-0.87	-0.73	0.81
C2251	Shikimates and Phenylpropanoids	Flavanones	-0.87	-0.73	0.73
C7071	Alkaloids	Pyridine alkaloids	-0.87	-0.74	0.82
C6121	Alkaloids	Capsaicins and Capsaicinoids	-0.87	0.74	-0.71
C3062	Terpenoids	Diacylglycerols	-0.87	0.74	-0.74
C1336	Polyketides	Bisnaphthalenes	-0.87	-0.74	0.81
C0210	Fatty acids	Wax monoesters	-0.87	-0.75	0.85
C0469	Terpenoids	Quassinoids	-0.87	-0.75	0.78
C2853	Amino acids and Peptides	Linear peptides	-0.87	0.76	-0.74
C1368	Shikimates and Phenylpropanoids	Arylnaphthalene and aryltetralin lignans	-0.87	-0.76	0.72
C0120	Alkaloids	Phenylalanine-derived alkaloids	-0.87	0.77	-0.77
C2032	Carbohydrates	Purine nucleosides	-0.87	-0.77	0.72
C1074	Alkaloids	Open-chain polyketides	-0.87	-0.77	0.85
C0065	Shikimates and Phenylpropanoids	Flavones	-0.87	-0.77	0.80
C0644	Alkaloids	Polyamines	-0.87	-0.78	0.87

			C0502	Fatty acids	Wax diesters	-0.87	-0.78	0.76
			C1841	Fatty acids	Dicarboxylic acids	-0.87	-0.78	0.83
			C0568	Alkaloids	Purine alkaloids	-0.87	-0.78	0.83
			C0572	Amino acids and Peptides	Tripeptides	-0.87	-0.78	0.72
			C0989	Fatty acids	Fatty alcohols	-0.87	-0.78	0.81
			C7412	Terpenoids	Hydrocarbons	-0.87	-0.79	0.83
			C3916	Amino acids and Peptides	Tripeptides	-0.87	0.81	-0.79
			C1869	Polyketides	Flavones	-0.87	-0.83	0.71
			C0672	Terpenoids	Cucurbitane triterpenoids	-0.87	-0.86	0.85
C5975	2	SA	C0745	Amino acids and Peptides	Aminoacids	0.82	-0.71	-0.87
C6099	2	SA	C1006	Amino acids and Peptides	Dipeptides	-0.78	0.72	-0.78
			C0210	Fatty acids	Wax monoesters	0.87	-0.71	-0.86
			C2032	Carbohydrates	Purine nucleosides	0.87	-0.71	-0.94
			C0767	Carbohydrates	Purine alkaloids	0.87	-0.71	-0.90
			C1006	Amino acids and Peptides	Dipeptides	-0.84	0.72	-0.77
			C0572	Amino acids and Peptides	Tripeptides	0.75	-0.71	-0.73
C5520	2	JA	C1779	Shikimates and Phenylpropanoids	Furofuranoid lignans	0.75	-0.71	-0.90
			C2130	Shikimates and Phenylpropanoids	Flavones	0.75	0.73	0.80
			C5162	Shikimates and Phenylpropanoids	Bisnaphthalenes	0.74	-0.71	-0.83
			C1482	Fatty acids	Triacylglycerols	0.74	0.71	0.76
			C1368	Shikimates and Phenylpropanoids	Arylnaphthalene and aryltetralin lignans	0.74	-0.71	-0.85

			C0469	Terpenoids	Quassinoids	0.74	-0.73	-0.76
			C1074	Alkaloids	Open-chain polyketides	0.74	-0.73	-0.90
			C5075	Shikimates and Phenylpropanoids	Flavonols	0.74	0.73	0.72
			C0850	Terpenoids	Dammarane and Protostane triterpenoids	0.74	-0.73	-0.85
			C2032	Carbohydrates	Purine nucleosides	0.74	-0.73	-0.94
			C0858	Amino acids and Peptides	Aminoacids	0.74	-0.73	-0.92
			C2130	Shikimates and Phenylpropanoids	Flavones	0.74	0.73	0.80
			C1869	Polyketides	Flavones	0.74	-0.75	-0.78
			C1466	Amino acids and Peptides	Aminoacids	0.74	-0.76	-0.77
			C0767	Carbohydrates	Purine alkaloids	0.74	-0.79	-0.90
			C1004	Alkaloids	Dipeptides	0.74	-0.79	-0.87
			C3137	Shikimates and Phenylpropanoids	Flavonols	0.74	0.80	0.78
			C3665	Alkaloids	Phenylethylamines	0.74	-0.80	-0.90
C5520	2	ABA	C0210	Fatty acids	Wax monoesters	0.88	-0.71	-0.89
			C2032	Carbohydrates	Purine nucleosides	0.88	-0.71	-0.83
			C0767	Carbohydrates	Purine alkaloids	0.88	-0.71	-0.80
C5520	2	JA-Ile	C0210	Fatty acids	Wax monoesters	0.75	-0.71	-0.75
			C2032	Carbohydrates	Purine nucleosides	0.75	-0.71	-0.83
			C0767	Carbohydrates	Purine alkaloids	0.75	-0.71	-0.85
C3899	2	IAA	C3665	Alkaloids	Phenylethylamines	-0.71	-0.70	0.85
			C0510	Fatty acids	Open-chain polyketides	-0.71	-0.71	0.78
			C1466	Amino acids and Peptides	Aminoacids	-0.71	-0.72	0.84
			C5075	Shikimates and Phenylpropanoids	Flavonols	-0.71	0.73	-0.85

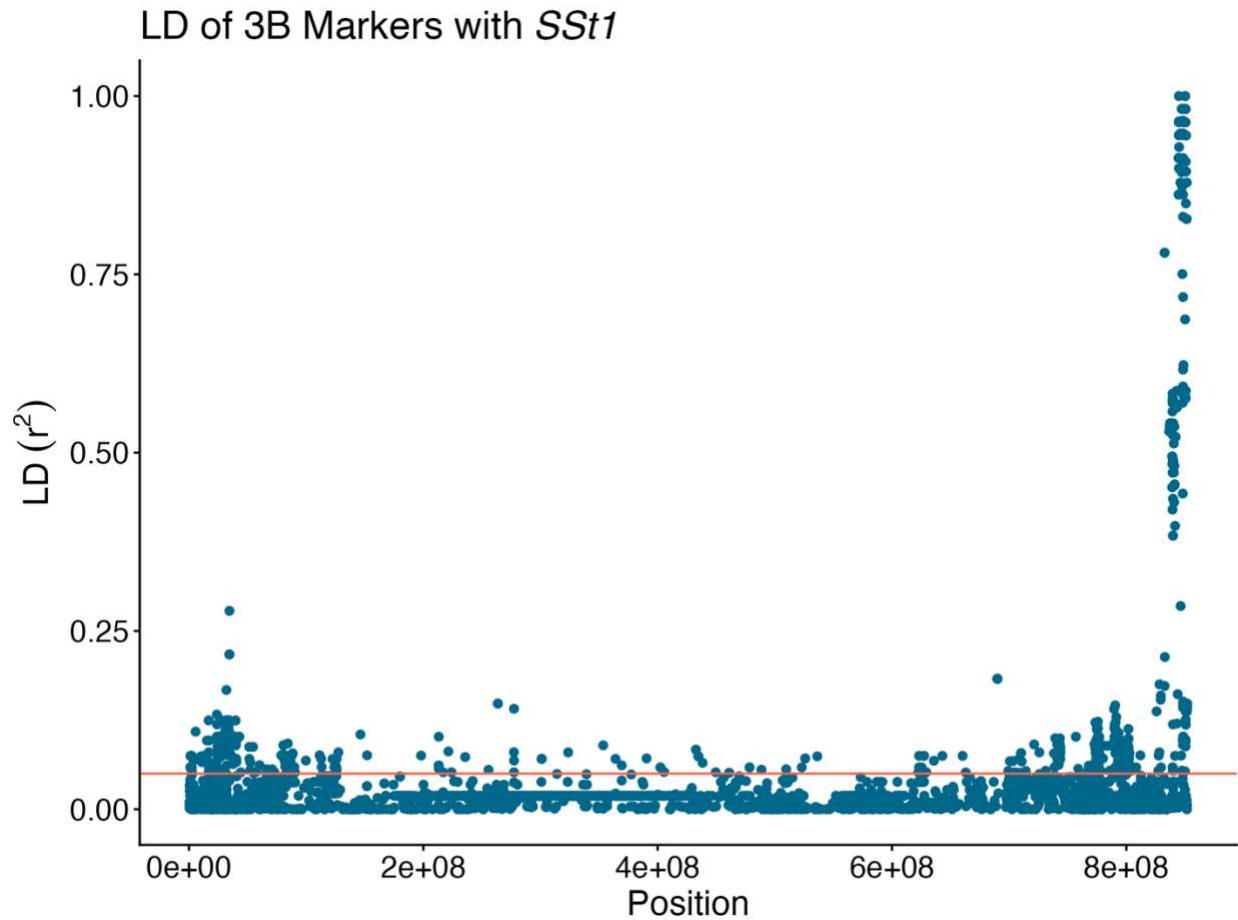
C5520	2	IAA	C0210	Fatty acids	Wax monoesters	-0.76	-0.71	0.77
			C2032	Carbohydrates	Purine nucleosides	-0.76	-0.71	0.83
			C0767	Carbohydrates	Purine alkaloids	-0.76	-0.71	0.75
C6099	2	IAA	C1006	Amino acids and Peptides	Dipeptides	0.80	0.72	0.88
C6153	2	IAA	C0611	Polyketides	Limonoids	0.85	0.71	0.83
			C6731	Shikimates and Phenylpropanoids	Flavones	0.85	-0.72	-0.78
			C0572	Amino acids and Peptides	Tripeptides	0.85	0.79	0.86
C6198	2	IAA	C0572	Amino acids and Peptides	Tripeptides	-0.76	-0.71	0.86
			C0510	Fatty acids	Open-chain polyketides	-0.76	-0.71	0.78
			C1779	Shikimates and Phenylpropanoids	Furofuranoid lignans	-0.76	-0.71	0.87
			C6731	Shikimates and Phenylpropanoids	Flavones	-0.76	0.73	-0.78
			C2130	Shikimates and Phenylpropanoids	Flavones	-0.76	0.73	-0.79
			C0568	Alkaloids	Purine alkaloids	-0.76	-0.74	0.80
C6182	3	SA	C0469	Terpenoids	Quassinoids	0.76	-0.70	-0.76
			C5075	Shikimates and Phenylpropanoids	Flavonols	0.76	0.72	0.73
			C1466	Amino acids and Peptides	Aminoacids	0.76	-0.73	-0.87
			C0120	Alkaloids	Phenylalanine-derived alkaloids	0.76	0.74	0.71
			C1841	Fatty acids	Dicarboxylic acids	0.76	-0.74	-0.79
			C6226	Polyketides	Macrolide lactones	0.76	0.76	0.72
			C7071	Alkaloids	Pyridine alkaloids	0.76	-0.78	-0.76
			C2130	Shikimates and Phenylpropanoids	Flavones	0.76	0.79	0.72

			C0502	Fatty acids	Wax diesters	0.76	-0.79	-0.78
			C0210	Fatty acids	Wax monoesters	0.76	-0.82	-0.74
C6612	3	SA	C5075	Shikimates and Phenylpropanoids	Flavonols	0.71	0.71	0.73
			C3665	Alkaloids	Phenylethylamines	0.71	-0.71	-0.84
			C3137	Shikimates and Phenylpropanoids	Flavonols	0.71	0.72	0.82
			C2130	Shikimates and Phenylpropanoids	Flavones	0.71	0.83	0.72
C6182	3	ABA	C0469	Terpenoids	Quassinoids	0.87	-0.70	-0.83
			C1869	Polyketides	Flavones	0.87	-0.70	-0.88
			C2032	Carbohydrates	Purine nucleosides	0.87	-0.72	-0.76
			C5075	Shikimates and Phenylpropanoids	Flavonols	0.87	0.72	0.87
			C6121	Alkaloids	Capsaicins and Capsaicinoids	0.87	0.73	0.82
			C1466	Amino acids and Peptides	Aminoacids	0.87	-0.73	-0.95
			C0644	Alkaloids	Polyamines	0.87	-0.73	-0.72
			C0858	Amino acids and Peptides	Aminoacids	0.87	-0.74	-0.73
			C1006	Amino acids and Peptides	Dipeptides	0.87	-0.75	-0.80
			C6226	Polyketides	Macrolide lactones	0.87	0.76	0.90
			C0065	Shikimates and Phenylpropanoids	Flavones	0.87	-0.76	-0.73
			C0767	Carbohydrates	Purine alkaloids	0.87	-0.77	-0.83
			C0568	Alkaloids	Purine alkaloids	0.87	-0.77	-0.78
			C2251	Shikimates and Phenylpropanoids	Flavanones	0.87	-0.77	-0.76
			C7071	Alkaloids	Pyridine alkaloids	0.87	-0.78	-0.84

			C0993	Shikimates and Phenylpropanoids	Bisnaphthalenes	0.87	-0.78	-0.91
			C0572	Amino acids and Peptides	Tripeptides	0.87	-0.78	-0.83
			C7412	Terpenoids	Hydrocarbons	0.87	-0.78	-0.77
			C2130	Shikimates and Phenylpropanoids	Flavones	0.87	0.79	0.94
			C0502	Fatty acids	Wax diesters	0.87	-0.79	-0.73
			C0611	Polyketides	Limonoids	0.87	-0.79	-0.74
			C0744	Carbohydrates	Purine nucleosides	0.87	-0.81	-0.73
			C0210	Fatty acids	Wax monoesters	0.87	-0.82	-0.73
C6612	3	ABA	C5075	Shikimates and Phenylpropanoids	Flavonols	0.75	0.71	0.87
			C3665	Alkaloids	Phenylethylamines	0.75	-0.71	-0.92
			C3137	Shikimates and Phenylpropanoids	Flavonols	0.75	0.72	0.84
			C2130	Shikimates and Phenylpropanoids	Flavones	0.75	0.83	0.94
C3752	3	IAA	C1074	Alkaloids	Open-chain polyketides	-0.73	-0.70	0.82
			C0611	Polyketides	Limonoids	-0.73	-0.71	0.75
			C4690	Terpenoids	Carotenoids (C40, $\leq$ - $\leq$)	-0.73	0.72	-0.78
			C2059	Polyketides	Open-chain polyketides	-0.73	-0.72	0.93
			C1368	Shikimates and Phenylpropanoids	Arylnaphthalene and aryltetralin lignans	-0.73	-0.73	0.92
			C0769	Shikimates and Phenylpropanoids	Triketide meroterpenoids	-0.73	-0.73	0.82
			C0672	Terpenoids	Cucurbitane triterpenoids	-0.73	-0.74	0.71
			C1004	Alkaloids	Dipeptides	-0.73	-0.75	0.79
			C0858	Amino acids and Peptides	Aminoacids	-0.73	-0.76	0.83

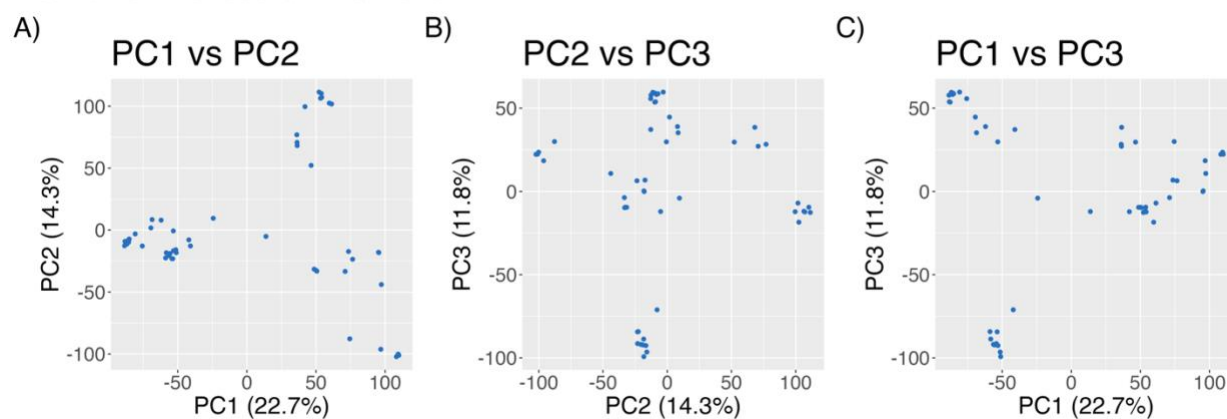
C0989	Fatty acids	Fatty alcohols	-0.73	-0.78	0.82
C1841	Fatty acids	Dicarboxylic acids	-0.73	-0.80	0.74
C0644	Alkaloids	Polyamines	-0.73	-0.83	0.80

CHAPTER VI SUPPLEMENTAL



Supplemental Figure 6.1. Linkage disequilibrium (LD) decay of markers on chromosome 3B with respect to the peak marker for the *SSt1* locus (S3B_844887153). The y-axis shows the r^2 of LD for the marker and the *SSt1* locus, the x-axis indicates the position on the 3B chromosome with the short arm of the chromosome on the left side of the axis and the long arm of the chromosome on the right side of the axis, and the red line indicates an r^2 of 0.05.

PCAs for Molecular Markers



Supplemental Figure 6.2. Principal component analysis for molecular markers showing PC1 vs PC2, PC2 vs PC3, and PC1 and PC3.

Supplemental Table 6.1. NIL group, breeding line names, and line label designation for all genotypes used in the study. The allele at the SSt1 locus is designated as hollow stem (WT) or solid stem (MT).

NIL Group	Genotype	SST1	Line Label
NIL1	CO20SF014W.1	WT	Line 1
	CO20SF014W.2	WT	Line 2
	CO20SF014W.3	WT	Line 3
	CO20SF014W.4	MT	Line 4
	CO20SF014W.5	MT	Line 5
NIL2	CO20SF036R.1	WT	Line 1
	CO20SF036R.2	WT	Line 2
	CO20SF036R.3	WT	Line 3
	CO20SF036R.4	MT	Line 4
NIL3	CO20SF040R.1	MT	Line 1
	CO20SF040R.10	MT	Line 2
	CO20SF040R.11	WT	Line 3
	CO20SF040R.2	WT	Line 4
	CO20SF040R.3	MT	Line 5
	CO20SF040R.4	WT	Line 6
	CO20SF040R.5	WT	Line 7
	CO20SF040R.6	WT	Line 8
	CO20SF040R.7	WT	Line 9
	CO20SF040R.8	MT	Line 10
	CO20SF040R.9	MT	Line 11
NIL4	CO20SF202R.1	WT	Line 1
	CO20SF202R.10	MT	Line 2
	CO20SF202R.2	WT	Line 3
	CO20SF202R.3	MT	Line 4
	CO20SF202R.4	WT	Line 5
	CO20SF202R.5	WT	Line 6
	CO20SF202R.6	MT	Line 7
	CO20SF202R.7	MT	Line 8
	CO20SF202R.8	MT	Line 9
	CO20SF202R.9	MT	Line 10
NIL5	CO20SF250R.1	WT	Line 1
	CO20SF250R.2	MT	Line 2
	CO20SF250R.3	MT	Line 3
	CO20SF250R.4	WT	Line 4
	CO20SF250R.5	WT	Line 5

	CO20SF250R.6	WT	Line 6
	CO20SF250R.7	MT	Line 7
NIL6	CO20SFD001M.1	WT	Line 1
	CO20SFD001M.2	WT	Line 2
	CO20SFD001M.3	MT	Line 3
	CO20SFD001M.4	WT	Line 4
NIL7	CO20SFD034R.1	WT	Line 1
	CO20SFD034R.2	MT	Line 2
	CO20SFD034R.3	MT	Line 3
	CO20SFD034R.4	MT	Line 4
	CO20SFD034R.5	MT	Line 5
NIL8	CO20SFD047R.1	MT	Line 1
	CO20SFD047R.2	MT	Line 2
	CO20SFD047R.3	MT	Line 3
	CO20SFD047R.4	WT	Line 4

20 **Supplemental Table 6.2.** Wald table for across environment model for all traits.

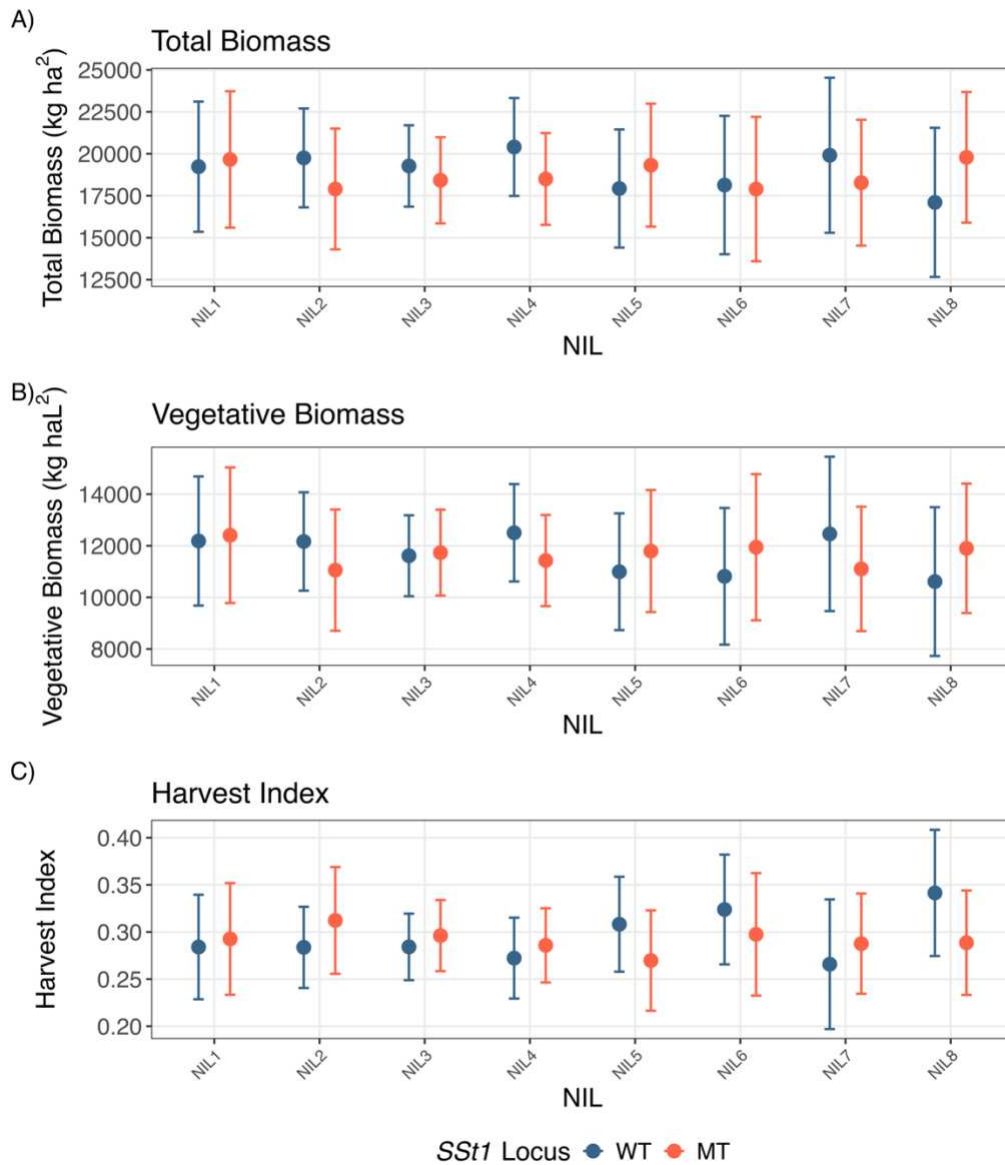
Trait	Variable	Df	Sum of Squares	Wald Statistic	P-Value	Significance
Total Biomass	(Intercept)	1	11216.50	11216.50	0.00E+00	***
	NIL	8	12.11	12.11	1.47E-01	-
	SST1	1	0.38	0.38	5.40E-01	-
	ENV	5	529.14	529.14	0.00E+00	***
	NIL:SST1	8	15.01	15.01	5.89E-02	-
	NIL:ENV	40	52.32	52.32	9.17E-02	-
	SST1:ENV	5	6.00	6.00	3.06E-01	-
	NIL:SST1:ENV	40	28.45	28.45	9.14E-01	-
	residual (MS)	NA	1.00	NA	NA	-
Vegetative Biomass	(Intercept)	1	10139.02	10139.02	0.00E+00	***
	NIL	8	6.29	6.29	6.14E-01	-
	SST1	1	0.30	0.30	5.84E-01	-
	ENV	5	783.28	783.28	0.00E+00	***
	NIL:SST1	8	14.33	14.33	7.35E-02	-
	NIL:ENV	40	53.49	53.49	7.51E-02	-
	SST1:ENV	5	3.14	3.14	6.79E-01	-
	NIL:SST1:ENV	40	31.86	31.86	8.17E-01	-
	residual (MS)	NA	1.00	NA	NA	-
Harvest Index	(Intercept)	1	11490.76	11490.76	0.00E+00	***
	NIL	8	6.89	6.89	5.48E-01	-
	SST1	1	0.51	0.51	4.74E-01	-
	ENV	5	615.46	615.46	0.00E+00	***
	NIL:SST1	8	12.32	12.32	1.37E-01	-
	NIL:ENV	40	63.91	63.91	9.54E-03	**
	SST1:ENV	5	5.54	5.54	3.53E-01	-
	NIL:SST1:ENV	40	32.49	32.49	7.95E-01	-

	residual (MS)	NA	1.00	NA	NA	-
Tillers	(Intercept)	1	13734.18	13734.18	0.00E+00	***
	NIL	8	30.75	30.75	1.56E-04	***
	SST1	1	0.02	0.02	8.84E-01	-
	ENV	5	1513.77	1513.77	0.00E+00	***
	NIL:SST1	8	15.20	15.20	5.54E-02	-
	NIL:ENV	40	133.17	133.17	6.08E-12	***
	SST1:ENV	5	7.21	7.21	2.06E-01	-
	NIL:SST1:ENV	40	46.46	46.46	2.24E-01	-
	residual (MS)	NA	1.00	NA	NA	-
Heading Date	(Intercept)	1	4298728.59	4298728.59	0.00E+00	***
	NIL	8	64.58	64.58	5.83E-11	***
	SST1	1	0.87	0.87	3.50E-01	-
	ENV	4	2334.64	2334.64	0.00E+00	***
	NIL:SST1	8	37.23	37.23	1.04E-05	***
	NIL:ENV	32	151.52	151.52	0.00E+00	***
	SST1:ENV	4	6.33	6.33	1.76E-01	-
	NIL:SST1:ENV	32	43.05	43.05	9.19E-02	-
	residual (MS)	NA	1.00	NA	NA	-
Plant Height	(Intercept)	1	111246.21	111246.21	0.00E+00	***
	NIL	8	13.51	13.51	9.54E-02	-
	SST1	1	0.91	0.91	3.40E-01	-
	ENV	5	896.05	896.05	0.00E+00	***
	NIL:SST1	8	28.06	28.06	4.63E-04	***
	NIL:ENV	40	80.25	80.25	1.65E-04	***
	SST1:ENV	5	7.31	7.31	1.99E-01	-
	NIL:SST1:ENV	40	49.33	49.33	1.48E-01	-
	residual (MS)	NA	1.00	NA	NA	-

Kernel Number Spike ⁻¹	(Intercept)	1	16031.84	16031.84	0.00E+00	***
	NIL	8	11.77	11.77	1.62E-01	-
	SST1	1	0.01	0.01	9.27E-01	-
	ENV	5	288.15	288.15	0.00E+00	***
	NIL:SST1	8	9.77	9.77	2.82E-01	-
	NIL:ENV	40	76.68	76.68	4.29E-04	***
	SST1:ENV	5	6.02	6.02	3.04E-01	-
	NIL:SST1:ENV	40	44.00	44.00	3.06E-01	-
	residual (MS)	NA	1.00	NA	NA	-
Grain Weight Spike ⁻¹	(Intercept)	1	7072.49	7072.49	0.00E+00	***
	NIL	8	13.59	13.59	9.30E-02	-
	SST1	1	0.17	0.17	6.81E-01	-
	ENV	5	296.57	296.57	0.00E+00	***
	NIL:SST1	8	15.49	15.49	5.03E-02	-
	NIL:ENV	40	80.86	80.86	1.39E-04	***
	SST1:ENV	5	11.62	11.62	4.04E-02	*
	NIL:SST1:ENV	40	33.80	33.80	7.44E-01	-
	residual (MS)	NA	1.00	NA	NA	-
Grain Yield	(Intercept)	1	35800.21	35800.21	0.00E+00	***
	NIL	8	37.24	37.24	1.04E-05	***
	SST1	1	1.32	1.32	2.50E-01	-
	ENV	5	1822.91	1822.91	0.00E+00	***
	NIL:SST1	8	33.09	33.09	5.93E-05	***
	NIL:ENV	40	145.26	145.26	7.28E-14	***
	SST1:ENV	5	10.82	10.82	5.50E-02	-
	NIL:SST1:ENV	40	66.89	66.89	4.86E-03	**
	residual (MS)	NA	1.00	NA	NA	-
Test Weight	(Intercept)	1	1297102.46	1297102.46	0.00E+00	***

	NIL	8	24.25	24.25	2.08E-03	**
	SST1	1	2.20	2.20	1.38E-01	-
	ENV	5	2209.79	2209.79	0.00E+00	***
	NIL:SST1	8	18.54	18.54	1.75E-02	*
	NIL:ENV	40	352.40	352.40	0.00E+00	***
	SST1:ENV	5	8.66	8.66	1.23E-01	-
	NIL:SST1:ENV	40	190.89	190.89	0.00E+00	***
	residual (MS)	NA	1.00	NA	NA	-
Thousand Kernel Weight	(Intercept)	1	36790.07	36790.07	0.00E+00	***
	NIL	8	403.46	403.46	0.00E+00	***
	SST1	1	0.10	0.10	7.48E-01	-
	ENV	5	626.00	626.00	0.00E+00	***
	NIL:SST1	8	37.44	37.44	9.54E-06	***
	NIL:ENV	40	78.21	78.21	2.86E-04	***
	SST1:ENV	5	4.15	4.15	5.28E-01	-
	NIL:SST1:ENV	40	66.06	66.06	5.89E-03	**
	residual (MS)	NA	1.00	NA	NA	-
Solidity	(Intercept)	1	17567.56	17567.56	0.00E+00	***
	NIL	8	708.98	708.98	0.00E+00	***
	SST1	1	950.17	950.17	0.00E+00	***
	ENV	5	44.92	44.92	1.50E-08	***
	NIL:SST1	8	209.92	209.92	0.00E+00	***
	NIL:ENV	40	56.06	56.06	4.73E-02	*
	SST1:ENV	5	30.41	30.41	1.23E-05	***
	NIL:SST1:ENV	40	48.29	48.29	1.73E-01	-
	residual (MS)	NA	1.00	NA	NA	-

Estimates for BM, VBM, and HI for NIL Group-by-*SSt1* Locus



Supplemental Figure 6.3. Estimates and confidence intervals for NIL group-by-*SSt1* interactions for (A) total biomass (kg ha⁻¹; BM), (B) vegetative biomass (kg ha⁻¹; VBM), and harvest index (HI). Blue is the hollow stem (WT) allele and red is the solid stem (MT) allele.