

THESIS

ANALYSIS OF WHEAT SPIKE CHARACTERISTICS USING IMAGE ANALYSIS,  
MACHINE LEARNING, AND GENOMICS

Submitted By

Mikayla Hammers

Department of Soil and Crop Sciences

In partial fulfillment of the requirements

For the Degree of Master of Science

Colorado State University

Fort Collins, Colorado

Summer 2022

Master's Committee:

Advisor: Esten Mason

Asa Ben-Hur  
Davina Rhodes  
Nathan Mueller

Copyright by Mikayla Hammers 2022

All Rights Reserved

## ABSTRACT

### ANALYSIS OF WHEAT SPIKE CHARACTERISTICS USING IMAGE ANALYSIS, MACHINE LEARNING, AND GENOMICS

Understanding genetics regulating yield component and spike traits can contribute to the development of new wheat cultivars. The flowering pathway in wheat is not entirely known, but spike architecture and its relationship with yield component traits could provide valuable information for crop improvement. Spikelets spike<sup>-1</sup> (SPS) has previously been positively associated with kernel number spike (KNS) and negatively correlated with thousand kernel weight, meaning a further understanding of SPS could help unlock full yield potential.

While genomics research has improved efficiency over time with the development of techniques such as genotyping by sequencing (GBS), phenotyping remains a labor and time intensive process, limiting the amount of phenomic data available for research. Recently, there has been more interest in generating high-throughput methods for collecting and analyzing phenotypic data. Imaging is a cheap and easily reproducible way to collect data at a specific maturity point or over time, and is a promising candidate for implementing deep learning algorithms to extract traits of interest.

For this study, a population of 594 soft red winter wheat (SRWW) inbred lines were evaluated for wheat spike characteristics over two years. Images of wheat spikes were taken in a controlled environment and used to train deep learning algorithms to count SPS. A total of 12,717 images were prepared for analysis and used to train, test, and validate a basic classification and regression convolutional neural network (CNN), as well as a VGG16 and VGG19 regression model. Classification had a low accuracy and did not allow for an assessment of error margins. Regression models were more accurate. Of the regression models, VGG16 had the lowest mean absolute error (MAE) (MAE = 1.09) and

mean squared error (MSE) (MSE = 2.08), and the highest coefficient of determination ( $R^2$ ) ( $R^2 = 0.53$ ) meaning it had the best fit of all models. The basic CNN was the next well fit model (MAE = 1.27, MSE = 2.61,  $r = 0.48$ ) followed by the VGG19 (MAE = 1.32, MSE = 2.98,  $r = 0.45$ ). With an average error of just above one spikelet, it is possible that counting methods could provide enough data with an accuracy high enough for use in statistical analyses such as genome wide association studies (GWAS), or genomic selection (GS).

A GWAS was used to identify markers associated with SPS and yield component traits, while demonstrating the use of genomic selection (GS) for prediction and screening of individuals across multiple breeding programs. The GWAS results indicated similar markers and genotypic regions underpinning both KNS and SPS on chromosome 6A and spike length and SPS on chromosome 7A. It was observed that favorable alleles at each locus were associated with higher KNS and SPS on chromosome 6A and longer wheat spikes with higher SPS on chromosome 7A. Significant markers on 7A were observed in the region near *WAP01*, the causal gene for SPS on the long arm of chromosome 7A, indicating they could be associated with that gene. GS results showed promise for whole genome selection, with the lowest prediction accuracy observed for heading date ( $r_{gs} = 0.30$ ) and the highest for spike area ( $r_{gs} = 0.62$ ). SPS showed prediction accuracies ranging from 0.33 to 0.42, high enough to aid in the selection process. These results indicate that knowledge of the flowering pathway and wheat spike architecture and how it relates to yield components could be beneficial for making selections and increasing grain yield.

## ACKNOWLEDGEMENTS

It has been an absolute privilege to pursue a graduate degree with an amazing support system. I would like to thank my parents, Rick and Kathy, for their unwavering love and support. Thank you for always encouraging me to be my best self, I could not be who I am without you. I would also like to thank my friends, you are the best family away from home I could ask for. My sincerest thank you to my advisor, Dr. Esten Mason, for his advice, guidance, and patience throughout my master's studies. I will always be thankful to have been given this opportunity. Thank you for always pushing and encouraging me to learn new things, even when I doubted myself. I would also like to thank my committee members Dr. Asa Ben-Hur, Dr. Davina Rhodes, and Dr. Nathan Mueller for their advice and insights as I maneuvered new areas of study. Thank you to the staff at the University of Arkansas Wheat Breeding and Genetics Program: David Moon, Zachary Winn, Dr. Dylan Larkin, and the undergraduate students for guiding me through the beginning of my degree and field days, you helped me feel less confused in the world of agriculture. Last, but certainly never least, I would like to thank the Colorado State University Wheat Breeding and Genetics Staff: Emily Hudson-Arns, Scott Seifert, John Stromberger, Amanda Amsberry, Hong Wang, Meenakshi Santra, Forrest Wold-MicGimsey, Meseret Wondifraw, and Selena Lopez for their constant support and encouragement.

## TABLE OF CONTENTS

ABSTRACT.....	ii
ACKNOWLEDGEMENTS.....	iv
CHAPTER I – LITERATURE REVIEW.....	1
Origins and Production of Wheat.....	1
Production.....	2
Wheat Breeding and Characteristics of Interest.....	2
Wheat Breeding.....	2
Adaptation Genes.....	4
Flowering Pathways – Comparing Wheat to Arabidopsis.....	8
Yield Components and Total Grain Yield.....	9
Spikelets Spike <sup>-1</sup> .....	10
Tools and Strategies for Breeding.....	14
QTLs and QTL Mapping.....	14
Genome-Wide Association Studies.....	16
Genomic Selection.....	17
High Throughput Phenotyping Using Image Analysis.....	19
Machine Learning.....	20
Deep Learning.....	22
REFERENCES.....	24
CHAPTER II – USING IMAGING FOR HIGH-THROUGHPUT PHENOTYPING IN WINTER WHEAT.....	36
INTRODUCTION.....	36
MATERIALS AND METHODS.....	39
Plant Materials.....	39
Experimental Design.....	39
Imaging.....	40

Image Analysis – Image J.....	41
Image Analysis – Deep Learning.....	42
RESULTS.....	44
Classification Model.....	45
Regression Model.....	45
DISCUSSION.....	48
CONCLUSSION.....	51
REFERENCES.....	52
CHAPTER III - GENOME-WIDE ASSOCIATION STUDY OF WHEAT SPIKE CHARACTERISTICS PHENOTYPED VIA IMAGING.....	55
INTRODUCTION.....	55
The Flowering Pathway in Wheat.....	55
Spikelets Spike <sup>-1</sup> .....	56
MATERIALS AND METHODS.....	58
Plant Materials.....	58
Trait Phenotyping.....	58
Genotyping.....	59
Statistical Analysis.....	59
GWAS and GS.....	61
RESULTS.....	66
Analysis of Phenotypic Data.....	66
Summary of Marker Trait Associations.....	70
MTA for Spikelets Spike <sup>-1</sup> and Related Traits.....	71
Candidate Genes for Underlying Identified MTA.....	72
Genomic Prediction of Spike Characteristics.....	73
DISCUSSION.....	73
CONCLUSION.....	75
REFERENCES.....	76

APPENDIX.....	82
Chapter II Supplemental.....	82
Chapter III Supplemental.....	90

LIST OF TABLES

**CHAPTER II**

**Table 2.1** Mean, standard deviation, and range for all 12,717 images prepared for use in the deep learning algorithm.....43

**Table 2.2** The mean and standard deviation of accuracy and loss for the classification model and mean absolute error (MAE) and mean squared error (MSE) means and standard deviations for each of the three regression models. ....46

**CHAPTER III**

**Table 3.1** Summary statistics and heritability of traits measured on the Historical Gulf Atlantic Wheat Nursery in Fayetteville, Arkansas for the 2019 and 2020 seasons and combined across years.....66

**APPENDIX**

**Supplemental Table 1** MrMLM results for all six models for 2019, 2020, and combined analysis.....90

LIST OF FIGURES

CHAPTER I

**Figure 1.1:** A wheat spike showing the alternating pattern of wheat spikelets on the rachis.....11

CHAPTER II

**Figure 2.1** A Phenotyping device allowed for a controlled environment and consistency for photos of wheat spikes. (A) A camera rests on top of the device, which was used on the same red background for each photo. The black interior of the cover and attached light strip create uniform lighting and minimal reflection from camera flash while taking pictures. (B) Controlled environment images of six wheat spikes from each plot. Each spike was inserted into clay by the peduncle, rachis side up. Spikelets were fully visible for each spike to allow for phenotyping. (C) Example input image for specified deep learning algorithms. Each image of six spikes was cropped down to images of individual spikes before having their size reduced and padding the images to ensure identical sizing.....42

**Figure 2.2** Comparisons of the accuracy and loss across iterations for the classification model and the (B) MAE and (C) MSE between the basic, VGG16, and VGG19 regression models over 10 iterations. The performance of each model varies over each iteration where no model was the best model for all ten iterations. VGG16 had the lowest MAE and MSE compared to the basic model and VGG19 model for more iterations than any other regression model.....47

**Figure 2.3** Regressions of predicted spikelet numbers versus true spikelet values in each image for the basic regression CNN, VGG16 regression, and VGG19 regression models. The x-axis is the true values of spikelets present in each image and the y-axis is the value of spikelets predicted by each respective model for each image. Displayed within each graph is the equation of the regression line and the coefficient of determination ( $R^2$ ).....48

CHAPTER III

**Figure 3.1** Distribution of best linear unbiased predictors (BLUPs) for (A) heading date (Julian days)(HD), (B) 1000 kernel weight (g)(TKW), (C) kernels spike<sup>-1</sup> (KNS), (D) spikelets spike<sup>-1</sup> (SPS), (E) spike width (mm)(SW), (F) spike length (mm) (SL), and (G) spike area (mm<sup>2</sup>) (SA) from the Historic Gulf Atlantic Wheat Nursery (HGAWN) for 2019, 2020, and the combined years.....68

**Figure 3.2** Heatmaps of best linear unbiased prediction (BLUP) correlations between heading date (HD), spike area (SA), spike length (SL), spike width (SW), spikelets spike<sup>-1</sup> (SPS), kernels spike<sup>-1</sup> (KNS), and 1000 kernel weight (TKW) for the Historic Gulf Atlantic Wheat Nursery (HGAWN) for the (A) 2019, (B) 2020, and (C) combined years. ....69

**Figure 3.3** (A) SPS across the population by allelic classes of S6A\_614663910, (B) KNS across the population by allelic classes of S6A\_614663910, (C) SPS across the population by allelic classes of S7A\_672045448, (D) SL across the population by allelic classes of S7A\_672045448.

Comparisons of phenotypes to alleles do not show large difference in phenotypes based on genetics.....72

**APPENDIX**

**Supplementary Figure 1.** Accuracy performance across all 10 iterations for 10 epochs of the basic five-layer classification CNN where the blue line is the performance of the training dataset and the orange line shows the performance of the test dataset. Lower accuracies indicate a lower estimation error from the true value.....82

**Supplementary Figure 2.** MAE performance across all 10 iterations for 10 epochs of the basic five-layer regression CNN where the blue line is the performance of the training dataset and the orange line shows the performance of the test dataset. Lower MAEs indicate a lower estimation error from the true value.....84

**Supplementary Figure 3.** MAE performance across all 10 iterations for 10 epochs of the regression VGG16 pre-trained model where the blue line is the performance of the training dataset and the orange line shows the performance of the test dataset. Lower MAEs indicate a lower estimation error from the true value.....86

**Supplementary Figure 4.** MAE performance across all 10 iterations for 10 epochs of the regression VGG19 pre-trained model where the blue line is the performance of the training dataset and the orange line shows the performance of the test dataset. Lower MAEs indicate a lower estimation error from the true value.....88

## CHAPTER I – LITERATURE REVIEW

### Origins and Production of Wheat

Wheat is a cereal crop first cultivated about 10,000 years ago during the Neolithic Revolution, the period when society started settled agriculture instead of solely hunting and gathering. The forms of wheat cultivated at this time were the diploid einkorn (*Triticum monococcum*) and the tetraploid emmer (*Triticum dicoccon*) (P. R. Shewry, 2009), wild varieties hand-picked by farmers for their desirable traits. Domestication of wheat is associated with two main differentiated characteristics resulting from mutation in the wheat genome. The first mutation in the *Br* (brittle rachis) locus on chromosomes 3A and 3B, resulted in a loss of spike shattering at maturity (J. Dubcovsky & Dvorak, 2007). In the wild, the spike shattering trait would allow for wheat kernels and thus progeny to disperse and grow. For domesticated wheat, kernels must be retained by the plant to allow for efficient harvest and utilization. The second domestication characteristic results from a dominant mutation in the *Q* locus on chromosome 5A (Luo, Yang, & Dvořák, 2000), which is controlled by *Tg* (tough glumes) loci. This mutation allows for free threshing of kernels from wheat heads as opposed to the hulled glumes present in early cultivated wheats (P. R. Shewry, 2009). Domesticated wheat has increased seed size, reduced tillers, reduced seed dormancy, and more erect growth compared to ancestor species (J. Dubcovsky & Dvorak, 2007).

Common wheat (*Triticum aestivum*) is an allopolyploid, the result of hybridization of cultivated emmer and the diploid wild grass *Triticum tauschii* (J. Dubcovsky & Dvorak, 2007). Emmer is the source of the A and B genomes in common wheat, while *T. tauschii* is the source of the D genome (J. Dubcovsky & Dvorak, 2007; P. R. Shewry, 2009). This most likely happened several times naturally and was selected by farmers for its superior grain yield (P. R.

Shewry, 2009). The allopolyploid nature of hexaploid wheat gives it several advantages over tetraploid wheat such as a higher adaptability to photoperiods and vernalization, tolerance to soil factors such as salt and low pH, resistance to pests and disease, and greater production potential (J. Dubcovsky & Dvorak, 2007).

### ***Production***

Wheat is one of the three most important cereal crops globally, with annual production exceeding 700 million tons (FAO, 2022). This is due in large part to its wide cultivation range, which spans from 67°N to 45°S (P. R. Shewry, 2009). Ninety-five percent of wheat grown is common wheat, which is used for breads, cookies, and pastries. The other 5% of wheat grown is durum wheat (*Triticum durum*), which is used for pasta, semolina products, and some flatbreads. A small amount of specialty wheats such as emmer and spelt are also grown (Allan, 1987). The five main wheat classes grown are hard red winter wheat, hard red spring wheat, soft red winter wheat, white wheat, and durum wheat. The type of wheat grown depends on environment, for example, spring wheats are adapted to cold, dry areas (Allan, 1987) or cultural and economic need for the crop. Wheat is also nutritionally valuable and provides beneficial dietary components such as protein, B vitamins, dietary fiber, and phytochemicals (Peter R. Shewry & Hey, 2015).

### **Wheat Breeding and Characteristics of Interest**

#### ***Wheat Breeding***

As wheat is one of the most widely used and versatile crops, there are many breeding programs globally that are dedicated to cultivar improvement. Adaptation is key for a wheat cultivar to be productive and profitable. Field trials are generally done in sites that include a

variety of different soil types, season times and temperatures, and precipitation events, across a breeding program's target production area (Allan, 1987). There are three main gene systems that influence adaptation, including the regulation of plant height (*Rht*), photoperiod response (*Ppd*) genes, and vernalization response (*Vrn*) (Blake et al., 2009). Collectively, these loci influence many quantitative traits that determine the ability of wheat to thrive in given environment.

Another factor that varies between environments is pests and disease, and resistant varieties are the most cost-effective means of control. Genetic resistance reduces the need for application of fungicides and insecticides, reducing overall input cost for a given crop (Wiseman & Webster, 1999). Resistance refers to the slowed time of a disease development (Allan, 1987) or the ability of a cultivar to produce more high quality output compared to non-resistant cultivars at the same infestation or infection levels (Wiseman & Webster, 1999). If a crop is less susceptible, the producer will not need to spend as much on pesticides or fungicides to ensure a healthy and high yielding crop. Vertical resistance, or resistance to a single pathogen strain, is normally a short-term solution. There is a recent push towards long-term or general resistance, which normally requires a plant to be resistant to all strains of a pathogen.

Breeding is necessary for both grain-yield and end-use quality. Enhancing the ability of wheat to tolerate extreme heat, drought, flooding, or nutrient deficiencies is vital to future wheat production under current and future climate variability and for more stable grain yield production against environmental elements that may not be consistent year to year (Allan, 1987). Arguably the most important characteristic to producers is the grain yield of a given cultivar. Breeding for grain yield is difficult due to the complexity of heritability. Grain yield is determined by a multitude of components and has a low to moderate heritability and is also strongly affected by environment (Allan, 1987). Released cultivars that show improved traits must still have

comparable yields compared to competitors, making genetic improvement for yield a large focus. Grain quality will influence how a given cultivar performs in the market. It can range from factors such as protein or mineral content to how the wheat performs in its given end use. For example, kernel hardness will affect the ability of wheat to be milled into flour. The ease of milling grain to flour and the percentage of grain that can be milled to flour increases value to end-users.

The development of stress tolerant and high yielding wheat cultivars is needed to meet future production requirements. The current global population of 7.7 billion people is projected to reach nearly 10 billion people by 2050 (UNIES, 2019). In order to meet growing caloric demands, it is estimated that the current annual yield improvement of 0.9 percent must increase to 2.4 percent annually to double wheat production by 2050 (Ray, Mueller, West, & Foley, 2013). Therefore, significant genetic gains must be made for long term food security for the growing population (Foley et al., 2011) and for increased consumption of meat and dairy (Godfray et al., 2010).

### ***Adaptation Genes***

Variation in *Ppd*, *Vrn*, and *Rht* loci allow for wheat to be adapted to a wide range of growing regions while also impacting grain yield and yield component traits. Photoperiodism is a plant's reaction to dark periods, resulting in flowering (Evans, 2016). Photoperiod allows plants to adapt to night length across a variety of latitudes (Allan, 1987), allowing for optimal flowering time based on geographical location (Shaw, Turner, & Laurie, 2012). Wild-type wheat is a long day (LD) plant, meaning it flowers more rapidly as day length increases (Shaw et al., 2012). Photoperiod is mainly controlled by the *Ppd-1* locus, with homoeologous copies of the gene in the short arm of chromosomes 2A, 2B, and 2D (Law, Sutka, & Worland, 1978; Snape,

Butterworth, Whitechurch, & Worland, 2001). The three genome alleles each effect photoperiod sensitivity in varying degrees (Bentley et al., 2011). *Ppd-D1* has the largest effect on photoperiod requirement, followed by *Ppd-B1*, and *Ppd-A1*, respectively (Chen et al., 2013; Worland, 1996). A photoperiod sensitive cultivar will require long day lengths to initiate flowering whereas a photoperiod insensitive (PI) cultivar reacts to temperature instead of day. (Shaw et al., 2012). Insensitive genotypes are represented by an “a” and are caused by deletions for both *Ppd-D1a* and *Ppd-A1a* (Beales, Turner, Griffiths, Snape, & Laurie, 2007; Wilhelm, Turner, & Laurie, 2008). While mutations for PI alleles are not found in *Ppd-B1a*, PI can be caused by an increased number of copy number variations (CNV) (Díaz, Zikhali, Turner, Isaac, & Laurie, 2012). Insensitive alleles are associated with more rapid flowering, which is associated with earlier expression of the wheat flowering time locus (*TaFT1*). Sensitive allele *Ppd-B1b* was associated with higher grain yield (Addison et al., 2016). Photoperiod genes regulate flowering time by responding to environmental stimuli, but are not the only adaptation gene that fills this role (Kamran, Iqbal, & Spaner, 2014).

The second major determinant of adaptation is vernalization requirement which is controlled in large part by the *VRN1*, *VRN2*, and *VRN3* genes (Jorge Dubcovsky et al., 2006; Yan et al., 2006). Vernalization will prevent winter wheat from flowering too soon in the season, which would cause freeze damage to the meristem and flowers (Yan et al., 2003). This adaptation alters the plants cold period requirements for flowering in different environments (Allan, 1987).

The vernalization requirements for a plant are partially determined by *VRN1* alleles in the 5A, 5B, and 5D chromosomes (X. Zhang et al., 2008). The candidate gene for *VRN1* is the MADS-box gene *API* (APETALA1) (Yan et al., 2003). *API* is a meristem identity gene in

Arabidopsis required for the transition from vegetative to reproductive growth. It was also found that at least two weeks of vernalization was required for *API* transcription in winter wheat but not spring wheat (Yan et al., 2003). Transcripts were observed after a plant was moved from a cold room (4°C) to room temperature, indicating that *API* is not a cold stress gene. The second vernalization gene, *VRN2*, has been mapped to chromosome 5A (Yan, 2004). The candidate gene for *VRN2* had similarities with CO-like proteins in Arabidopsis and was named *ZCCT1*. *ZCCT1* transcripts were present in the apices before vernalization but not after (Yan, 2004). The reduction of *ZCCT1* transcripts and increase in *API* transcripts is consistent with epistatic interactions between *VRN1* and *VRN2* (Yan, 2004; Yan et al., 2003). Allelic variation in *ZCCT1* was also associated with a spring growth habit, indicating the role of *ZCCT1* in differentiating winter and spring classes (Yan, 2004). The third vernalization gene *VRN3* was mapped to the short arm of chromosome 7B (Yan et al., 2006). *VRN3* is in complete linkage with the orthologous Arabidopsis flowering time (FT) gene. There is also a relationship between polymorphisms in FT, transcript levels, and flowering time of the plant (Yan et al., 2006)

*VRN* genes work together in a regulatory pathway based on temperature and photoperiod. Before vernalization requirements have been met, *VRN2* is upregulated to suppress flowering while *VRN1* and *VRN3* are downregulated (Jorge Dubcovsky et al., 2006) *VRN2* will be downregulated after vernalization requirements are reached and there is a transition from SD to LD, signaling the end of winter months (Yan et al., 2006). Once long days are reached, *VRN3* will be upregulated and will upregulate *VRN1* in return, initiating flowering of the plant. Upregulation of *VRN1* will in turn downregulate *VRN2*, completing the regulatory feedback loop (Yan et al., 2006). Wheat responds linearly to vernalization until its requirement is reached, and the vernalization requirement increases with copy number variation of the *VRN* genes (Díaz et

al., 2012). If a cultivar has a dominant *Vrn1* and *Vrn3* allele, it will not have a vernalization requirement and is a spring wheat. However, if a cultivar has a dominant *Vrn2* allele it will be a winter wheat (Kamran et al., 2014). Varying alleles for these traits cause variance in growth stages and timing of flowering between cultivars and could potentially impact yield traits.

The third growth determinant of adaptation is the effect of semi-dwarfing (*Rht*) genes that became popular after the Green Revolution (Lozada [Unpublished], Carter, & Mason). *Rht* genes improved wheat productivity by increasing resistance to lodging in combination with increased harvest index (Gale & Youssefian, 1985). Due to the shorter stature of modern cultivars, a larger number of assimilates produced by photosynthesis are available to supply the developing grains as opposed to the stems or vegetative tissue (Chapman, Mathews, Trethowan, & Singh, 2006). Dwarfing genes are identified as either gibberellin (GA) sensitive or GA-insensitive (Gale & Youssefian, 1985). They are categorized between sensitive and insensitive based on their response to applied GA phytohormones, a class of hormones characterized for controlling stem elongation (Pearce et al., 2011).

A large number of *Rht* genes have been identified across the wheat genome, each with varying effects on plant height. *Rht-B1* and *Rht-D1* code for DELLA proteins, which are regulators that repress growth before an increase in GA causes targeted degradation (Pearce et al., 2011). Reduced GA sensitivity has been detected for mutations altering the function of DELLA proteins in Arabidopsis and rice, indicating a potential cause for dwarfism in wheat (Pearce et al., 2011). Further dwarfing alleles for *Rht-B1* and *Rht-D1* were identified, each having varying effects on plant height. A more severe dwarfing allele, *Rht-B1c*, is caused by a frame insertion causing an addition of amino acids to the coded DELLA protein (Pearce et al., 2011). Markers developed for *Rht-B1b* on chromosome 4B and *Rht-D1d* on chromosome 4D are

“perfect” markers representing the base pairs responsible for dwarfing and are associated with an increased yield (Ellis, Spielmeier, Gale, Rebetzke, & Richards, 2002).

### ***Flowering Pathways- Comparing Wheat to Arabidopsis***

Flowering pathways have been extensively studied in *Arabidopsis thaliana*. Like winter wheat, *Arabidopsis* also has a vernalization requirement, meaning it could be useful in the understanding of the wheat flowering pathway. *Arabidopsis* has four major pathways involved in flowering; these include vernalization and photoperiod pathways, which respond to environmental stimuli, and autonomous and gibberellin (GA) pathways which function independently. These pathways are defined as regulating repressors that activate flowering (Boss, Bastow, Mylne, & Dean, 2004). Vernalization and the autonomous pathway downregulate *FLOWERING LOCUS C (FLC)*, a MADS-box gene involved in flowering repression, to help initiate flowering (Searle et al., 2006). *Arabidopsis* follows a circadian clock pattern by using photoreceptors to regulate flowering (Carré, 2001). The photoperiod pathway responds to long days to initiate floral transition by increasing *CONSTANS (CO)* expression which activates *FT* expression (Thomas, 2006). Once *FT* has been activated it interacts with the bZIP transcription factor FD and activates meristem identity genes *APETALA1 (API)* and *LEAFY (LFY)* (Abe, 2005). *FT* and the GA pathway increase expression of *SUPPRESSOR OF OVEREXPRESSION OF CO1(SOCl)*, which in turn activates *LFY* expression through *AGAMOUS-LIKE 24(AGL24)*, a MADS-box transcription factor (Moon, Lee, Kim, & Lee, 2005). These mechanisms allow for optimal flowering time in *Arabidopsis*. For reproduction, the genetics of the flowering cycle passed on in the gametes will allow the new plant to flower in the appropriate time for their environment (Boss et al., 2004)

The main photoperiod and autonomous pathways are conserved between *Arabidopsis* and wheat, with earliness per se (*Eps*) in wheat being the equivalent of the autonomous pathway in *Arabidopsis* (Shitsukawa et al., 2007), with potential evolutionary differences between species in flowering time (*FT*) (Higgins, Bailey, & Laurie, 2010). While *FLC* is the main flowering repressor in *Arabidopsis*, it is absent in cereals where *VRN2* acts largely as a repressor (Yan, 2004). *FLC*, *SOC1*, *API*, *AGL24* and other MADS-box genes in *Arabidopsis* play an important role in the transition to flowering, indicating similar genes may be important in other species. Unlike wheat *API*, wheat *SOC1* (*WSOC1*) expression was not effected by photoperiod or vernalization and was upregulated by application of gibberellin A<sub>3</sub> (GA<sub>3</sub>) (Shitsukawa et al., 2007). This indicated that *WSOC1* could be part of a GA pathway in wheat, as *SOC1* is part of the GA pathway in *Arabidopsis*. While the flowering pathways of wheat and *Arabidopsis* are not identical, understanding the pathways for *Arabidopsis* could allow us to better understand the underlying mechanisms in other species.

### ***Yield Components and Total Grain Yield***

Grain yield in wheat is quantitatively inherited. However, according to heritability estimates, morphological traits that affect grain yield may be more heritable than yield itself. (Erkul, Unay, & Konak, 2010) Narrow-sense heritability estimates, denoted by  $h^2$  (Bernardo, 2020), indicate the magnitude of genetic gain transmitted from the parents to the progeny (Erkul et al., 2010). Narrow-sense heritability is the amount of allelic variation as a proportion of total expressed phenotypic variation (Bernardo, 2020) Higher heritability estimates help the breeder know what traits will likely pass on to progeny (Khan, Salim, & Ali, 2003). Total grain yield is the combination of each yield component trait. Since many yield component traits have negative

correlations, it can become difficult to increase yield in newer cultivars and current research focuses on looking for the genetic components of individual traits.

The three major adaptation gene systems not only determine growth and development in a variety of environments, they also play a role in the expression of yield components.

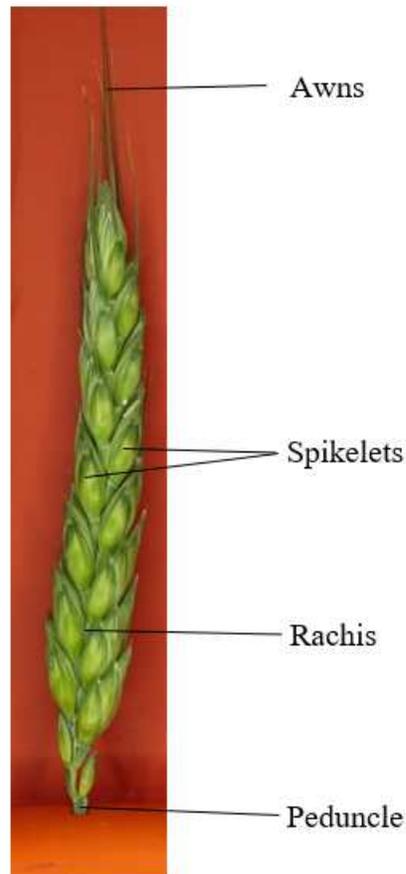
Photoperiod sensitivity, vernalization, and plant height underpin spike development and have been linked to yield component traits (Cao et al., 2019). The quantitative trait loci (QTL) *Vrn-A1* on the long arm of chromosome 5A has been associated with yield components such as spikelet spike<sup>-1</sup>(SPS), number of tillers, and 50 grain weight (K. Kato, Miura, & Sawada, 2000). *Ppd-D1* and *Rht-B1* genes have also been identified as candidate genes for QTL rich clusters for thousand kernel weight (TKW), number of kernels spike<sup>-1</sup>, and spike number (Cao et al., 2019), indicating the pleiotropic effects of these loci.

While there are genetic components associated with individual yield components, the phenotypic traits are also correlated with each other. Spike length is positively correlated with both SPS and grain number (Cao et al., 2019) as well as heading date and flowering time (Alqudah et al., 2020). Spikelet number and grain number are also positively correlated; however, spikelet spike<sup>-1</sup> and thousand kernel weight are negatively correlated (Cao et al., 2019). Grain number is also negatively correlated with TKW and grain size, indicating if a plant spends more energy producing more kernels, it may have less energy and assimilate to produce larger sized kernels.

### ***Spikelets Spike<sup>-1</sup>***

As wheat spikes (Fig. 1.1) are the flower of the plant, there are multiple physiological and environmental factors that affect their formation. Spikelets spike<sup>-1</sup> (SPS) can be defined as the

groups of florets positioned at each rachis node with an alternating pattern up the spike (Koppolu & Schnurbusch, 2019). SPS is positively correlated with yield and is potentially an important morphological trait for increasing total yield (Chen, X. Cheng, et al., 2020) with a linear relationship with total grain yield up to 32 spikelets (Rawson, 1970).



**Figure 1.1.** A wheat spike showing the alternating pattern of wheat spikelets on the rachis.

*Vrn*, *Ppd*, and earliness per se (*Eps*) gene systems control most aspects of pre-anthesis growth phases in wheat (Herndl, White, Graeff, & Claupein, 2008). Each gene system contributes to variation of heading time with 70 to 75% coming from *Vrn*, 20 to 25% from *Ppd*, and approximately 5% coming from *Eps* (Stelmakh, 1998). Vernalization is the main

determinant of heading date and is a regulatory factor in the transition from vegetative growth to reproductive growth. The *Vrn-A1* gene has pleiotropic effects on total yield, tiller number, grain weight, and spikelet number, indicating the importance of flowering time (K. Kato et al., 2000). Photoperiod alleles also affect flowering, meaning they also play a role in SPS (Chen, Cheng, et al., 2020). The *Ppd-1*(*Ppd-1a*) insensitive allele accelerates the spikelet initiation rate, thus decreasing SPS (Z. Chen et al., 2020; Zhaoyan Chen et al., 2020; Wurchum, Leiser, Langer, Tucker, & Longin, 2018). *Ppd-1* and vernalization alleles are responsible for regulating *FLOWERING LOCUS T* (FT) expression, a heading date gene. FT can manipulate SPS by altering the duration of the spikelet initiation phase, which may also be influenced by *Eps*. Unlike *Ppd* and *Vrn*, *Eps* genes respond independently of environmental stimuli (Worland, 1996) and make up variation that occurs after *Ppd* and *Vrn* requirements have been fulfilled (H. Kato, Taketa, Ban, Iriki, & Murai, 2001). Since *Eps* has such a small effect on FT, it is usually mapped using QTLs rather than major genes (K. Kato, Miura, & Sawada, 1999). There is evidence supporting the influence of heading date on SPS and spikelet differentiation (Z. Chen et al., 2020; Miura & Worland, 1994). A study using near isogenic lines (NILs) showed that plants with the late *Eps* allele *Eps-A<sup>m</sup> 1-1* produced 8.7 more SPS than plants with the early allele (Lewis, Faricelli, Appendino, Valárik, & Dubcovsky, 2008). The complete flowering pathway of wheat is unknown, however, information from other species and homologs aid in the understanding.

Once the plant is at an appropriate time to flower, other genes play a role in determining inflorescence architecture. Once flowering has been initiated, the shoot apical meristem (SAM) transitions into an inflorescence meristem (IM)(Boss et al., 2004). The IM in grasses initiates floral meristems (FMs) where the florets are formed. Florets contain the reproductive organs of

flowers. A set of multiple florets occur at each rachis node are surrounded by leaf-like bracts called glumes and make up one spikelet (Koppolu & Schnurbusch, 2019; Sakuma, Salomon, & Komatsuda, 2011). The IM is determinate and becomes the terminal spike meristem (Koppolu & Schnurbusch, 2019). There are several genes involved in spike morphology, potentially having an impact on yield components.

Adaptation genes related to flowering are known to have pleiotropic effects on spike morphology, however, a variety of candidate genes have also been discovered. One cloned gene identified by multiple studies is the Q gene, a domestication gene coding for an APETALA2(AP2) family transcription factor (Z. Chen et al., 2020). The Q gene has been linked to ear morphology and was identified as a QTL for grain yield and ear grain weight (K. Kato et al., 2000). Cultivars containing mutations resulting in loss of Q function exhibit reduced spikelet numbers and increased florets spikelet<sup>-1</sup> (Z. Chen et al., 2020). *TaMOC1*, the wheat ortholog of rice *MOC1* regulates auxiliary stem initiation and has been linked to an increased number of SPS in several environments (B. Zhang et al., 2015). *TaMOC1-7a HapH* has been identified as haplotype of the gene with positive contributions to SPS. *TaAPO-A1* is another candidate gene that is an ortholog from rice ABERRANT PANICLE ORGANIZATION (APO). *TaAPO-A1* was identified as a candidate gene for several yield component traits including TKW, kernel spike<sup>-1</sup>, and SPS (Cao et al., 2019). One spikelet rachis node<sup>-1</sup> is standard in wheat, although supernumerary spikelets still occur on some heads (Dobrovolskaya et al., 2015) While multiple QTL studies have been done for SPS, there are a lacking number of major QTL as well as limited research of introgression of candidate genes.

Number of SPS is significantly positively correlated with number of kernels spike<sup>-1</sup>, however, kernels spike<sup>-1</sup> is negatively correlated with TKW (Cao et al., 2019). With more

assimilate being put towards creating a larger number of kernels, there is less available to increase the size of the kernels. In general, cultivars that produce a higher number of kernels will have a lower TWK. Since TKW is an economically important trait, a higher value cultivar would increase SPS while maintaining or increasing TKW. It has been indicated that this may be possible through gene pyramiding (B. Zhang et al., 2015). Sucrose non-fermenting1-related protein kinase (SnRK) is a serine/threonine specific protein kinase involved in signal transduction pathways of plants. SnRK can also affect carbon metabolism, thus increasing TKW. Gene pyramiding TaSnRK2.10-C and TaMOC1-7A (*HapH*) resulted in an increase for both SPS and TKW in eight out of 10 environments. The combination of both favorable alleles may allow for the improvement of both yield component traits instead of having the typical negative relationship (B. Zhang et al., 2015).

## **Tools and Strategies for Breeding**

### ***QTLs and QTL Mapping***

Breeders use different methods to aid the selection of parental lines for new populations or for gene introgression through marker assisted backcrossing. The first step in many genetic studies is to identify quantitative trait loci (QTL) associated with a trait of interest, utilizing DNA markers, usually single nucleotide polymorphisms (SNPs) (Collard, Jahufer, Brouwer, & Pang, 2005), to construct a linkage map of the genome (M. E. Goddard & Hayes, 2007). Linkage maps help identify chromosome regions that have QTL for quantitative traits or find regions controlling simple traits (Collard et al., 2005). The presence or absence of these markers and their effect on phenotypic measurements can be used to locate QTL associations with quantitative traits. A QTL can be considered either major or minor depending on its estimated effect on the phenotype (Collard et al., 2005). While major QTLs are generally considered QTLs

explaining 10% or more of the phenotypic variation, they are less common. Most variation is explained by the combination of several minor QTL. Markers that are positionally close to the gene of interest have high linkage disequilibrium (LD) and have lower recombination frequencies, meaning they are more likely to be inherited together. Recombination frequencies greater than 0.5 indicate that the trait and gene region are unlinked. They are either far apart on the same chromosome or located on a separate chromosome. QTLs that are closely linked can be utilized for marker assisted selection.

Once a QTL is located, high resolution mapping may be performed by adding more markers in the QTL region to locate the specific gene or genes associated with the traits. Significant markers must also be validated across different cultivars and populations with differing parental lines to ensure their efficacy to predict the phenotypic outcome. As gene markers are identified, they can determine parental combinations that are more likely to have desired traits (Collard et al., 2005). MAS can be twice as effective as phenotypic selection and provides several advantages (Larkin, Lozada, & Mason, 2019). MAS allows you to select for traits that are difficult to phenotype due to expense or low expression, trait selection that relies on phenotyping in certain environments or growth stages and allows selection of either monogenic traits or quantitative traits. However, it is difficult to validate QTLs across different environments and breeding programs, meaning multiple mapping populations would need to be developed by breeding programs in differing regions for selection to be accurate. Phenotyping a multitude of lines is time consuming for breeders and comes at a large cost. Because of these shortcomings, other methods are preferable for the targeted selection of quantitative traits.

## ***Genome-Wide Association Studies***

Another mapping method used for identifying marker trait associations is genome wide association studies (GWAS). A GWAS detects associations between genotypes and phenotypes within a population (Visscher et al., 2017). Instead of measuring the contribution of QTL on the phenotype, a GWAS allows genes associated with a phenotype to be located by whole genome genotyping using single nucleotide polymorphisms (SNPs). Marker trait associations (MTA) can be identified using GWAS with diverse germplasm, which differs from the biparental populations used for locating QTLs (Hamblin, Buckler, & Jannink, 2011). A GWAS relies on linkage disequilibrium between SNPs and an association with a measured phenotype (Visscher et al., 2017).

One study conducted a GWAS for 22 traits using 10,653 significant SNPs from 96 wheat lines (Alqudah et al., 2020). Three of these traits were related to spike morphology. Flowering time and spike length each had 5 MTAs detected while spikelets spike<sup>-1</sup> (SPS) had less than 5. Twenty-four SNPs in the study had intra-chromosomal interactions that controlled most of the agronomic traits phenotyped. Of the three spike morphology traits, flowering time was the only trait associated with these 24 SNPs, three of which occurred on chromosome 2A while the fourth occurred on chromosome 7A. A genotype-phenotype network showed relationships between SPS, spike length, grain number, and flowering time. Spike length was also related to spike weight and flowering time was also related to plant height and heading date. Another study conducted a GWAS for five spike morphology traits using a population of 25 synthetic hexaploidy wheat lines, 80 landraces, and 87 cultivars. One hundred and eighty-four significant SNPs were detected, with 51 and 28 related to spike length and SPS respectively (Liu et al., 2018). SNP clustering from this study identified two haplotypes for spike length on

chromosomes 2A, 2B, and 2D. Haplotypes Hsl-2A-2 and Hsl-2B-2 had significantly longer spikes than their counterparts, Hsl-2A-1 and Hsl-2B-1. However, there was no significant difference in spike length between the Hsl-2D-1 and Hsl-2D-2 haplotypes (Liu et al., 2018). A spike length SNP cluster on 6A contained four associated haplotypes, where haplotype Hsl-6A-4 displayed significantly longer spike than haplotypes Hsl-6A-1~3. A kernel spike<sup>-1</sup> and SPS cluster on 7B included two associated haplotypes. Hkps/sn-7B-2 was found to be the advantage haplotype over Hkps/sn-7B-1, but was only identified in landraces and none of the cultivars examined (Liu et al., 2018). These studies indicate the success of utilizing GWAS for identifying genetic regions associated with spike traits, which could allow breeders to further understand the genetic background of spike morphology.

### ***Genomic Selection***

Genomic selection (GS) is a modified form of MAS in which markers across the genome are used to identify breeding values based on genotypes (Ganal, Plieske, Hohmeyer, Polley, & Order, 2019). The first step in genomic selection is developing a training population. This is created using a panel of genotyped and phenotyped developed using genome wide markers (Larkin et al., 2019). The training population is then used to train a model that calculates genomic estimated breeding values (GEBVs). GEBVs account for all markers within the genotype, allowing for a larger amount of variation to be captured in GS compared to MAS (Newell & Jannink, 2014). The GEBVs are then used to make selections from a validation population which has been genotyped but not phenotyped. The more closely related the training population is to the validation population, the more accurate GS will be (Spindel et al., 2015). This method allows for progeny with promising genetics to be retained prior to phenotyping each line (Ganal et al., 2019).

Genomic selection (GS) first gained traction in animal breeding programs (Meuwissen, Hayes, & Goddard, 2001) until high-throughput genotyping allowed for better implementation in plant breeding programs. It is more cost effective to implement GS in earlier generations, making the accuracy of prediction important (Bassi, Bentley, Charmet, Ortiz, & Crossa, 2016). For success in any given crop, the model, training population, linkage disequilibrium, and selection candidates need to be adjusted for the highest accuracy (Goddard, 2009). Selection response depends on training population size and relatedness of the training population to the validation population (Lozada, Mason, Sarinelli, & Brown-Guedira, 2019). When the markers and QTL are closely linked, the size of the training population will determine accuracy (Goddard, 2009). However, if they are not closely linked, accuracy needs to be improved by having a training population more closely related to the selection candidates. GS has been successfully implemented for quantitative traits in wheat, each with varying results. Multi-environment and multi-trait analysis GS models have been generated for selection of wheat characteristics (Ward et al., 2019). A genomic best linear unbiased prediction (GBLUP) multi-environment model was tested against one cross validation set that would test accuracy under addition of new genotypes (CV1) and a second cross validation set that would test accuracy is phenotypes for some genotypes were not collected across all environments in a given year (CV2). The GBLUP models using CV1 and CV2 had coefficients of determination of 0.42 and 0.87 respectively. The model using CV1 had a mean prediction ability of 0.37 for grain yield across environments while the model using CV2 were greater than 0.90 (Ward et al., 2019). Even though grain yield prediction was lower using the CV1, the model still had prediction abilities of 0.55 or higher for other characteristics such as plant height, grain test weight, and thousand kernel weight (TKW) (Ward et al., 2019). These higher prediction abilities still demonstrate the potential of using GS

for yield related traits. While these results show promise for using GS as a genomic prediction method for complex traits in crops, accurate phenotypic data for training populations remains key for increased model accuracy (Bhat et al., 2016).

### ***High Throughput Phenotyping Using Image Analysis***

A phenotype is the resulting characteristic of a genotype and is needed to observe genetic associations with physiological outcome. The development of next-generation sequencing has greatly increased the efficiency of genomics and accuracy of analysis for plant breeding (Koboldt, Steinberg, Larson, Wilson, & Mardis, 2013). While the increased number of sequenced genomes is a large advancement for the genomics side of research, phenotyping still lags behind (Houle, Govindaraju, & Omholt, 2010). Since phenotypic information is needed to discover marker-trait associations, GWAS and MAS methods in breeding are negatively affected by the phenotyping bottleneck (Furbank & Tester, 2011; Minervini, Scharr, & Tsafaris, 2015). There has recently been a push to create high-throughput phenotyping techniques to increase the rate at which we can understand phenotypes or pleiotropy, a single gene producing several unrelated effects, responsible for complex traits in plants such as yield (Houle et al., 2010). Advancements in phenotyping technology can also help increase the amount of data used to help us understand complex traits such as yield.

Using imaging as a tool for high-throughput phenotyping has become more accessible due to the lower cost of cameras and open-source coding and tools for required analysis (Gehan & Kellogg, 2017). Imaging can be used at macroscopic and microscopic level and with regular cameras or infra-red cameras depending on the scientific question being answered (Noah Fahlgren Maximilian Feldman Malia, 2015). Images can be taken at a single maturity point or may utilize time lapse imaging for things such as color, growth, or leaf shape using a Raspberry

Pi, a credit card sized computer that can streamline collection and analysis of image data (Minervini, Giuffrida, Perata, & Tsaftaris, 2017; Tovar et al., 2018). Depending on funding and ingenuity, phenotyping devices can be generated in lab by researchers for little to no cost compared to paying private companies for phenotyping services, making phenotyping using imaging accessible (Gehan & Kellogg, 2017).

While adequate and appropriate imaging techniques are crucial, data analysis is the more difficult aspect of high-throughput phenotyping. Like most genomics work, a small background in basic computing is beneficial. The first step is to be able to differentiate the relevant aspects of the image, such as leaves of spikelets, from the background (Gehan & Kellogg, 2017). This can be done using open source programs such as ImageJ, an imaging platform that can be customized using macros (Caroline, Wayne, & Kevin, 2012; Hartmann, Czauderna, Hoffmann, Stein, & Schreiber, 2011), or PlantCV, and imaging platform customized using basic Python (Noah Fahlgren Maximilian Feldman Malia, 2015). The type of data needed from the images will dictate if imaging platform tools are capable of carrying out the data analysis, or if more computer science-based automation such as machine learning could be trained to extract information from the images (Gehan & Kellogg, 2017). The most important aspect is ensuring the analysis is reproducible, precise, and accurate to ensure reliable data and results. With a higher demand for new varieties, it is critical to implement high-throughput phenotyping capabilities to improve the breeding process.

### ***Machine Learning***

Taking time to generate images is considered low-throughput (Gehan & Kellogg, 2017), and analyzing images manually is tedious, more prone to error, and time consuming (Stuart et al., 2019). Machine learning (ML) algorithms can increase efficiency by automating the analysis

portion of imaging (Gehan & Kellogg, 2017; Tsaftaris, Minervini, & Scharr, 2016), especially with large datasets (Rousseau, Dee, & Pridmore, 2015). ML is a computerized model that can learn patterns from data and can make decisions (Singh, Ganapathysubramanian, Singh, & Sarkar, 2016). There are several methods for generating a ML model depending on the task (Pridmore, French, & Pound, 2012). However, the main idea behind each method is the ability to find generalized trends or similarities and dissimilarities between images provided (Singh et al., 2016).

The first step of generating a ML model is preparing the training dataset. The data must be presented to the ML algorithm in a way it may be understood and used, known as feature extraction (Tsaftaris et al., 2016). Preparing the images could include minor changes such as cropping images, increasing contrast, or removing the background (Singh et al., 2016). This removes noise from the images, allowing more reliable data to be obtained (Tsaftaris et al., 2016). To finish preparing the dataset, a learning process must be chosen. ML algorithms can be trained using supervised or unsupervised training methods. For supervised learning, images are labeled in a way that helps the computer identify information in the image (LeCun, Bengio, & Hinton, 2015). The model will map output labels to the input image it receives (Singh et al., 2016). In unsupervised learning, the training data remains unlabeled before going through the learning process. In this method the model will identify structures in the image. Allowing the model to select what information it finds meaningful to use without input from the user. Finally, a semi-supervised model can be used. This means that parts of the training dataset images are labeled and part of it remains unlabeled (Singh et al., 2016)

A modeling objective must also be chosen for the ML method (Singh et al., 2016). A discriminative model is trained to discriminate between two different data patterns, many of

which are taught using supervised learning (Singh et al., 2016). Since these models are built for a predetermined task, they do not learn the features of objects in the image. A generative model is able to take the data pattern from the input dataset and generate synthetic images, allowing the model to perform several decision tasks simultaneously (Singh et al., 2016). Discriminative models usually perform better than generative models when larger amounts of training data are available, especially for classification tasks (Singh et al., 2016).

After a ML model is trained, a new dataset must be used to validate the algorithm. This dataset is known as the validation dataset and can be part of the population used in the training dataset or may come from a different population entirely (Singh et al., 2016). This tests the consistency and reliability of the model. Once accuracy of the model is deemed to be high enough, it may be regularly used for analysis (Tsaftaris et al., 2016).

### ***Deep Learning***

Deep learning(DL) is an emerging sub-section of ML where the features are not designed by human engineers, but instead learned from data using a general purpose learning procedure (LeCun et al., 2015). DL is able to perform image-based tasks such as segmentation, detection, localization, and classification (LeCun et al., 2015). While it is not commonly found in plant phenotyping literature, it excels at tackling large datasets with more complex data analytics questions (Ubbens & Stavness, 2017). DL is an example of the representation method, a method in which raw data can be fed to the model and have the representations required for detection or classification discovered automatically (LeCun et al., 2015). DL has multiple different layers of non-linear modules that start with raw data and transform the representation at each level until the output layer is reached. Convolutional neural networks (CNNs) are a class of deep learning methods composed of several different layers of connected processors called neurons

(Schmidhuber, 2015). The convolutional layer and pooling layers apply filters to the input data and represents the presence or absence of edges at particular orientations of the image and detect motifs in these arrangements, despite slight differences in edge positions (LeCun et al., 2015; Ubbens & Stavness, 2017). The third layer type is the fully connected layer. These final connected layers receive the output data from the convolutional and pooling layers, reshaped into a feature vector, as their input. The output layer is activated according to the type of analysis being performed, such as classification or regression (Ubbens & Stavness, 2017). DL does not require the preprocessing of images like traditional ML and has been able to outperform traditional ML techniques for tasks such as detection and segmentation (Girshick, 2015). Due to its ability to handle large datasets and complex analysis, DL is a proposed candidate for plant phenotyping (Tsaftaris et al., 2016). DL has been shown to greatly improve tasks such as leaf counting from images taken in controlled environments (Ubbens & Stavness, 2017). While hand created pipelines can be used for leaf counting, CNNs are able to account for varying factors of objects it is counting such as varying shapes or overlapping, making it optimal for analysis of different individual plants (Ubbens & Stavness, 2017).

## REFERENCES

- Abe, M. (2005). FD, a bZIP Protein Mediating Signals from the Floral Pathway Integrator FT at the Shoot Apex. *Science (American Association for the Advancement of Science)*, 309(5737), 1052-1056.
- Addison, C. K., Mason, R. E., Brown-Guedira, G., Guedira, M., Hao, Y., Miller, R. G., et al. (2016). QTL and major genes influencing grain yield potential in soft red winter wheat adapted to the southern United States. *Euphytica*, 209(3), 665-677.
- Allan, R. E. (1987). Wheat. In W. R. Fehr (Ed.), *Principles of Cultivar Development*: Collier MacMillan.
- Alqudah, A. M., Haile, J. K., Alomari, D. Z., Pozniak, C. J., Kobiljski, B., & Börner, A. (2020). Genome-wide and SNP network analyses reveal genetic control of spikelet sterility and yield-related traits in wheat. *Scientific reports*, 10(1), 1-12.
- Bassi, F. M., Bentley, A. R., Charmet, G., Ortiz, R., & Crossa, J. (2016). Breeding schemes for the implementation of genomic selection in wheat (*Triticum* spp.). *Plant science (Limerick)*, 242, 23-36.
- Beales, J., Turner, A., Griffiths, S., Snape, J. W., & Laurie, D. A. (2007). A Pseudo-Response Regulator is misexpressed in the photoperiod insensitive Ppd-D1a mutant of wheat (*Triticum aestivum* L.). *Theoretical and applied genetics*, 115(5), 721-733.
- Bentley, A. R., Turner, A. S., Gosman, N., Leigh, F. J., Maccaferri, M., Dreisigacker, S., et al. (2011). Frequency of photoperiod-insensitive Ppd-A1a alleles in tetraploid, hexaploid and synthetic hexaploid wheat germplasm. *Plant breeding*, 130(1), 10-15.
- Bernardo, R. (2020). *Breeding for quantitative traits in plants*. Woodbury, Minn: Stemma Press.

- Bhat, J. A., Ali, S., Salgotra, R. K., Mir, Z. A., Dutta, S., Jadon, V., et al. (2016). Genomic Selection in the. *Front Genet*, 7, 221.
- Blake, N. K., Lanning, S. P., Martin, J. M., Doyle, M., Sherman, J. D., Naruoka, Y., et al. (2009). Effect of Variation for Major Growth Habit Genes on Maturity and Yield in Five Spring Wheat Populations. *Crop science*, 49(4), 1211-1220.
- Boss, P., Bastow, R., Mylne, J., & Dean, C. (2004). Multiple Pathways in the Decision to Flower: Enabling, Promoting, and Resetting. *The Plant cell*, 16(suppl\_1), S18-S31.
- Cao, P., Liang, X., Zhao, H., Feng, B., Xu, E., Wang, L., et al. (2019). Identification of the quantitative trait loci controlling spike-related traits in hexaploid wheat (*Triticum aestivum* L.). *Planta*, 250(6), 1967-1981.
- Caroline, A. S., Wayne, S. R., & Kevin, W. E. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nature methods*, 9(7), 671.
- Carré, I. A. (2001). Day-length perception and the photoperiodic regulation of flowering in *Arabidopsis*. *J Biol Rhythms*, 16(4), 415-423.
- Chapman, S. C., Mathews, K. L., Trethowan, R. M., & Singh, R. P. (2006). Relationships between height and yield in near-isogenic spring wheats that contrast for major reduced height genes. *Euphytica*, 157(3), 391-397.
- Chen, Cheng, Chai, Wang, Du, Bian, et al. (2020). Pleiotropic QTL influencing spikelet number and heading date in common wheat (*Triticum aestivum* L.). *Theor Appl Genet*, 133(6), 1825-1838.
- Chen, Cheng, X., Chai, L., Wang, Z., Du, D., Bian, R., et al. (2020). Pleiotropic QTL influencing spikelet number and heading date in common wheat (*Triticum aestivum* L.). *Theor Appl Genet*, 133(6), 1825-1838.

- Chen, F., Gao, M., Zhang, J., Zuo, A., Shang, X., & Cui, D. (2013). Molecular characterization of vernalization and response genes in bread wheat from the Yellow and Huai Valley of China. *BMC plant biology*, *13*(1), 199-199.
- Chen, Z., Cheng, X., Chai, L., Wang, Z., Du, D., Bian, R., et al. (2020). Pleiotropic QTL influencing spikelet number and heading date in common wheat (*Triticum aestivum* L.). *Theor Appl Genet*, *133*(6), 1825-1838.
- Chen, Z., Cheng, X., Chai, L., Wang, Z., Du, D., Wang, Z., et al. (2020). Pleiotropic QTL influencing spikelet number and heading date in common wheat (*Triticum aestivum* L.). *Theoretical and applied genetics*, *133*(6), 1825-1838.
- Collard, B. C. Y., Jahufer, M. Z. Z., Brouwer, J. B., & Pang, E. C. K. (2005). An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. *Euphytica*, *142*(1-2), 169-196.
- Dobrovolskaya, O., Pont, C., Sibout, R., Martinek, P., Badaeva, E., Murat, F., et al. (2015). FRIZZY PANICLE drives supernumerary spikelets in bread wheat. *Plant physiology (Bethesda)*, *167*(1), 189-199.
- Dubcovsky, J., & Dvorak, J. (2007). Genome Plasticity a Key Factor in the Success of Polyploid Wheat Under Domestication. *Science (American Association for the Advancement of Science)*, *316*(5833), 1862-1866.
- Dubcovsky, J., Loukoianov, A., Fu, D., Valarik, M., Sanchez, A., & Yan, L. (2006). Effect of Photoperiod on the Regulation of Wheat Vernalization Genes VRN1 and VRN2. *Plant molecular biology*, *60*(4), 469-480.

- Díaz, A., Zikhali, M., Turner, A. S., Isaac, P., & Laurie, D. A. (2012). Copy Number Variation Affecting the Photoperiod-B1 and Vernalization-A1 Genes Is Associated with Altered Flowering Time in Wheat (*Triticum aestivum*). *PLoS one*, 7(3), e33234-e33234.
- Ellis, M., Spielmeier, W., Gale, K., Rebetzke, G., & Richards, R. (2002). "Perfect" markers for the Rht-B1b and Rht-D1b dwarfing genes in wheat. *Theoretical and applied genetics*, 105(6), 1038-1042.
- Erkul, A., Unay, A., & Konak, C. (2010). Inheritance of yield and yield components in a bread wheat (*Triticum aestivum* L.) cross (Vol. 15, pp. 137-140): Turkish Journal of Field Crops.
- Evans, R. (2016). Photoperiodism (pp. 209-211).
- FAO. (2022). FAO Cereal Supply and Demand Brief: FAO.
- Foley, J. A., Ramankutty, N., Brauman, K. A., Cassidy, E. S., Gerber, J. S., Johnston, M., et al. (2011). Solutions for a cultivated planet. *Nature*, 478(7369), 337-342.
- Furbank, R. T., & Tester, M. (2011). Phenomics – technologies to relieve the phenotyping bottleneck. *Trends in plant science*, 16(12), 635-644.
- Gale, M. D., & Youssefian, S. (1985). Dwarfing genes in wheat. *Proj. plant Breed*, 1, 1-35.
- Ganal, M. W., Plieske, J., Hohmeyer, A., Polley, A., & Order, M. S. (2019). High-throughput genotyping for cereal research and breeding *Applications of Genetic and Genomic Research in Cereals* (pp. 3-17): Woodhead Publishing Series in Food Science, Technology and Nutrition.
- Gehan, M. A., & Kellogg, E. A. (2017). High-throughput phenotyping. *American journal of botany*, 104(4), 505-508.
- Girshick, R. (2015). Fast R-CNN (pp. 1440-1448): IEEE.

- Goddard. (2009). Genomic selection: prediction of accuracy and maximisation of long term response. *Genetica*, 136(2), 245-257.
- Goddard, M. E., & Hayes, B. J. (2007). Genomic selection. *Journal of animal breeding and genetics (1986)*, 124(6), 323-330.
- Godfray, H. C., Beddington, J. R., Crute, I. R., Haddad, L., Lawrence, D., Muir, J. F., et al. (2010). Food security: the challenge of feeding 9 billion people. *Science*, 327(5967), 812-818.
- 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.
- Hartmann, A., Czauderna, T., Hoffmann, R., Stein, N., & Schreiber, F. (2011). HTPheno: An image analysis pipeline for high-throughput plant phenotyping. *BMC bioinformatics*, 12(1), 148-148.
- Herndl, M., White, J. W., Graeff, S., & Claupein, W. (2008). The impact of vernalization requirement, photoperiod sensitivity and earliness per se on grain protein content of bread wheat (*Triticum aestivum* L.). *Euphytica*, 163(2), 309-320.
- Higgins, J. A., Bailey, P. C., & Laurie, D. A. (2010). Comparative Genomics of Flowering Time Pathways Using *Brachypodium distachyon* as a Model for the Temperate Grasses. *PloS one*, 5(4), e10065-e10065.
- Houle, D., Govindaraju, D. R., & Omholt, S. (2010). Phenomics: the next challenge. *Nature reviews. Genetics*, 11(12), 855-866.
- Kamran, A., Iqbal, M., & Spaner, D. (2014). Flowering time in wheat (*Triticum aestivum* L.): a key factor for global adaptability. *Euphytica*, 197(1), 1-26.

- Kato, H., Taketa, S., Ban, T., Iriki, N., & Murai, K. (2001). The influence of a spring habit gene, *Vrn-D1*, on heading time in wheat. *Plant breeding*, *120*(2), 115-120.
- Kato, K., Miura, H., & Sawada, S. (1999). Detection of an earliness per se quantitative trait locus in the proximal region of wheat chromosome 5AL. *Plant breeding*, *118*(5), 391-394.
- Kato, K., Miura, H., & Sawada, S. (2000). Mapping QTLs controlling grain yield and its components on chromosome 5A of wheat. *Theoretical and applied genetics*, *101*(7), 1114-1121.
- Khan, A. S., Salim, I., & Ali, Z. (2003). Heritability of various morphological traits in wheat. *Int. J. Agric. Biol*, *5*(2), 138-140.
- Koboldt, Daniel C., Steinberg, Karyn M., Larson, David E., Wilson, Richard K., & Mardis, E. R. (2013). The Next-Generation Sequencing Revolution and Its Impact on Genomics. *Cell (Cambridge)*, *155*(1), 27-38.
- Koppolu, R., & Schnurbusch, T. (2019). Developmental pathways for shaping spike inflorescence architecture in barley and wheat. *J Integr Plant Biol*, *61*(3), 278-295.
- Larkin, Lozada, & Mason. (2019). Genomic Selection—Considerations for Successful Implementation in Wheat Breeding Programs. *Agronomy (Basel)*, *9*(9), 479.
- Law, C. N., Sutka, J., & Worland, A. J. (1978). A Genetic study of day-length response in wheat. *Heredity*, *41*(2), 185-191.
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature (London)*, *521*(7553), 436-444.
- Lewis, S., Faricelli, M. E., Appendino, M. L., Valárik, M., & Dubcovsky, J. (2008). The chromosome region including the earliness per se locus *Eps-Am1* affects the duration of

- early developmental phases and spikelet number in diploid wheat. *Journal of experimental botany*, 59(13), 3595-3607.
- Liu, J., Xu, Z., Fan, X., Zhou, Q., Cao, J., Wang, F., et al. (2018). A Genome-Wide Association Study of Wheat Spike Related Traits in China. *Frontiers in plant science*, 9, 1584-1584.
- Lozada, D. N., Mason, R. E., Sarinelli, J. M., & Brown-Guedira, G. (2019). Accuracy of genomic selection for grain yield and agronomic traits in soft red winter wheat. *BMC genetics*, 20(1), 82-82.
- Lozada [Unpublished], D., Carter, A., & Mason, R. Unlocking yield potential of wheat: influence of major growth habit and adaptation genes.
- Luo, M. C., Yang, Z. L., & Dvořák, J. (2000). The Q locus of Iranian and European spelt wheat. *Theoretical and applied genetics*, 100(3), 602-606.
- Meuwissen, T. H. E., Hayes, B. J., & Goddard, M. E. (2001). Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. *Genetics (Austin)*, 157(4), 1819-1829.
- Minervini, M., Giuffrida, M. V., Perata, P., & Tsiftaris, S. A. (2017). Phenotiki: an open software and hardware platform for affordable and easy image-based phenotyping of rosette-shaped plants. *The Plant journal : for cell and molecular biology*, 90(1), 204-216.
- Minervini, M., Scharr, H., & Tsiftaris, S. A. (2015). Image Analysis: The New Bottleneck in Plant Phenotyping [Applications Corner]. *IEEE signal processing magazine*, 32(4), 126-131.
- Miura, H., & Worland, A. (1994). Genetic control of vernalization, day-length response, and earliness per se by homoeologous group-3 chromosomes in wheat. *Plant Breeding*, 113(2), 160-169.

- Moon, J., Lee, H., Kim, M., & Lee, I. (2005). Analysis of flowering pathway integrators in *Arabidopsis*. *Plant Cell Physiol*, *46*(2), 292-299.
- Newell, M. A., & Jannink, J.-L. (2014). Genomic Selection in Plant Breeding. *Crop Breeding*, *1145*, 117-130.
- Noah Fahlgren Maximilian Feldman Malia, A. G. M. S. W. C. S. D. W. B. S. T. H. C. J. M. a. N. W. I. K. T. F. (2015). A Versatile Phenotyping System and Analytics Platform Reveals Diverse Temporal Responses to Water Availability in *Setaria*. *Molecular plant*, *8*(10), 1520-1535.
- Pearce, S., Saville, R., Vaughan, S. P., Chandler, P. M., Wilhelm, E. P., Sparks, C. A., et al. (2011). Molecular Characterization of Rht-1 Dwarfing Genes in Hexaploid Wheat. *Plant physiology (Bethesda)*, *157*(4), 1820-1831.
- 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.
- Rawson, H. M. (1970). Spikelet number, its control and relation to yield per ear in wheat (Vol. 23, pp. 1-15): Australian Journal of Biological Sciences.
- Ray, D. K., Mueller, N. D., West, P. C., & Foley, J. A. (2013). Yield trends are insufficient to double global crop production by 2050. *PloS one*, *8*(6), e66428.
- Rousseau, D., Dee, H., & Pridmore, T. (2015). Imaging Methods for Phenotyping of Plant Traits (pp. 61-74). New Delhi: Springer India.
- Sakuma, S., Salomon, B., & Komatsuda, T. (2011). The domestication syndrome genes responsible for the major changes in plant form in the Triticeae crops. *Plant Cell Physiol*, *52*(5), 738-749.

- Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural networks*, *61*, 85-117.
- Searle, I., He, Y., Turck, F., Vincent, C., Fornara, F., Kröber, S., et al. (2006). The transcription factor FLC confers a flowering response to vernalization by repressing meristem competence and systemic signaling in Arabidopsis. *Genes & development*, *20*(7), 898-912.
- Shaw, L. M., Turner, A. S., & Laurie, D. A. (2012). The impact of photoperiod insensitive Ppd-1a mutations on the photoperiod pathway across the three genomes of hexaploid wheat (*Triticum aestivum*). *The Plant journal : for cell and molecular biology*, *71*(1), 71-84.
- Shewry, P. R. (2009). Wheat. *Journal of Experimental Botany*, *60*(6), 1537-1553.
- Shewry, P. R., & Hey, S. J. (2015). The contribution of wheat to human diet and health. *Food and energy security*, *4*(3), 178-202.
- Shitsukawa, N., Ikari, C., Mitsuya, T., Sakiyama, T., Ishikawa, A., Takumi, S., et al. (2007). Wheat SOC1 functions independently of WAP1/VRN1, an integrator of vernalization and photoperiod flowering promotion pathways. *Physiologia plantarum*, *130*(4), 627-636.
- 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.
- Snape, J. W., Butterworth, K., Whitechurch, E., & Worland, A. J. (2001). Waiting for fine times: genetics of flowering time in wheat. *Euphytica*, *119*(1), 185-190.
- Spindel, J., Begum, H., Akdemir, D., Virk, P., Collard, B., Redoña, E., et al. (2015). Genomic Selection and Association Mapping in Rice (*Oryza sativa*): Effect of Trait Genetic Architecture, Training Population Composition, Marker Number and Statistical Model on

- Accuracy of Rice Genomic Selection in Elite, Tropical Rice Breeding Lines. *PLoS genetics*, *11*(2), e1004982-e1004982.
- Stelmakh, A. F. (1998). Genetic systems regulating flowering response in wheat. *Euphytica*, *100*(1), 359-369.
- Stuart, B., Dominik, K., Thorben, K., Christoph, N. S., Bernhard, X. K., Carsten, H., et al. (2019). ilastik: interactive machine learning for (bio)image analysis. *Nature methods*, *16*(12), 1226-1232.
- Thomas, B. (2006). Light signals and flowering. *Journal of experimental botany*, *57*(13), 3387-3393.
- Tovar, J. C., Hoyer, J. S., Lin, A., Tielking, A., Callen, S. T., Elizabeth Castillo, S., et al. (2018). Raspberry Pi-powered imaging for plant phenotyping. *Applications in plant sciences*, *6*(3), e1031-n/a.
- Tsaftaris, S. A., Minervini, M., & Schar, H. (2016). Machine Learning for Plant Phenotyping Needs Image Processing. *Trends in plant science*, *21*(12), 989-991.
- Ubbens, J. R., & Stavness, I. (2017). Deep Plant Phenomics: A Deep Learning Platform for Complex Plant Phenotyping Tasks. *Frontiers in plant science*, *8*, 1190-1190.
- UNIES, U. N. (2019). World Population Prospects 2019: Highlights. *UN Dep. Econ. Soc. Aff. Popul. Div.*
- Visscher, P. M., Wray, N. R., Zhang, Q., Sklar, P., McCarthy, M. I., Brown, M. A., et al. (2017). 10 Years of GWAS Discovery: Biology, Function, and Translation. *American journal of human genetics*, *101*(1), 5-22.

- Ward, B. P., Brown-Guedira, G., Tyagi, P., Kolb, F. L., Sanford, D. A., Sneller, C. H., et al. (2019). Multienvironment and Multitrait Genomic Selection Models in Unbalanced Early-Generation Wheat Yield Trials. *Crop science*, 59(2), 491-507.
- Wilhelm, E. P., Turner, A. S., & Laurie, D. A. (2008). Photoperiod insensitive Ppd-A1a mutations in tetraploid wheat (*Triticum durum* Desf.). *Theoretical and applied genetics*, 118(2), 285-294.
- Wiseman, B., & Webster, J. (1999). *Economic, environmental, and social benefits of resistance in field crops*. Lanham, Md: Entomological Society of America.
- Worland, A. J. (1996). The influence of flowering time genes on environmental adaptability in European wheats. *Euphytica*, 89(1), 49-57.
- Wurchum, T., Leiser, W., Langer, S., Tucker, M., & Longin, C. (2018). Phenotypic and genetic analysis of spike and kernel characteristics in wheat reveals long-term genetic trends of grain yield components. (Vol. 131, pp. 2071-2084): *Theoretical and Applied Genetics*.
- Yan, L. (2004). The Wheat VRN2 Gene Is a Flowering Repressor Down-Regulated by Vernalization. *Science (American Association for the Advancement of Science)*, 303(5664), 1640-1644.
- Yan, L., Fu, D., Li, C., Blechl, A., Tranquilli, G., Bonafede, M., et al. (2006). The wheat and barley vernalization gene VRN3 is an orthologue of FT. *Proceedings of the National Academy of Sciences - PNAS*, 103(51), 19581-19586.
- Yan, L., Loukoianov, A., Tranquilli, G., Helguera, M., Fahima, T., & Dubcovsky, J. (2003). Positional cloning of the wheat vernalization gene VRN1. *Proceedings of the National Academy of Sciences - PNAS*, 100(10), 6263-6268.

Zhang, B., Liu, X., Xu, W., Chang, J., Li, A., Mao, X., et al. (2015). Novel function of a putative MOC1 ortholog associated with spikelet number per spike in common wheat. *Scientific reports*, 5(1), 12211-12211.

Zhang, X., Xiao, Y., Zhang, Y., Xia, X., Dubcovsky, J., & He, Z. (2008). Allelic variation at the vernalization genes Vrn-A1, Vrn-B1, Vrn-D1, and Vrn-B3 in Chinese wheat cultivars and their association with growth habit. *Crop Science*, 48(2), 458-470.

## **CHAPTER II: PHENOTYPING AND PREDICTING WHEAT SPIKE CHARACTERISTICS USING IMAGE ANALYSIS AND MACHINE LEARNING**

### **INTRODUCTION**

Phenotypes are the resulting characteristics of a genotype and genotype by environment interaction. Accurate phenotyping is vital to genetic improvement through breeding and for discovery of marker-trait associations, genome-wide association studies, marker-assisted selection, and genomic selection (Furbank and Tester 2011, Minervini, Scharr and Tsafaris 2015). While advancements in technology for generating high-density genotypic data have increased the efficiency of genomic analysis (Koboldt et al. 2013), phenotyping methods are still lagging (Houle, Govindaraju and Omholt 2010). There has recently been a push to create high-throughput phenotyping (HTP) techniques to increase the rate at which we can understand and predict complex traits such as grain yield. Multiple HTP approaches are now being evaluated and deployed in plant science research, including sensors, unoccupied ariel vehicles (UAVs), and imaging.

Recently UAVs have gained popularity for in-field phenotyping as remote sensing capabilities improve (Araus et al. 2018). UAVs allow for images or videos of large areas of land to be taken in short periods of time, allowing for several field trials to be phenotyped in a single day. UAVs are highly customizable given the ability to choose sensors and cameras attached. Using this remote sensing technique allows for the observation of physiological and spectral traits while also limiting the amount of labor, time, and cost associated (Krause et al. 2020). Sensors on drones can generate heat map and reflectance data which can be used for vegetative indices or chlorophyll content to predict plant health (Huete et al. 2002, Condorelli et al. 2018).

While UAVs allow efficient in-field data collection, they have a high entry cost and do not allow for highly detailed images of individual plants.

Images taken in controlled settings take more time and labor to collect, but require lower-cost equipment and can provide varying images ranging from a microscopic to macroscopic level. Imaging techniques are highly reproducible and cover a variety of applications ranging from microscopic to macroscopic subjects (Furbank & Tester, 2011; Minervini et al. 2015), capturing a single maturity point in time, or creating time-lapse data throughout a field season (Fahlgren and Maximilian Feldman Malia 2015, Gehan and Kellogg 2017). While the collection of image data is important, an efficient analysis pipeline is also critical to maximize output and not just input of data.

Machine Learning (ML) is a computerized model that can learn patterns from data and make decisions (Singh et al. 2016) While there are multiple types of models used for machine learning, the main idea behind computer vision is the detection of similarities and dissimilarities between the images provided. ML has gained popularity for use in HTP in plant breeding programs due to the ability to automate the analysis process and increase efficiency (Gehan and Kellogg 2017, Tsaftaris, Minervini and Scharr 2016). Models can be trained using supervised methods, with labels for images during the time of training, or unsupervised methods, which allows the model to decide the most meaningful features of an image (Singh et al. 2016).

Training an ML model requires your dataset to be divided into three different parts including: 1) a training set where the model will learn how to identify relevant aspects based on its task, 2) a validation set for testing the accuracy of the model, and 3) a test set. The validation set allows for fine-tuning of the model's hyperparameters to increase accuracy to the desired level before the test set is used. The test set is part of the dataset that the model has never seen

and was not used for training (Singh et al. 2016). For supervised learning, training sets labeled with the correct information are an input to the model so that it can learn the associations between the provided image and the associated label. Unsupervised learning does not provide any labels with the validation set and allows the algorithm to make its own conclusions.

A subset of ML, called deep learning, uses feature vectors not designed by human engineers (LeCun, Bengio and Hinton 2015). Deep learning models have different layers of nonlinear modules that can transform data into different representations at each layer until an output is reached. This structure allows deep learning models to excel with large datasets and more complex data analytics and is commonly found in plant phenotyping literature (Ubbens and Stavness 2017). Convolutional neural networks (CNNs) are a class of deep learning methods where the processing layers are convolutional layers and pooling layers (LeCun et al. 2015). The final layer of a CNN is the fully connected layer, which takes the output from the processing layers and reshapes them back into a single column for processing based on the type of model being used, such as classification or regression (Ubbens and Stavness 2017). CNNs have the ability to find edges and motifs in images that improve their ability for image analysis for certain plant phenotyping tasks (Tsaftaris et al. 2016) such as counting (Ubbens and Stavness 2017). Deep learning can be used for several image-based tasks including segmentation, regression, classification, and detection (LeCun et al. 2015). Regression can be used for counting tasks over other methods, such as classification, to have a better understanding of how close the predicted values are to the true values of an image, providing an error rate that can help assess model accuracy.

Increasing the amount of phenotypic data available for wheat spike characteristics within populations would allow for further genetic analysis of the flowering pathway, which is not fully

understood in wheat. The use of imaging and image analysis for plant phenotyping is also transferable to other crop species, making it a useful area of research and improvement. The objective of this study was to evaluate soft red winter wheat (SRWW) genotypes using imaging techniques to phenotype spikelets spike<sup>-1</sup> and develop a deep learning model for high-throughput analysis of wheat spike images.

## **MATERIALS AND METHODS**

### ***Plant Materials***

The genetic material used in this study was the Historic Gulf Atlantic Nursery (HGAWN), a population consisting of 594 soft red winter wheat lines from public breeding institutions located in the southeastern United States. The HGAWN included varieties from the University of Arkansas (n=103), Louisiana State University (n=109), University of Georgia (n=105), North Carolina State University (n=104), Texas A&M University (n=60), Virginia Polytechnic Institute (n=44), Clemson University (n=19), and the United States Department of Agriculture Agricultural Research Service (USDA-ARS, n=9).

### ***Experimental design***

For this study, the HGAWN was evaluated during the 2019 and 2020 growing seasons at the Milo J. Shult Agricultural Research & Extension Center in Fayetteville, Arkansas. Plots were drill seeded at a rate of 250 seed m<sup>-2</sup> in a randomized complete block design with two replications. Each plot consisted of a single 1.20 m long row with 0.38 m horizontally between rows and 0.60 vertically between each range of plots. During both seasons, pre-plant recommendations were followed for phosphorus and potassium and 100 kg hectare<sup>-1</sup> of nitrogen in the form of urea was applied in a split application in the spring (February and March).

Applications of Harmony<sup>®</sup> Extra (0.28 kg Ha<sup>-1</sup>) and Axial<sup>®</sup> (0.6 kg Ha<sup>-1</sup>) were applied for control of annual ryegrass (*Lolium multiflorum* L.) and other weed species.

### ***Imaging***

An imaging device developed by 1, was used to achieve consistent lighting and camera height across wheat spike image capture (Figure 2.1A). The imaging device was designed for a Canon<sup>®</sup> Powershot G1 X Mark II camera. A 4.5 cm circular hole was drilled into the bottom of a 0.46-liter bucket using a 4.5 cm drill bit on an electric drill. The hole was drilled into the center of the bottom of the bucket with the mouth of the bucket facing downward. After drilling, the edges were filed down to protect the camera lens from scratches. A 2 cm hole was then drilled into the side of the bucket approximately 10 cm above the base of the mouth while the mouth of the bucket faced downwards.

The inside of the bucket was painted by spraying with RUST-OLEUM<sup>®</sup> Camouflage Ultra Flat Black so that the interior was fully coated. The paint was cured for five minutes and was coated until exterior light was no longer visible through the walls of the bucket. After the final coat of paint had cured, tape lights from a 1.8 m GoodEarth<sup>®</sup> Self-Adhesive Tape Lighting Kit were cut to fit the circumference of the bucket. The protective paper strip was removed from the back of the lights, allowing it to be directly applied to the inside of the bucket. The lights were adhered to the interior of the bucket starting at the 2 cm hole, making sure they remained parallel to the mouth of the bucket. Any necessary paint touch ups were performed using an artistic brush.

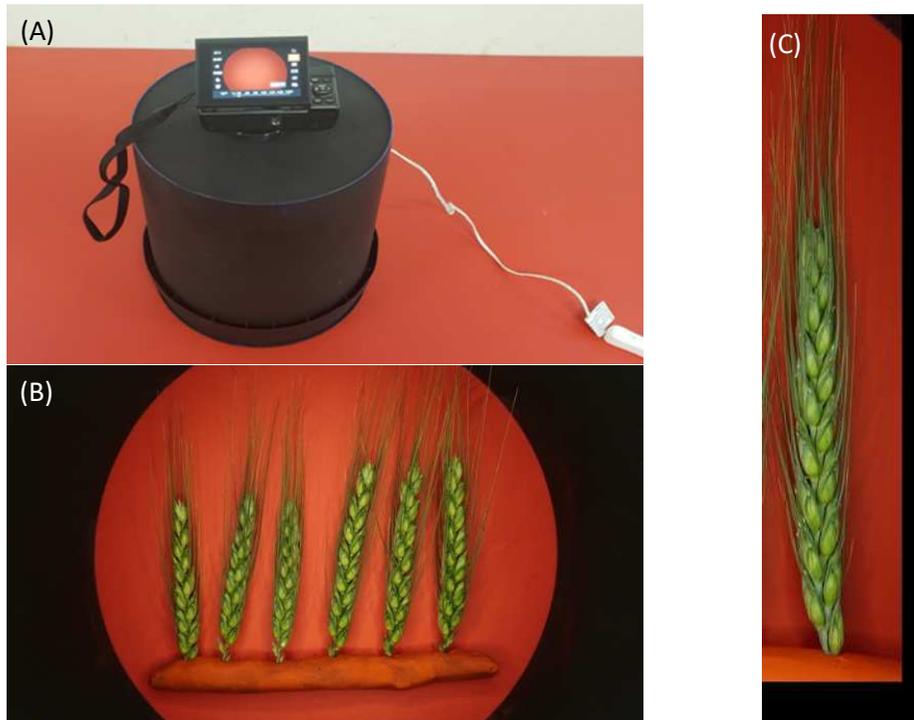
For an imaging surface, a 1.29 m by 0.81 m maker board with a smooth particle board back was used. The board was painted red using Classic Red Valspar Ultra<sup>®</sup> Interior Flat Paint

until the color was opaque and texture was gone. The red background was chosen for visibility of the heads during imaging. Each image was taken on the marker board surface using the imaging cap and the Canon® Powershot G1 X Mark II camera. Each of the images were 20.1 megapixels and had a focal length of 9 millimeters, and F-stop of f/4.0, and a shutter speed of 1/60 seconds at an ISO of 200 with a white balance of 3000 Kelvin. This allowed for clarity in the images and definition of the spike for analysis.

### ***Image Analysis – Image J***

For determination spikelets spike<sup>-1</sup> (SPS), six heads were collected from each plot at 30 to 33 days after the plot was fully headed but before the onset of senescence. All images were taken the same day as sampling. The peduncle of all six heads was inserted into a piece of red Van Aken® Plastalina Modeling Clay, ensuring visibility of the heads in the image. Each head was placed in the clay in a uniform manner so that the rachis faced upwards towards the lens to ensure each individual spikelet could be observed (Figure 2.1B). SPS was manually measured using ImageJ version 1.52o (Caroline, Wayne and Kevin 2012).

For SPS counting, each visible spikelet was counted manually starting from the base of the spike and counting the alternating spikes as they go up the rachis using the multi-point tool in ImageJ. The total number of spikelets in each image was then downloaded to a CSV file. The number of times an image name occurred in the CSV file corresponded with the number of selections that had been made on the image. This number was then divided by the number of heads in the image to obtain the average SPS for that genotype.



**Figure 2.1.** A Phenotyping device allowed for a controlled environment and consistency for photos of wheat spikes. (A) A camera rests on top of the device, which was used on the same red background for each photo. The black interior of the cover and attached light strip create uniform lighting and minimal reflection from camera flash while taking pictures. (B) Controlled environment images of six wheat spikes from each plot. Each spike was inserted into clay by the peduncle, rachis side up. Spikelets were fully visible for each spike to allow for phenotyping. (C) Example input image for specified deep learning algorithms. Each image of six spikes was cropped down to images of individual spikes before having their size reduced and padding the images to ensure identical sizing.

### ***Image Analysis – Deep Learning***

For optimal use in a deep learning algorithm, the images needed to be prepared and simplified. First, each image of six wheat heads was repeatedly cropped to create six separate photos of individual spikes. The cropped photos were padded to the same size of 140 by 600 pixels (Figure 2.1C). The number of spikelets in each image was counted using the multitool on ImageJ, recorded in a CSV file, and used as the associated label for the image. After preparation,

12,717 total images were used for model development. The mean spikelets per image was 17.5 with a range of 11 to 32 spikelets (Table 2.1). Summary statistics for spikelets for prepared images are shown in table 1.

**Table 2.1.** Mean, standard deviation, and range for all 12,717 images prepared for use in the deep learning algorithm.

	Mean	Standard Deviation	Range
Spikelets per Image	17.5	2.02	11 - 32

Four different CNNs were trained for this study, one classification CNN and three regression CNNs were trained in this study: The classification model and one of the regression models had five alternating sets of convolutional and max-pooling layers The last two regression models were a pre-trained Visual Geometry Group (VGG)16 application, and a pre-trained VGG19 application (Simonyan and Zisserman 2014). VGG16 and VGG19 models have been utilized in several image counting studies (Khaki et al., 2020) (Ubbens and Stavness 2017) with the VGG19 model having a larger initial layer size, three additional convolutional layers, and two additional filter sizes. The loss function for each model evaluates differences between values predicted by the model and their true value. Each of the models was trained using the python tensorflow package (Abadi et al. 2016) through the Google Colab application (Bisong 2019). Both the VGG16 and VGG19 models were trained using the 1,000 class ImageNet dataset of over one million images (Russakovsky et al. 2015) and made available through Keras (Chollet 2015). Image arrays for the first model were normalized by dividing them by 255 to put vector values between 0 and 1, while images for the VGG16 and VGG19 models were preprocessed using their respective TensorFlow Keras preprocessing functions. Images were randomly divided into three different groups, 70% were used in a training set while 20% were used for the test set

and the remaining 10% was reserved as a validation set. Sparse categorical crossentropy was used as the loss function for the classification model and is given by

$$J(w) = -\frac{1}{N} \sum_{i=1}^N [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)]$$

Where  $w$  is the model parameters,  $y_i$  is the true label, and  $\hat{y}_i$  is the predicted label. Sparse categorical cross entropy was used over cross entropy since there are a large amount of label values, all of which are integers. The MSE was used as the loss function for all three regression models and is given by

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2$$

Where  $N$  is the number of samples,  $y_i$  is the true value of the label, and  $\hat{y}_i$  is the predicted label.. Mean absolute error (MAE) was the second metric used as to evaluate model performance and each model was run for 10 epochs. MSE was used to evaluate model fit and MAE was used to evaluate the average deviation of the estimated values for spikelets from the true value in the image. The mean and standard deviation for each metric was calculated from 10 iterations of the model.

## RESULTS

While classification models show how well a model can accurately categorize an image, regression models demonstrate the error in estimation. In this study, three different regression CNNs were trained to predict the number of spikelets on a single spike in an image. The three models included a basic CNN with five sets of a convolutional layer followed by a max-pooling layer, a pretrained VGG16 model, and a pretrained VGG19 model. The models were then run for

20 iterations for robustness and variability, each model was trained using the same training and test sets. Mean average error (MAE) and mean squared error (MSE) were used for the evaluation of average error in spikelet estimation and fit of the model, respectively. Results for each of the models is presented in table 1.

### **Classification model**

The classification model had a mean accuracy of 0.27 with a standard deviation of 0.02 and a loss of 1.79 with a standard deviation of 0.05. On average, under 30% of images presented to the model receive an accurately predicted label, however, the model had consistent accuracy across iterations. Accuracy measures for classification models do not represent an error rate to understand the difference in predicted values and true values. The accuracy of the model can give misleading information about the usefulness of the model because it fails to communicate the difference between the true value and predicted values. Because of this, the regression models were evaluated for error and fit.

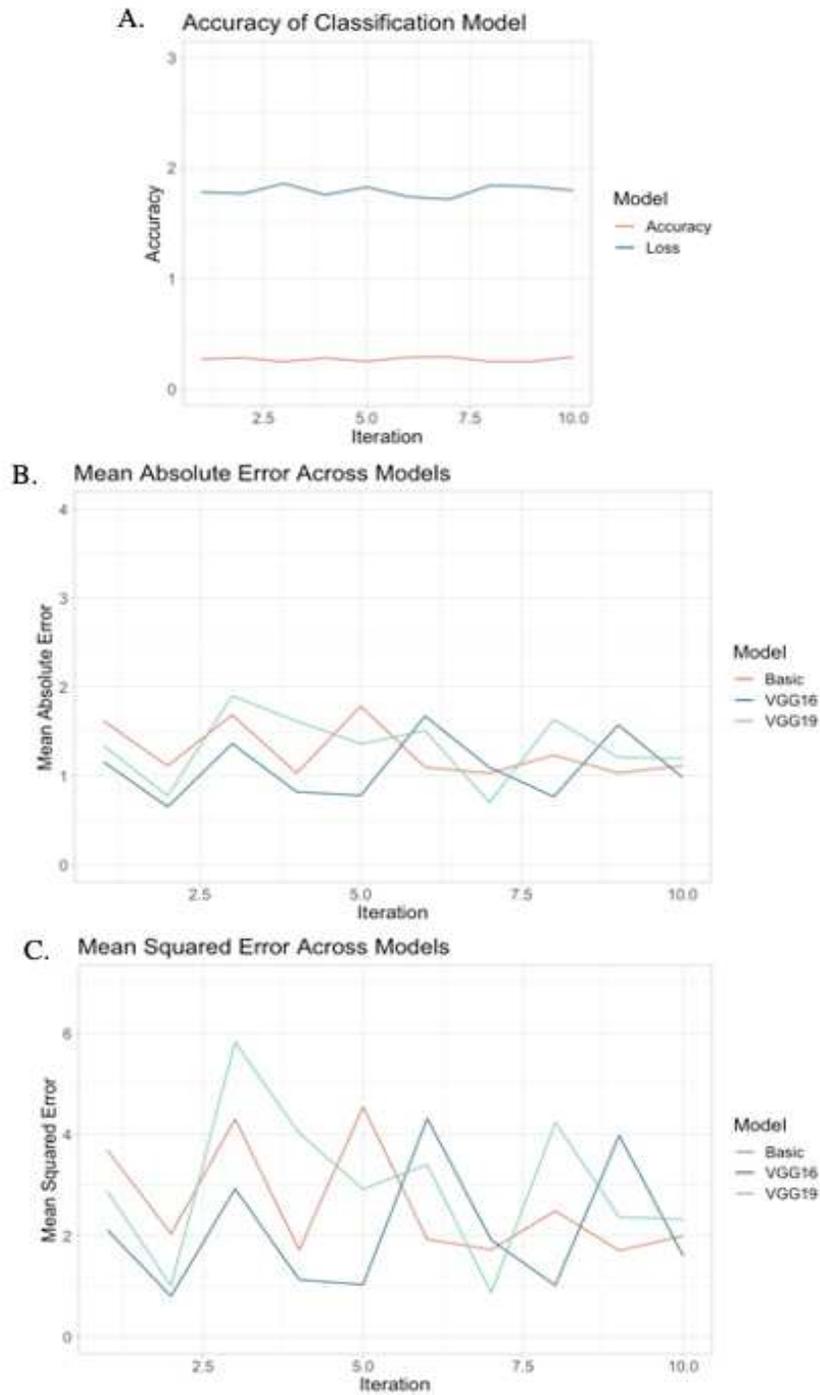
### **Regression model**

For the regression models, the basic CNN model had a mean MAE for the model was 1.27 with a standard deviation of 0.30 and an MSE of 2.61 with a standard deviation of 1.12. The VGG16 model had an average MAE of 1.09 and a standard deviation 0.35 and an MSE of 2.08 with a standard deviation of 1.26 (Table 2.2). The VGG19 model had an average MAE of the model was 1.32 with a standard deviation of 0.38 and an MSE of 2.98 with a standard deviation of 1.49. The MAEs indicate that the VGG16 model had the lowest error between the true value of spikelets in an image and the value the model estimated, being an average of 1.09 spikelets off from the true value. For five of the ten iterations the model was run, the MAE of the VGG16

model was less than one with the lowest being 0.65. The standard deviation of MAE was nearly equal for all three models with 0.30 for the basic model, 0.35 for VGG16, and 0.36 for VGG19. Because the standard deviations were so similar, yet VGG16 had a much lower MAE, VGG16 is the most robust model and has the best chance of having a consistently lower error over time than the other two models presented. The mean squared error (MSE) allows for comparison of the fit of the models, with the VGG16 having the lowest MAE and MSE of all models, showing it had the best fit ( $R^2 = 0.53$ ) comparatively as well as the lowest error. VGG19 had both the highest MAE and MSE, indicating it is the least successful of the three models. While VGG16 did not have the lowest standard deviations for all model MSEs, it still has the most consistently low MSE across iterations. Figure 2 depicts the MAE and MSE for all models across the ten iterations they were run to show an error comparison over time. VGG19 also showed exceptional variation across the 10 epochs of each iteration and varied at which epoch optimal training occurred. VGG19 was the least optimal of all three models ( $R^2 = 0.45$ ), the basic regression CNN had the second best fit ( $R^2 = 0.48$ ). Regressions comparing predicted values to true values of spikelets are shown in figure 3.

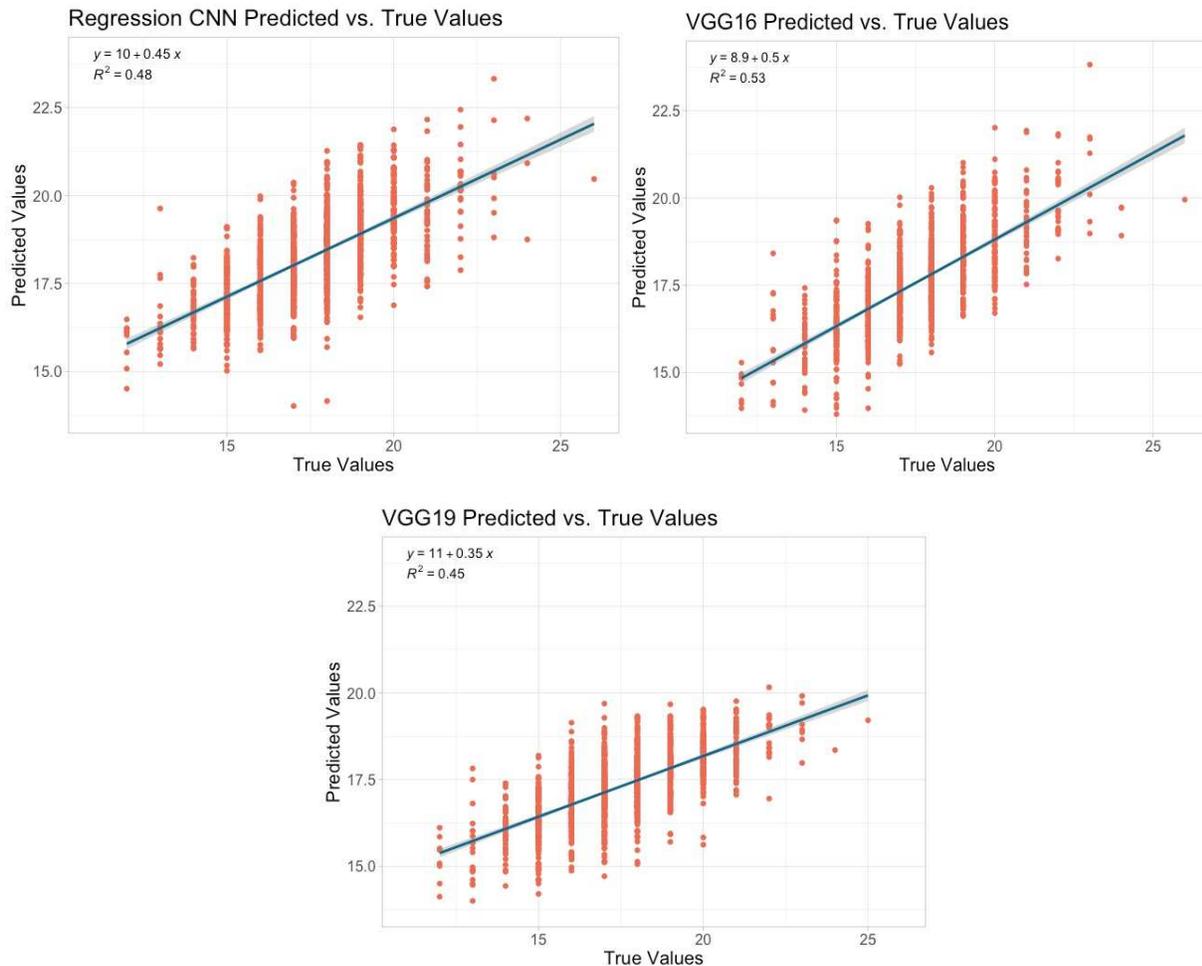
**Table 2.2.** The mean and standard deviation of accuracy and loss for the classification model and mean absolute error (MAE) and mean squared error (MSE) means and standard deviations for each of the three regression models.

Model	Accuracy	Loss
<u>Classification</u>		
Basic Model	$0.27 \pm 0.02$	$1.79 \pm 0.05$
Model	MAE	MSE
<u>Regression</u>		
Basic Model	$1.27 \pm 0.30$	$2.61 \pm 1.12$
VGG16	$1.09 \pm 0.35$	$2.08 \pm 1.26$
VGG19	$1.32 \pm 0.36$	$2.98 \pm 1.49$



**Figure 2.2.** Comparisons of the accuracy and loss across iterations for the classification model and the the (B) MAE and (C) MSE between the basic, VGG16, and VGG19 regression models over 10 iterations. The performance of each model varies over each iteration where no model was the best model for all ten

iterations. VGG16 had the lowest MAE and MSE compared to the basic model and VGG19 model for more iterations than any other regression model.



**Figure 2.3.** Regressions of predicted spikelet numbers versus true spikelet values in each image for the basic regression CNN, VGG16 regression, and VGG19 regression models. The x-axis is the true values of spikelets present in each image and the y-axis is the value of spikelets predicted by each respective model for each image. Displayed within each graph is the equation of the regression line and the coefficient of determination ( $R^2$ ).

## DISCUSSION

Phenotyping through imaging promises a cheap, accurate, and easily reproducible means of collecting data from lines in a plant breeding program. Developing means of phenotyping that allow for a larger amount of data collection is crucial to improving the efficiency and accuracy of

a breeding pipeline. The use of deep learning for high throughput phenotyping in plant breeding is promising, but further development and fine-tuning of hyperparameters is needed to generate optimal models. The objective of this study was to demonstrate the capabilities of imaging for the high-throughput phenotyping of wheat heads using deep learning models to analyze spikelets spike<sup>-1</sup>. Moving forward with current research, models should be evaluated using validation data and tested in analyses that can be compared to previously performed research. Regression models showed more potential than classification models for counting algorithms based on the inability of classification models to take error into account. It has been shown that the amount of data produced is more important than the accuracy of the data (Lane and Murray 2021), but the acceptable threshold for analyses in plant breeding programs must be discovered prior to implementation.

Counting has several uses in agriculture and is commonly seen in ML models designed in other studies such as counting leaves (Ubbens and Stavness 2017, Miao et al. 2021) and corn stalks in fields (Khaki et al. 2020). Counting is often used for features such as leaf number due to its correlations with various traits depending on the species. Generating a model for SPS not only provides a resource for phenotyping wheat spike characteristics but can also contribute to the development of models designed for similar tasks in agriculture. Both regression (Miao et al. 2021, Ubbens and Stavness 2017) and detection-based methods (Khaki et al. 2020) have been implemented using sets of annotated images. Miao et al., 2021 had an agreement rate ranging from 0.33 to 0.45, which is the proportion of perfect predictions, and MSEs ranging from 0.92 to 1.72. This is most comparable to the basic classification model used for counting spikelets, which had an average accuracy rate of 0.27 and an MSE of 1.79. Ubbens and Stavness, 2017 used regression to count *Arabidopsis* leaves with MAEs of 0.41 and 0.61 as well as tobacco

leaves with an MAE of 0.61. Similar structure to the basic CNNs performed in this study allowed for a prediction value less than one leaf away from the true value. This indicates that more complex models may be necessary to increase accuracy for wheat spikes since there is more room for obstruction in the wheat spike images.

While an error rate just over one spikelet is still fair, additional error in the model could be due to spikelets being more difficult to differentiate in some images. Spikelets get smaller as they reach the top of the spike, and are often partially obscured if awned. These differences in the spikes within each image could have hindered the ability of the model to learn certain patterns of the spikes. An increasing the number of counting models generated increases the number of annotated images available for training, making the use of pre-trained models or transfer learning more accessible for counting algorithms (Wang, Sun and Wang 2017). Use of these methods could help further improve the accuracy of the models, allowing for the efficient and reliable use of ML models for high-throughput phenotyping.

Another method commonly used for crowd counting, density mapping, could also be used for counting objects in an image (Gao et al. 2020). Density mapping works for counting tasks by identifying everywhere in an image it detects the object it has been trained to recognize, and has been used in agricultural settings for phenotyping using images in field Fields(Khaki et al. 2020). The number of annotation points on the density maps are able to be counted, giving the number of times an object of interest appears in the image. The ability for a model to be trained where the objects of interest are annotated could increase the accuracy of detection. Khaki et al. 2020, was able to overcome error caused by overlapping by using crowd counting methods with manually annotated images. Accuracies of corn stalk counting were measured across several different growth stages. Accuracies were highest between growth stages V2 and V4, which is

when each plant has emerged and has about three leaves. For the six models tested, MAEs ranged from 1.39 to 4.43 and MSEs ranged from 3.57 to 24.4. Even once plants and leaves were overlapping in stages V5 and V6, MAEs still ranged from 1.91 to 6.24 with MSEs ranging from 5.38 to 47.61. While accuracies decreased with an increase of overlapping, the comparable average errors indicate the robustness of this model, even with images not taken in a controlled setting. Density mapping could be useful in our study by training a density mapping model to detect spikelets on a wheat spike, which could potentially improve the accuracy of counting spikelets.

## **CONCLUSION**

This study used imaging techniques to analyze wheat spike characteristics and developed deep learning models for high-throughput analysis of wheat spike images. A comparison of classification and regression deep learning methods found that regression is more beneficial for observing error rates. CNN regression models were able to obtain an average error of 1.09 spikelets in the best fit model with spikelet values in analyzed images ranging from 11 spikelets to 32 spikelets. The use of detection-based methods could further improve this accuracy, increasing the utility of ML models as a means of HTP within breeding programs. Many methods are implemented by plant breeders for phenotyping crop traits, with little currently existing for wheat spike characteristics. Improving the efficiency of analyzing these traits will improve breeders' ability to discover and understand genetic components behind wheat spike architecture and their relationship to yield component traits. This will allow for the development of more environmentally efficient and higher-yielding cultivars to meet future demands.

## REFERENCES

- Abadi, M., A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean & M. Devin (2016) Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *arXiv preprint arXiv:1603.04467*.
- Araus, J. L., S. C. Kefauver, M. Zaman-Allah, M. S. Olsen & J. E. Cairns (2018) Translating High-Throughput Phenotyping into Genetic Gain. *Trends in plant science*, 23, 451-466.
- Bisong, E. 2019. Google Colaboratory. In *Building Machine Learning and Deep Learning Models on Google Cloud Platform: A Comprehensive Guide for Beginners*, 59-64. Berkeley, CA: Apress.
- Caroline, A. S., S. R. Wayne & W. E. Kevin (2012) NIH Image to ImageJ: 25 years of image analysis. *Nature methods*, 9, 671.
- Chollet, F. 2015. Keras. GitHub.
- Condorelli, G. E., M. Maccaferri, M. Newcomb, P. Andrade-Sanchez, J. W. White, A. N. French, G. Sciara, R. Ward & R. Tuberosa (2018) Comparative Aerial and Ground Based High Throughput Phenotyping for the Genetic Dissection of NDVI as a Proxy for Drought Adaptive Traits in Durum Wheat. *Frontiers in plant science*, 9, 893-893.
- Fahlgren, N. & A. G. M. S. W. C. S. D. W. B. S. T. H. C. J. M. a. N. W. I. K. T. F. S. T. Maximilian Feldman Malia (2015) A Versatile Phenotyping System and Analytics Platform Reveals Diverse Temporal Responses to Water Availability in Setaria. *Molecular plant*, 8, 1520-1535.
- Furbank, R. T. & M. Tester (2011) Phenomics – technologies to relieve the phenotyping bottleneck. *Trends in plant science*, 16, 635-644.
- Gao, G., J. Gao, Q. Liu, Q. Wang & Y. Wang (2020) Cnn-based density estimation and crowd counting: A survey. *arXiv preprint arXiv:2003.12783*.

- Gehan, M. A. & E. A. Kellogg (2017) High-throughput phenotyping. *American journal of botany*, 104, 505-508.
- Houle, D., D. R. Govindaraju & S. Omholt (2010) Phenomics: the next challenge. *Nature reviews. Genetics*, 11, 855-866.
- Huete, A., K. Didan, T. Miura, E. P. Rodriguez, X. Gao & L. G. Ferreira (2002) Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote sensing of environment*, 83, 195-213.
- Khaki, S., H. Pham, Y. Han, W. Kent & L. Wang (2020) High-throughput image-based plant stand count estimation using convolutional neural networks. *arXiv preprint arXiv:2010.12552*.
- Koboldt, Daniel C., Karyn M. Steinberg, David E. Larson, Richard K. Wilson & E. R. Mardis (2013) The Next-Generation Sequencing Revolution and Its Impact on Genomics. *Cell (Cambridge)*, 155, 27-38.
- Krause, M. R., S. Mondal, J. Crossa, R. P. Singh, F. Pinto, A. Haghhighattalab, S. Shrestha, J. Rutkoski, M. A. Gore, M. E. Sorrells & J. Poland (2020) Aerial high-throughput phenotyping enables indirect selection for grain yield at the early generation, seed-limited stages in breeding programs. *Crop science*, 60, 3096-3114.
- Lane, H. M. & S. C. Murray (2021) High throughput can produce better decisions than high accuracy when phenotyping plant populations. *Crop science*, 61, 3301-3313.
- LeCun, Y., Y. Bengio & G. Hinton (2015) Deep learning. *Nature (London)*, 521, 436-444.
- Miao, C., A. Guo, A. M. Thompson, J. Yang, Y. Ge & J. C. Schnable (2021) Automation of leaf counting in maize and sorghum using deep learning. *Plant phenome journal*, 4, n/a.

- Minervini, M., H. Scharf & S. A. Tsafaris (2015) Image Analysis: The New Bottleneck in Plant Phenotyping [Applications Corner]. *IEEE signal processing magazine*, 32, 126-131.
- Russakovsky, O., J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg & L. Fei-Fei (2015) ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision*, 115, 211-252.
- Simonyan, K. & A. Zisserman (2014) Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.
- Singh, A., B. Ganapathysubramanian, A. K. Singh & S. Sarkar (2016) Machine Learning for High-Throughput Stress Phenotyping in Plants. *Trends in plant science*, 21, 110-124.
- Tsafaris, S. A., M. Minervini & H. Scharf (2016) Machine Learning for Plant Phenotyping Needs Image Processing. *Trends in plant science*, 21, 989-991.
- Ubbens, J. R. & I. Stavness (2017) Deep Plant Phenomics: A Deep Learning Platform for Complex Plant Phenotyping Tasks. *Frontiers in plant science*, 8, 1190-1190.
- Wang, G., Y. Sun & J. Wang (2017) Automatic Image-Based Plant Disease Severity Estimation Using Deep Learning. *Computational intelligence and neuroscience*, 2017, 2917536-8.
- Winn, Z. J., D. L. Larkin, J. T. Murry, D. E. Moon & R. E. Mason (2021) Phenotyping Anther Extrusion of Wheat Using Image Analysis. *Agronomy (Basel)*, 11, 1244.

## CHAPTER III: GENOME-WIDE ASSOCIATION STUDY OF WHEAT SPIKE CHARACTERISTICS PHENOTYPED VIA IMAGING

### INTRODUCTION

#### The Flowering Pathway in Wheat

Flowering in plants is a regulated system controlled by several signaling pathways. In wheat (*Triticum aestivum* L.) these pathways determine flowering time, flower formation, and anthesis, which factor into the development of kernels and final grain yield (Mouradov, Cremer, & Coupland, 2002). While flowering pathways have been extensively studied in model systems such as *Arabidopsis thaliana*, they are not yet fully characterized in crops, including wheat. Gene families in wheat that contribute to variation in the timing of head emergence include the *Vrn* (70 to 75%), *Ppd* (20 to 25%), and *Eps* (5%) loci (Stelmakh, 1998). These gene families impact important traits in wheat, most notably flowering time, but can also be seen in the morphology of the spike including the SW, length, area, and number of spikelets.

Vernalization is the main determinant of heading date and is a regulatory factor in the transition from vegetative to reproductive growth. Wheat has three vernalization genes: *VRN1*, *VRN2*, *VRN3* (Dubcovsky et al., 2006; Yan, 2004). *VRN1* has three homoeologous copies *Vrn-A1*, *Vrn-B1*, and *Vrn-D1* on chromosomes 5A, 5B, and 5D, respectively. *VRN2* on chromosome 4A (Yan, 2004) and *VRN3* on chromosome 7B (Yan et al., 2006) have a smaller effect on vernalization than *VRN1* (Goncharov, 2004). The *Vrn-A1* gene has pleiotropic effects on total yield, tiller number, grain weight, and spikelet number, indicating the importance of flowering time (K. Kato, Miura, & Sawada, 2000). Homoeologous copies of the *Ppd* gene, *Ppd-A1*, *Ppd-B1*, and *Ppd-D1* are located on chromosomes 2A, 2B, and 2D, respectively (Scarath & Law, 1983; Snape, Butterworth, Whitechurch, & Worland, 2001) with *Ppd-D1* having the largest effect on photoperiod sensitivity followed by *Ppd-B1* and *Ppd-A1* (Worland, 1996). Photoperiod

alleles also affect flowering and play a role in spikelets spike<sup>-1</sup> (SPS) (Chen, Cheng, et al., 2020). The *Ppd-1* (*Ppd-A1a*) insensitive allele accelerates the spikelet initiation rate, thus decreasing SPS (Z. Chen et al., 2020; Zhaoyan Chen et al., 2020; Wurchum, Leiser, Langer, Tucker, & Longin, 2018). *Ppd-1* and vernalization loci are responsible for regulating *FLOWERING LOCUS T* (*FT*) expression, a heading date gene. *FT* can manipulate SPS by altering the duration of the spikelet initiation phase, which may also be influenced by *Eps*. Unlike *Ppd* and *Vrn*, *Eps* genes respond independently of environmental stimuli (Worland, 1996) and make up variation that occurs after *Ppd* and *Vrn* requirements have been fulfilled (H. Kato, Taketa, Ban, Iriki, & Murai, 2001). Since *Eps* have such a small effect on *FT*, they are usually mapped as quantitative trait loci (QTLs) rather than major genes (K. Kato, Miura, & Sawada, 1999). There is evidence supporting the influence of heading date on SPS and spikelet differentiation (Z. Chen et al., 2020; Miura & Worland, 1994). A study using near isogenic lines (NILs) showed that plants with the late *Eps* allele *Eps-A<sup>m</sup> 1-1* produced 8.7 more SPS than plants with the early allele (Lewis, Faricelli, Appendino, Valárik, & Dubcovsky, 2008). While the complete flowering pathway of wheat is unknown, information from other species and homologs aid in the understanding, allowing us to make more informed breeding decisions that affect economic traits of interest. To determine the remaining genetic components of the flowering pathway, additional phenotyping of wheat spikes is needed.

### ***Spikelets Spike<sup>-1</sup>***

As wheat spikes are the flower of the plant, there are multiple physiological and environmental factors that affect their formation. Spikelets spike<sup>-1</sup> (SPS) can be defined as the groups of florets positioned at each rachis node with an alternating pattern up the spike (Koppolu & Schnurbusch, 2019). Inflorescence architecture is determined by inflorescence meristems and floral meristems (Boss, Bastow, Mylne, & Dean, 2004), which form the reproductive organs of the spike (Koppolu & Schnurbusch, 2019; Sakuma, Salomon, & Komatsuda, 2011). These morphological traits are of interest because of their relationship to grain yield. Spikelets contain several individual florets and produce from two to four kernels per spikelet. Dimensional traits are important as the length of the rachis influences the number of

spikelets that can develop. SPS is positively correlated with grain number and is potentially an important morphological trait for increasing total grain yield (Chen, X. Cheng, et al., 2020) with a linear relationship with total grain yield up to 32 spikelets (Rawson, 1970).

In soft red winter wheat (SRWW), previous research showed a significant variety effect for the spike morphology traits spike area ( $F_{0.05,48,48} = 2.90, p = 0.0002$ ), spikelets spike<sup>-1</sup> ( $F_{0.05,48,48} = 2.06, p = 0.0069$ ), and spike SW ( $F_{0.05,48,48} = 1.98, p = 0.0096$ ) (Winn, Larkin, Murry, Moon, & Mason, 2021). Each of these traits regressed against area per spike also showed significant regressions for spike length ( $F_{0.05,1,48} = 15.25, R^2 = 0.2450, p = 0.0003$ ), spike width ( $F_{0.05,1,48} = 6.72, R^2 = 0.1251, p = 0.0127$ ), and approximate area per spike ( $F_{0.05,1,48} = 22.87, R^2 = 0.3273, p < 0.0001$ ). These results indicate that spike dimensions, area, and spikelets spike<sup>-1</sup> are beneficial traits to research in SRWW

Genome-wide association studies (GWAS) is a method used for identifying association between a molecular marker and a phenotype of interest general utilizing single nucleotide polymorphisms (SNPs) across the entire genome (Visscher et al. 2017). Using genome wide SNPs is useful for identifying small effect markers for traits. Genomic selection (GS) utilizes both genotypes and phenotypes to estimate breeding values within a population (Ganal et al., 2019). GS uses markers from across the genome, allowing for the use of small effect alleles for estimations. GS can increase the rate of genetic gain by allowing for the selection of traits earlier in a breeding cycle (Jannick et al. 2010). GWAS and GS are beneficial tools for the identification of genetic components related to flowering and yield component traits and the selection of these favorable traits within a program.

To better understand the genetics of wheat spike morphology, a panel of 594 diverse soft red winter wheat lines adapted to the southeastern United States was phenotyped using image analysis for spike width, spike length area spike<sup>-1</sup> (APS), and spikelets spike<sup>-1</sup>(SPS). Phenotypic data and genotyping-by-sequencing (GBS) data were used to perform GWAS to find significant marker-trait associations (MTA) related to spike width, length, APS, and SPS; The MTAs drawn from the GWAS will be used to inform selections of parental lines via genomic selection.

## **MATERIALS AND METHODS**

### **Plant Materials**

The Historic Gulf Atlantic Wheat Nursery (HGAWN) is a population consisting of 594 soft red winter wheat lines from public breeding institutions in the southeastern United States. Of these, 103 were developed by the University of Arkansas, 109 from Louisiana State University, 105 from the University of Georgia, 104 from North Carolina State University, 60 from Texas A&M University, 44 from Virginia Polytechnic Institute and State University, 19 from Clemson University, and 9 from the United States Department of Agriculture Agricultural Research Service (USDA-ARS) in Raleigh, North Carolina.

The HGAWN was planted during the 2018-2019 and 2019-2020 growing seasons in October of the previous year in a randomized complete block design with two replications in Fayetteville, Arkansas at the Milo J. Shult Agricultural Research and Extension Center. Each experimental unit (plot) consisted of a single row measuring 1.20 m in length with 0.38 m between adjacent plots. Nutrient management was determined by soil sampling. For 2019, 67 kilograms of urea and 0.28 kilograms per hectare of Harmony® Extra were applied in late February. In early March, Axial® was applied at a rate of 0.6 kilograms per hectare to control for ryegrass. Urea was applied again in late March at a rate of 33 kilograms per hectare.

### **Trait phenotyping**

Heading date (HD) was collected as the date on which half the spikes on the primary tiller of each plot had emerged from the boot during flowering. Yield components were determined from a sample of 10 spike-bearing culms collected from each plot after physiological maturity. Each sample was threshed and kernels counted using a DATA Count Jr. seed counter (DATA technologies). Kernels were weighed

to determine kernel weight spike<sup>-1</sup> and a 1000 kernel weight (TKW). The number of kernels was divided by number of spikes threshed to get average kernel number spike<sup>-1</sup> (KNS). Total kernels harvested was weighed for total kernel weight measurements. Imaging of wheat spikes and the analysis of traits follow methods detailed in chapter 2 of this thesis and by (Winn et al., 2021). Briefly, spike characteristics were measured from images taken 30 to 33 days after heading with a Canon® Powershot G1 X Mark II camera. Six spikes were harvested from each plot for imaging. Each spike was imaged rachis side up and ImageJ version 1.52o (Schneider, Rasband, & Eliceiri, 2012) was used to collect spike width (SW), spike length (SL), spike area (SA), and spikelets spike<sup>-1</sup> (SPS).

### **Genotyping**

All varieties in the HGAWN population were genotyped using a genotyping-by-sequencing (GBS) approach. Library prep was done in Raleigh, North Carolina by the USDA Eastern Regional Small Grains Genotyping Lab and sequencing was done at the sequencing lab at North Carolina State University. Deoxyribonucleic acid (DNA) was extracted using a Mag-Bind® Plus kit from Omega Bio-tek using the instructions included by the manufacturer. DNA was quantified using Quant-iT™ PicoGreen® dsDNA Assay Kits and standardized to 20 nanograms per microliter concentrations. GBS libraries were prepared using the Pst1-Msp1 or the Pst1-Mse1 restriction enzymes. Adapters were ligated to each line and multiplexed at 192-plex to create libraries, which were then sequenced on a lane of an Illumina Hi-Seq 2500 sequencer. Detected single nucleotide polymorphisms (SNPs) were aligned to the RefSeq v1.0 wheat reference genome using the trait analysis by association, evolution, and linkage (TASSEL) GBS v2 pipeline with a 64 base kmer length and a minimum kmer count of 5 (Bradbury et al., 2007) .

### **Statistical Analysis**

All analysis were performed in R version 4.1.1. Individual site-year and multi-site-year BLUPs were calculated using ASReml-R version 4.1.0.160 (Gilmour, Gogel, Cullis, Welham, & Thompson,

2015). For individual site-years, data were analyzed using 10 spatial models with the best fit model ultimately being an anisotropic exponential variance model. This was modeled using the equation:

$$y = X\beta + Z\mu + \varepsilon$$

Where X is the design matrix for fixed effects and  $\beta$  is a vector of coefficients for fixed effects, Z is a design matrix for random effects and  $\mu$  is a vector of coefficients for random effects, and  $\varepsilon$  is a vector of errors with an anisotropic exponential variance model with the structure:

$$C_{ij} = \phi_1^{|x_i - x_j|} \phi_2^{|y_i - y_j|}$$

Where x and y are vectors of coordinates and  $|\phi_1| < 1$  and  $|\phi_2| < 1$ . The mean of each trait was used as a fixed effects and the genotypes were used as random effects (Gilmour et al., 2015).

Multi-site-year analysis treated pairs of years and locations as individual environments and was modeled using the equation:

$$y = X\beta + Z\mu + \varepsilon$$

Where X is the design matrix for fixed effects and  $\beta$  is a vector of coefficients for fixed effects, Z is a design matrix of coefficients for random effects and  $\mu$  is a vector of random effects, and  $\varepsilon$  is a vector of errors. The mean of each trait and the environment were treated as fixed effects in the model, and genotype by year interaction where constant correlation between pairs of environments was assumed was treated as the random effect.

Correlations were calculated using functions in base R. Heritability (or repeatability) within each site-year was determined using the general broad-sense heritability equation:

$$H_{Standard} = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_r^2}$$

Where  $\sigma_g^2$  is the genetic variance and  $\sigma_r^2$  is the residual variance. Due to unbalanced sampling across years, multi-site-year heritability was estimated using the method proposed by Cullis, Smith, and Coombes, 2006 in the lme4 version 1.1-28 package (Bates, Mächler, Bolker, & Walker, 2015). The following formula was used:

$$\bar{H}_{Cullis}^2 = 1 - \frac{\bar{v}_{\Delta}^{BLUP}}{2\sigma_g^2}$$

Where  $\sigma_g^2$  is the genotypic variance and  $\bar{v}_{\Delta}^{BLUP}$  is the mean variance of a difference of two genotypic best linear unbiased predictions (BLUPs).

### ***GWAS and GS***

The BLUPs generated for yield components and spike morphological traits were used in the genome wide association studies (GWAS). GWAS were performed using multi-locus models from the mrMLM and mrMLM.GUI version 4.0.2 packages (Y.-W. Zhang et al., 2020). A principal component analysis (PCA) was used to account for population structure using the LEA R package version 3.6.0 (Frichot, François, & O'Meara, 2015) and a kinship matrix was generated through mrMLM.GUI to account for relatedness. A mixed linear model (MLM) approach allowed for the replacement of the typical Bonferroni correction to create a less stringent selection criterion. This was done by replacing it

with other techniques for Bayesian LASSO or empirical Bayes method, or only implementing it using the effective markers for base MLMs (Wang et al., 2016). Six different models were employed in the mrMLM package. Additive effects were calculated by fitting a regression model with the marker as the independent variable and the phenotype of interest as the dependent variable where the slope of the model describes the additive effect.

Random MLMs (RMLMs) follow the basic model

$$y = X\alpha + Z_k\gamma_k + \xi + \varepsilon$$

Where  $y$  is a vector of phenotypic values for all individuals,  $X$  is an incident matrix for non-genetic fixed effects,  $\alpha$  is a vector of coefficients for fixed effects,  $Z_k$  is a vector of individuals with the  $k_{th}$  SNP,  $\gamma_k$  is the marker effect of the marker,  $k$ ,  $\xi$  is a vector of polygenic effects for each genotype ( $\xi \sim N(0, K_\phi^2)$ ) with a normal distribution of mean zero and variance  $K_\phi^2$ , and  $\varepsilon$  is residual error (Wang et al., 2016).

Multi-Locus Random-SNP-Effect Mixed Linear Model (mrMLM) is based on the random effect mixed linear model (RMLM)(Wang et al., 2016), which used a modified Bonferroni adjustment by dividing the significance value by an effective value of SNPs instead of using all SNPs. The RMLM is used like a first stage screening analysis in the mrMLM model to obtain all markers that have a P-value less than 0.01, and eliminate any consecutive makers to eliminate collinearity. Of all remaining markers, those that pass the modified Bonferroni adjustment are used to conduct a likelihood ratio test (LRT). Any markers that do not pass the Bonferroni adjustment or have a LOD score higher than 1.5 for the LRT are

treated as fixed in the model while all other markers are treated as random. FAST multi-locus random-SNP-effect Mixed Linear (FASTmrMLM), proposed by Tamba and Zhang, 2018, uses a different model transformation and least angle regression methods for identifying potentially associated SNPs, reducing the runtime of the mrMLM method.

Model FAST multi-locus random-SNP-effect EMMA (FASTmrEMMA), proposed by Wen et al., 2017, is also a two-stage GWAS with three alterations in the first stage to enable a faster runtime. A new matrix transformation was implemented to multiply the original MLM to whiten the covariance matrix of the polygenic matrix and environmental noise, a polygenic to residual variance ratio was fixed to all single marker genome tests, and all non-zero eigenvalue vectors were specified to one. In the second stage, all SNPs selected in the first stage are placed into an MLM and estimated using expectation and maximization empirical Bayes (EMEB) for quantitative trait nucleotide (QTN) identification.

Iterative Sure Independence Screening EM-Bayesian least absolute shrinkage and selection operator (EM-BLASSO), proposed by Tamba, Ni, and Zhang, 2017, is based on the linear mixed model

$$Y = \sum_{j=1}^q X_j \beta_j + \sum_{k=1}^p Z_k \gamma_k + \varepsilon$$

Where  $X_j$  is the design matrix and  $\beta_j$  is the  $j^{th}$  non-QTN effect,  $Z_k$  is the corresponding incidence matrix determined by genotypes for the locus  $k$ ,  $\gamma_k$  is the vector of SNP effects for locus  $k$ , and  $\varepsilon$  is the residual error term. The first stage of EM-BLASSO screens SNPs using correlations between predictors and responses. Kinship and population structure are treated as fixed effects in the model to correct phenotypes. A modified version of sure independence screening smoothly clipped absolute deviation (SIS-SCAD) is used to select models based on the order of their significant marginal correlation. In the

second stage, the estimation stage, the EM-Bayesian LASSO algorithm is used to filter and estimate true effects using

$$Y = X\beta + \sum_{k=1}^{\tau} Z_k\gamma_k + \varepsilon$$

Where uncorrected phenotype values are used and denoted by  $Y$ , meaning  $Z_k$  denotes the  $k$ th SNP value and  $X$  is the overall mean and population structure. The variance components and residual variance estimator are designed so that at convergence  $\hat{\gamma}_k = E(\gamma_k)$  and the prediction error for  $\gamma_k$  is  $Var(\gamma_k)$ .

Polygene-Background-Control-Based Least Angle Regression plus Empirical Bayes (pLARmEB), proposed by J. Zhang et al., 2017, uses least angle regression and empirical Bayes for GWAS. The genotypic model is expressed by

$$y = 1\mu + W\alpha + Z\gamma + u + \varepsilon$$

Where  $y$  is the phenotypic value of the  $i^{th}$  individual from sample size  $n$ ,  $1$  is a  $n \times 1$  vector,  $\mu$  is the total average,  $W$  is the design matrix for  $\alpha$  and  $\alpha$  is the fixed effect population structure,  $Z$  is the design matrix for  $\gamma$  and  $\gamma$  is random QTN effects,  $u$  is polygenic effects, and  $\varepsilon$  is the residual error term. LARS was used for selecting SNPs that were most likely associated with traits of interest. Estimations are then made using the empirical Bayes reduced model, which is expressed by

$$y = X\beta + Z\gamma + \varepsilon$$

Where  $y$  is the phenotypic value,  $X$  is the design matrix for fixed effects,  $\beta$  is the vector of fixed effects,  $Z$  is the design matrix for random effects,  $\gamma$  is the vector of random effects, and  $\varepsilon$  is the residual error term. The variable selection. By LARS allows the number of variables in the model to be few enough for estimation using empirical Bayes.

Kruskal-Wallis test with Empirical Bayes (PKWmEB), proposed by Ren, Wen, Dunwell, and Zhang, 2018, follows the same genetic model as pLARmEB, but also contains a polygenic background correction. Based on the polygenic background model, the Kruskal-Wallis test is used to detect if a SNP is associated with the trait of interest. Empirical Bayes and the likelihood ratio test are then used for marker effect estimation.

For genomic selection, version 4.6.1 of the R package rrBLUP (Endelman, 2011) was used to calculate genomic estimated breeding values (GEBVs) using the mixed linear model:

$$y = X\beta + [Z \ 0]g + \varepsilon$$

Where  $\beta$  is a vector of fixed effects,  $g$  is a vector of random genotypic values with covariance structure  $G = Var[g]$ . Residuals follow the structure  $Var[\varepsilon_i] = R_i\sigma_\varepsilon^2$  with  $R_i = 1$  as the default. The population mean was included as the fixed effect. Model accuracy was assessed using fivefold cross-validation, where genotypes are assigned to one of five folds. The four remaining folds train the model and attempt to predict the GEBVs of the fifth fold. The predicted GEBVs and observed BLUPs are then plotted against each other to find the prediction accuracy of the model. Cross-validation was run for 20 cycles.

## RESULTS

### Analysis of phenotypic data

Summary statistics for individual and multi-site-year analyses are shown in Table 3.1 and Figure 3.1.

Trait heritability ranged from 0.52 to 0.90, 0.64 to 0.84 and 0.49 to 0.85 for the 2019, 2020, and combined analyses, respectively.

**Table 3.1.** Summary statistics and heritability of traits measured on the Historical Gulf Atlantic Wheat Nursery in Fayetteville, Arkansas for the 2019 and 2020 seasons and combined across years.

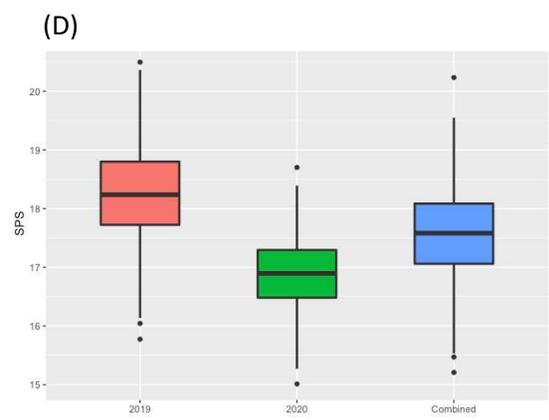
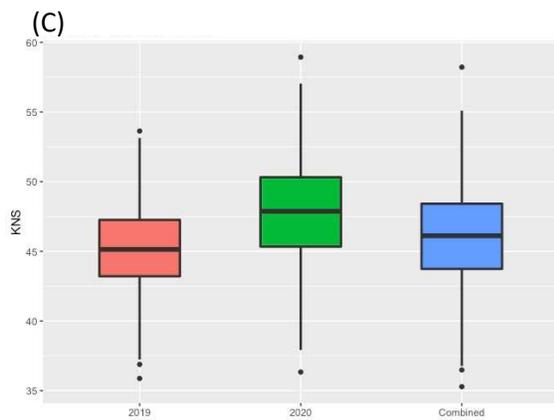
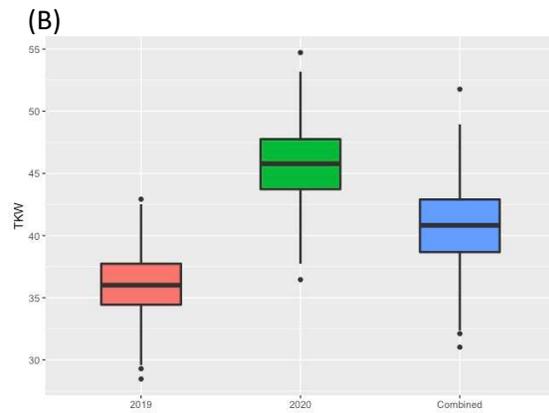
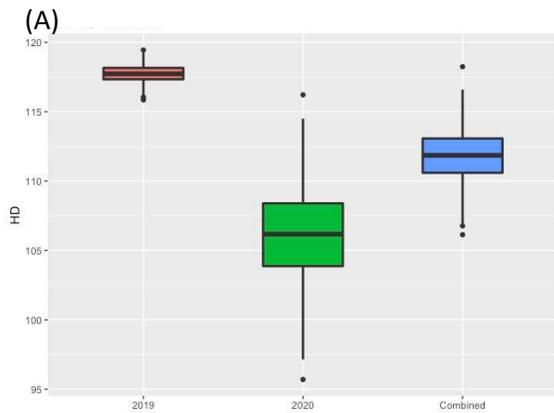
Trait	Mean	Range	Heritability	$\sqrt{H}$	$r_{gs}$
<u>2019</u>					
Heading (Julian days)	117.9	113 - 125	0.79	0.89	0.31
1000 kernel weight (g)	36.1	29.1 - 113.1	0.52	0.72	0.38
Kernels spike <sup>-1</sup>	45.4	24.4 - 81.2	0.58	0.76	0.25
Spikelets spike <sup>-1</sup>	18.25	12.8 - 21.2	0.62	0.79	0.33
Spike width (mm)	9.9	8.8 - 13.5	0.62	0.79	0.38
Spike length (mm)	86.3	65.4 - 131.4	0.90	0.95	0.33
Spike area (mm <sup>2</sup> )	1073	584 - 2321	0.62	0.79	0.56
<u>2020</u>					
Heading (Julian days)	105.934	111 - 123	0.84	0.92	0.35
1000 kernel weight (g)	45.767	18.1 - 51.1	0.64	0.80	0.54
Kernels spike <sup>-1</sup>	47.437	16.6 - 77.4	0.80	0.89	0.51
Spikelets spike <sup>-1</sup>	17.002	13.3 - 23.2	0.69	0.83	0.35
Spike width (mm)	10.439	7 - 13.2	0.69	0.83	0.40
Spike length (mm)	87.505	57.8 - 127.6	0.79	0.89	0.50
Spike area (mm <sup>2</sup> )	1079.201	448 - 2537	0.64	0.80	0.60
<u>Combined analysis</u>					
Heading (Julian days)	111.9	97 - 125	0.49	0.70	0.30
1000 kernel weight (g)	40.9	18.1 - 60.6	0.85	0.92	0.53
Kernels spike <sup>-1</sup>	46.3	19.5 - 81.2	0.73	0.85	0.43
Spikelets spike <sup>-1</sup>	17.6	12.8 - 23.3	0.78	0.88	0.42

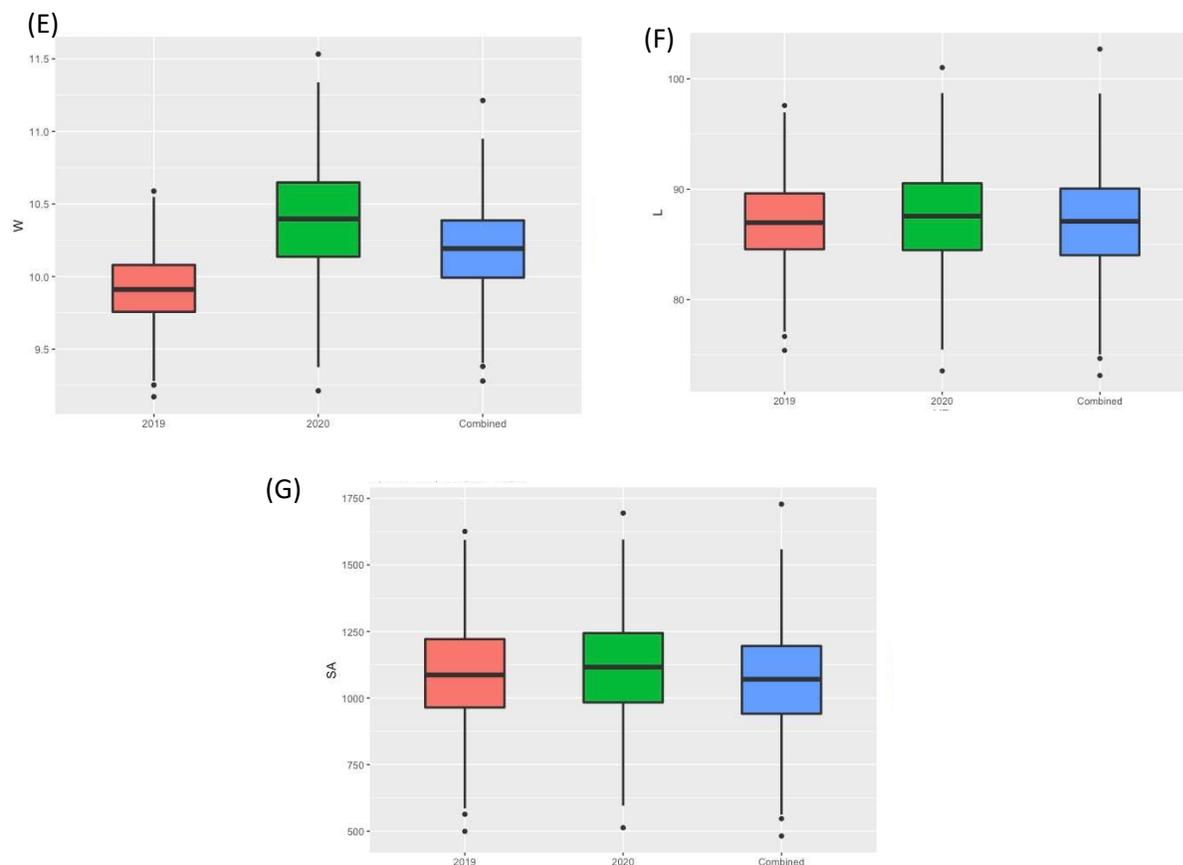
Spike width (mm)	10.1	7 - 13.5	0.63	0.79	0.46
Spike length (mm)	86.9	59.3 - 131.4	0.76	0.87	0.50
Spike area (mm <sup>2</sup> )	1075	448 - 2537	0.85	0.92	0.62

H = Heritability

$\sqrt{H}$  = Square root of heritability

$r_{gs}$  = Genomic selection accuracy





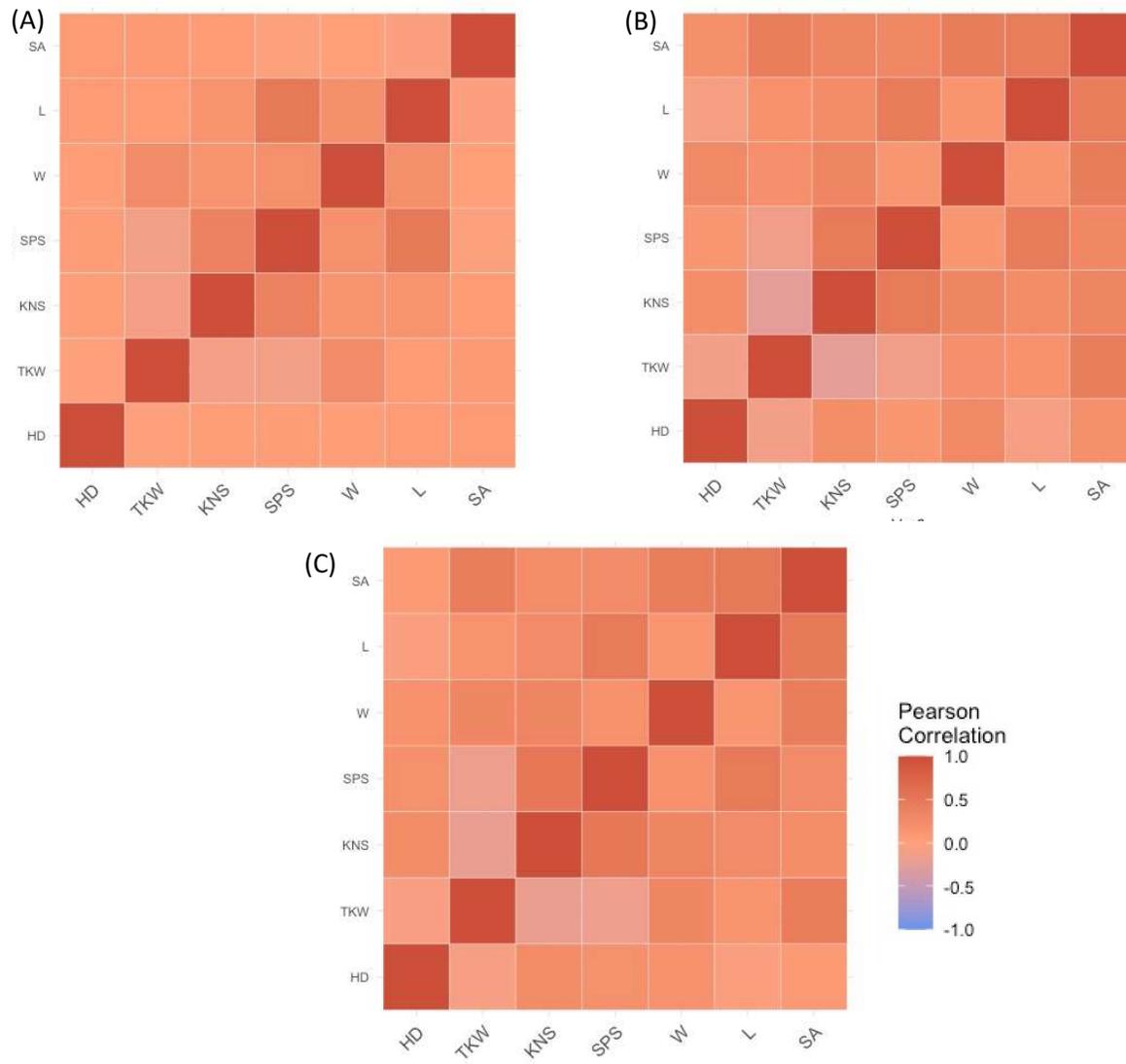
**Figure 3.1.** Distribution of best linear unbiased predictors (BLUPs) for (A) heading date (Julian days)(HD), (B) 1000 kernel weight (g)(TKW), (C) kernels spike<sup>-1</sup> (KNS), (D) spikelets spike<sup>-1</sup> (SPS), (E) spike width (mm)(SW), (F) spike length (mm) (SL), and (G) spike area (mm<sup>2</sup>) (SA) from the Historic Gulf Atlantic Wheat Nursery (HGAWN) for 2019, 2020, and the combined years.

Phenotypic correlations were generally consisted across seasons. In 2019, TKW was negatively correlated with SPS ( $r = -0.11$ ,  $p < 0.01$ ) and KNS ( $r = -0.096$ ,  $p < 0.05$ ). KNS was positively correlated with SPS ( $r = 0.37$ ,  $p < 0.0001$ ), spike width (SW) ( $r = 0.14$ ,  $p < 0.001$ ) and spike length (SL) ( $r = 0.49$ ,  $p < 0.001$ ) and was negatively correlated with TKW ( $r = -0.09$ ,  $p < 0.05$ ). SL and SPS had the highest correlation of ( $r = 0.49$ ,  $p < 0.0001$ ). SA was not significantly correlated with any traits. (Figure 3.2A).

In 2020, TKW was again negatively correlated with SPS ( $r = -0.14$ ,  $p < 0.007$ ) and KNS ( $r = -0.24$ ,  $p < 0.0001$ ). KNS was positively correlated with SPS ( $r = 0.47$ ,  $p < 0.0001$ ) and SL ( $r = 0.25$ ,  $p < 0.0001$ ).

0.0001). KNS and SW had the highest correlation ( $r = 0.459, p < 0.0001$ ), and SL and SPS had the second highest correlation ( $r = 0.45, p < 0.0001$ ). All yield component traits had a significant relationship with other yield component traits (Figure 3.2B).

When combined across years, TKW was negatively correlated with SPS ( $r = -0.15, p < 0.001$ ) and KNS ( $r = -0.19, p < 0.0001$ ). KNS was positively correlated with SPS ( $r = 0.49, p < 0.0001$ ) and with SL ( $r = 0.27, p < 0.0001$ ). SPS and KNS had the highest correlation and each had a significant relationship with all yield component traits (Figure 3.2C).



**Figure 3.2.** Heatmaps of best linear unbiased prediction (BLUP) correlations between heading date (HD), spike area (SA), spike length (SL), spike width (SW), spikelets spike<sup>-1</sup> (SPS), kernels spike<sup>-1</sup> (KNS), and 1000 kernel weight (TKW) for the Historic Gulf Atlantic Wheat Nursery (HGAWN) for the (A) 2019, (B) 2020, and (C) combined years.

### **Summary of Marker Trait Associations**

For the combined analysis, 32 MTA were identified for HD, 33 for TKW, 19 for KNS, 38 for SPS, 26 for SA, 19 for SW, and 26 for SL. All seven traits had MTA identified by all six models except for HD and SW, where no MTA were identified by FASTmrEMMA. The least stringent models were pLARmEB and ISIS EM-BLASSO, which identified 47 and 53 MTA across all traits, respectively. FASTmrEMMA was the most stringent of the models with only 10 MTA being identified between all seven traits. While the results presented herein will focus on SPS and related MTA, the full results for significant MTAs are in Supplemental Table 1.

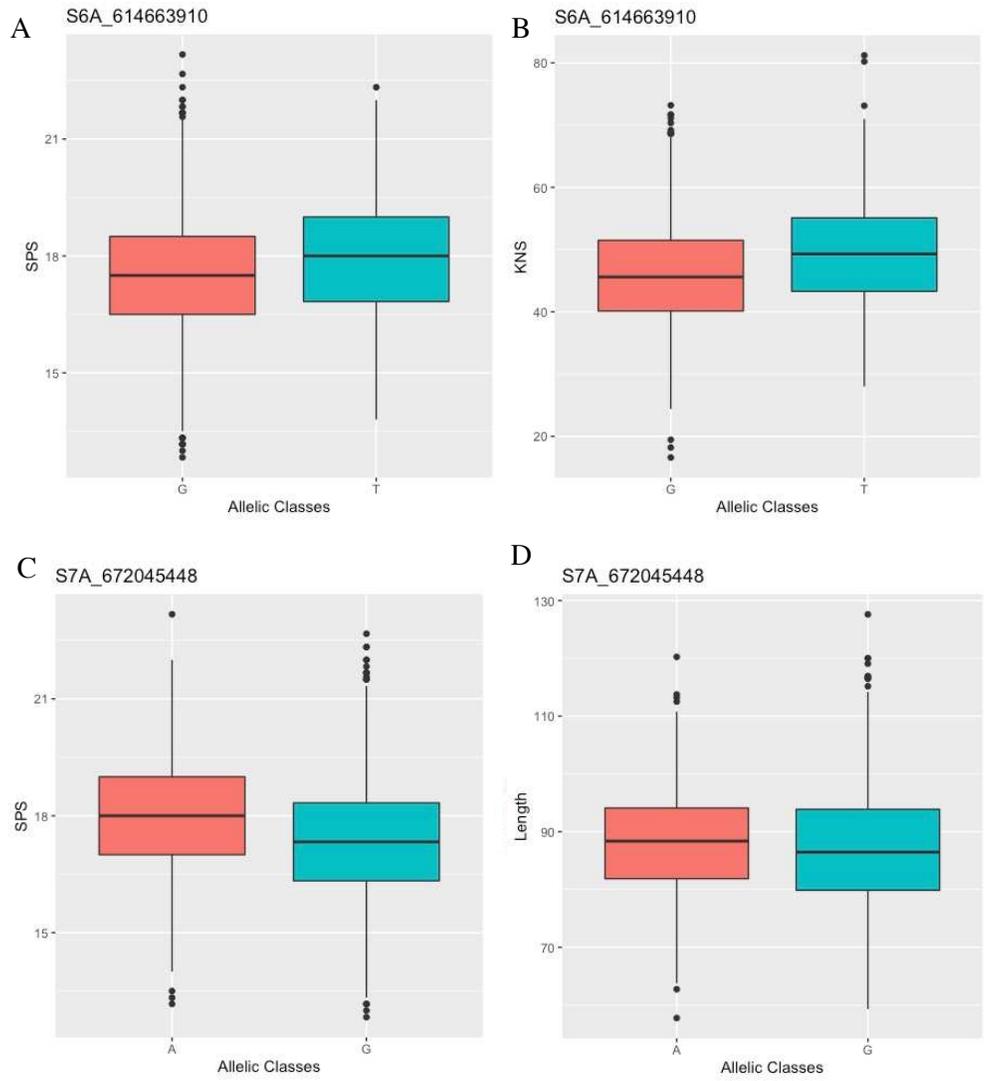
### ***MTA for Spikelets Spike<sup>-1</sup> and Related Traits***

SPS had a large number of MTA identified, with several part of a large peak on chromosome 7A. SNP S7A\_672854561 was the only SNP in the peak to be identified by multiple models and had an LOD ranging from 6.69 to 7.16 and the minor allele had an additive effect of -0.77 spikelets. S7A\_672045448 and S7A\_672148737 were the most significant SNPs in the peak with LOD of 11.65 and 10.31 and minor allele additive effects of 0.65 and 0.68 spikelets, respectively. Significant markers on 7A generally explained less than  $R^2 < 10\%$ , except for S7A\_673066304 which had an  $R^2 = 11.1\%$  and a minor allele additive effect of -0.74. One SNP on chromosome 5A, S5B\_590471022, was identified by all six models and had LOD ranging from 4.35 to 7.51 and had and the minor allele had an additive effect of 0.21 spikelets. All identified MTA on chromosome 5B were small-effect and each had an  $R^2 < 5\%$ .

SNP S7A\_672045448 was significant for both SL and SPS and was identified for SL by four of the six models. LOD ranged from 4.94 to 6.97 with a minor allele additive effect of 1.2 mm for SL with a LOD of 11.65. S6A\_614663910 on chromosome 6A was identified for both KNS and SPS with LODs of 6.1 and 5.87, respectively. S6A\_614663910 had a minor allele additive effect of 3.82 kernels for KNS and an additive effect of 0.44 spikelets for SPS. Differences in phenotypes between allelic classes for SPS and KNS with marker S6A\_614663910 and SPS and SL for marker S7A\_672045448 are shown in figure 4. An analysis of variance (ANOVA) test showed that the BLUPs were significantly different between allelic classes. For S7A\_672045448, SL had the lowest significant difference ( $p < 0.05$ ) while SPS for both markers and KNS for S7A\_672045448 all had  $p$  values less than 0.0001. This indicates that a T allele at S6A\_614663910 confers higher SPS and KNS while an A allele at S7A\_672045448 increases SL and SPS. QTLs S5B\_590471022 and S7A\_673066304 were also identified in the 2019 and combined analyses and QTLs S7A\_672148737 and S7A\_672854561 were identified in the 2020 and combined analyses.

### **Candidate genes underlying identified MTA**

SNPs were searched in the International Wheat Genome Sequencing Consortium (IWGSC) Chinese Spring wheat RefSeq v2.1 (Zhu et al., 2021) to look for potential candidate genes. SNP S7A\_672045448 was the only marker location associated with a gene, TraesCS7A03G1143700. The gene is a reverse strand that is 1,161 base pairs long and encodes for an F-box like domain superfamily protein, meaning it governs protein to protein interactions. All other significant markers on 7A were not associated with candidate genes, but were located before the promotor region of *WAP01*, a gene influencing spikelet number.



**Figure 3.3.** (A) SPS across the population by allelic classes of S6A\_614663910, (B) KNS across the population by allelic classes of S6A\_614663910, (C) SPS across the population by allelic classes of S7A\_672045448, (D) SL across the population by allelic classes of S7A\_672045448. Comparisons of phenotypes to alleles do not show large difference in phenotypes based on genetics.

### Genomic Prediction of Spike Characteristics

GS accuracies (Table 3.1) were lower in 2019 compared to 2020 and the combined analyses. SA had the highest prediction accuracies, ranging from 0.56 to 0.62. HD had the lowest average accuracy with the highest being 0.35 for 2020 and the lowest being 0.30 for combined years. SPS had moderate

prediction accuracy ranging from 0.33 in 2019 to 0.42 for combined years. Moderately high accuracies were expected due to the highly quantitative nature of yield component traits. While these traits are complexly related, these results indicate the use of genomic selection for spike architecture traits such as SPS, SW, SL, and SA. For all three analyses, the square root of heritability ( $\sqrt{H}$ ) was higher than  $r_{gs}$  for all traits, indicating heritability indices may provide a higher selection accuracy for these traits.

## DISCUSSION

Trait correlations were similar to those previously reported (K. Kato et al., 2000). TKW was negatively correlated with KNS, SPS, and HD for both years. However, KNS and SPS are positively correlated with each other. This is consistent with the idea that for a spike to produce more kernels, less assimilate is available to have larger size and weight for kernels produced by the spike (Mangini et al., 2018). While yield component traits are complexly inherited, understanding the relationships between them helps aid in breeding decisions.

This region on chromosome 7A has been linked to SPS in previous studies (Wurchum et al., 2018; Xu et al., 2013; Zhai et al., 2016) and was cloned as *WAPO-AI* (Kuzay et al., 2022). *WAPOI* is expressed in the inflorescence, spikelet, and floral meristems, and allelic variation was found to influence SPS. Loss of function in *WAPO-AI* results in abnormal spike morphology. *WAPOI* was mapped between flanking markers AX-109397893 and IWA5913 at 673,854,124 bp and 674,276,906 bp, respectively (Kuzay et al., 2019) with the promoter region beginning at 674,080,862 bp (Kuzay et al., 2022). This puts all three markers on chromosome 7A for this study, within the range of 672,045,448 bp to 673,066,304 bp, before the promoter region of the gene. This could be due to recombination effects, or gene clustering of new candidate genes of interest not discovered with the initial *WAPO-AI* QTL region (Kuzay et al., 2019) which has been observed in similar studies (Xu et al., 2013). One marker (S7A\_672045448) had an underlying protein-coding reverse sequence gene, while no other markers were located close to any exons included in the v2.1 reference sequence (Alaux et al., 2018). Aside from markers on chromosome 7A,

studies identified QTL and markers of interest for spikelets per spike and chromosome 5B as well (Cao et al., 2019). Studies identifying significant QTL and MTA for SPS commonly find only large associations with markers near *WAP0-1A* (Kuzay et al., 2022; Kuzay et al., 2019). Use of the gene as a fixed effect may allow for the emergence of other markers across multiple studies.

The goal of this study was to improve understanding of the flowering pathway of wheat by identifying underlying genomic regions associated with wheat spike characteristics. The use of mixed linear models enabled more significant regions in the genome to be identified, including two markers associated with both SPS and other spike architecture and yield component traits.

Grain yield component traits and spike architecture traits are complexly inherited and are more difficult to predict. Comparisons of GS accuracy and  $\sqrt{H}$  indicated that heritability is a better prediction index for all traits in this study. However, GS accuracy was still moderately high for these traits, providing it with some value. Accuracies for GS could potentially be improved by using large effect markers and fixed effects in the model, and can still serve as a selection criterion for unphenotyped material.

A major limitation to this study and others involving phenotyping wheat spikes is the large amount of time and labor required to phenotype wheat spikes. There are currently no automated methods available, leaving the manual phenotyping of spikes or images of spikes as the only methods of data collection. This hinders the power of analysis by limiting the amount of phenotypic data available.

## CONCLUSION

The results support that the *WAP01* gene influences SPS in SRWW within the HGAWN population. While showing the correlations between traits such as SL and KNS on SPS, it also showed influence from markers that were significantly related to SPS and SL or SPS and KNS. Development of MTAs behind spike architecture and yield component traits will help further understanding of the relationship between the two. Knowledge of the spike architecture and its effects on yield could aid in

breeding endeavors. Phenotyping wheat spikes in the field takes extensive time and labor, making it difficult to perform at a large scale. However, GS can be used as a method for the prediction of individuals in SRWW germplasm. The understanding of genetic controls for flowering and spike development, as well as their predictions, could benefit plant breeders seeking ways to maintain or increase yield as other traits diminish it.

## REFERENCES

- Alaux, M., Rogers, J., Letellier, T., Flores, R., Alfama, F., Pommier, C., et al. (2018). Linking the International Wheat Genome Sequencing Consortium bread wheat reference genome sequence to wheat genetic and phenomic data. *Genome Biology*, *19*(1), 111.
- Bates, Mächler, Bolker, & Walker. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of statistical software*, *67*(1), 1-48.
- Boss, P., Bastow, R., Mylne, J., & Dean, C. (2004). Multiple Pathways in the Decision to Flower: Enabling, Promoting, and Resetting. *The Plant cell*, *16*(suppl\_1), S18-S31.
- Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., & Buckler, E. S. (2007). TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics*, *23*(19), 2633-2635.
- Cao, P., Liang, X., Zhao, H., Feng, B., Xu, E., Wang, L., et al. (2019). Identification of the quantitative trait loci controlling spike-related traits in hexaploid wheat (*Triticum aestivum* L.). *Planta*, *250*(6), 1967-1981.
- Chen, Cheng, Chai, Wang, Du, Bian, et al. (2020). Pleiotropic QTL influencing spikelet number and heading date in common wheat (*Triticum aestivum* L.). *Theor Appl Genet*, *133*(6), 1825-1838.
- Chen, Cheng, X., Chai, L., Wang, Z., Du, D., Bian, R., et al. (2020). Pleiotropic QTL influencing spikelet number and heading date in common wheat (*Triticum aestivum* L.). *Theor Appl Genet*, *133*(6), 1825-1838.

- Chen, Z., Cheng, X., Chai, L., Wang, Z., Du, D., Bian, R., et al. (2020). Pleiotropic QTL influencing spikelet number and heading date in common wheat (*Triticum aestivum* L.). *Theor Appl Genet*, *133*(6), 1825-1838.
- Chen, Z., Cheng, X., Chai, L., Wang, Z., Du, D., Wang, Z., et al. (2020). Pleiotropic QTL influencing spikelet number and heading date in common wheat (*Triticum aestivum* L.). *Theoretical and applied genetics*, *133*(6), 1825-1838.
- Cullis, B. R., Smith, A. B., & Coombes, N. E. (2006). On the Design of Early Generation Variety Trials with Correlated Data. *Journal of agricultural, biological, and environmental statistics*, *11*(4), 381-393.
- Dubcovsky, J., Loukoianov, A., Fu, D., Valarik, M., Sanchez, A., & Yan, L. (2006). Effect of Photoperiod on the Regulation of Wheat Vernalization Genes VRN1 and VRN2. *Plant molecular biology*, *60*(4), 469-480.
- Endelman, J. B. (2011). Ridge Regression and Other Kernels for Genomic Selection with R Package rrBLUP. *The plant genome*, *4*(3), 250-255.
- Frichot, E., François, O., & O'Meara, B. (2015). LEA: An R package for landscape and ecological association studies. *Methods in ecology and evolution*, *6*(8), 925-929.
- Gilmour, A. R., Gogel, B. J., Cullis, B. R., Welham, S. J., & Thompson, R. (2015). ASReml User Guide Release 4.1 Functional Specification, . Hemel Hempstead, UK: VSN International Ltd.
- Goncharov, N. P. (2004). Response to vernalization in wheat: its quantitative or qualitative nature. *Cereal research communications*, *32*(3), 323-330.
- Kato, H., Taketa, S., Ban, T., Iriki, N., & Murai, K. (2001). The influence of a spring habit gene, Vrn-D1, on heading time in wheat. *Plant breeding*, *120*(2), 115-120.

- Kato, K., Miura, H., & Sawada, S. (1999). Detection of an earliness per se quantitative trait locus in the proximal region of wheat chromosome 5AL. *Plant breeding*, *118*(5), 391-394.
- Kato, K., Miura, H., & Sawada, S. (2000). Mapping QTLs controlling grain yield and its components on chromosome 5A of wheat. *Theoretical and applied genetics*, *101*(7), 1114-1121.
- Koppolu, R., & Schnurbusch, T. (2019). Developmental pathways for shaping spike inflorescence architecture in barley and wheat. *J Integr Plant Biol*, *61*(3), 278-295.
- Kuzay, S., Lin, H., Li, C., Chen, S., Woods, D. P., Zhang, J., et al. (2022). WAPO-A1 is the causal gene of the 7AL QTL for spikelet number per spike in wheat. *PLoS genetics*, *18*(1), e1009747-e1009747.
- Kuzay, S., Xu, Y., Zhang, J., Katz, A., Pearce, S., Su, Z., et al. (2019). Identification of a candidate gene for a QTL for spikelet number per spike on wheat chromosome arm 7AL by high-resolution genetic mapping. *Theoretical and applied genetics*, *132*(9), 2689-2705.
- Lewis, S., Faricelli, M. E., Appendino, M. L., Valárik, M., & Dubcovsky, J. (2008). The chromosome region including the earliness per se locus Eps-Am1 affects the duration of early developmental phases and spikelet number in diploid wheat. *Journal of experimental botany*, *59*(13), 3595-3607.
- Mangini, G., Gadaleta, A., Colasuonno, P., Marcotuli, I., Signorile, A. M., Simeone, R., et al. (2018). Genetic dissection of the relationships between grain yield components by genome-wide association mapping in a collection of tetraploid wheats. *PloS one*, *13*(1), e0190162-e0190162.

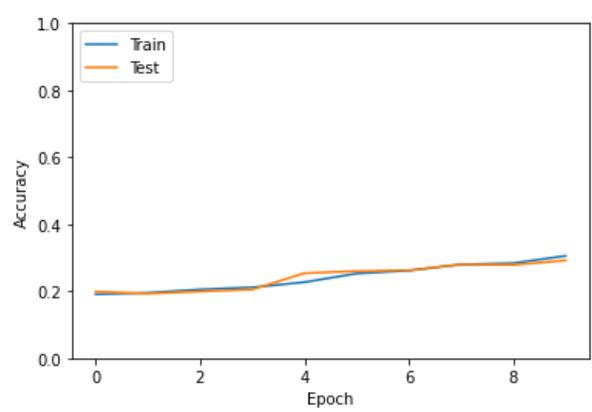
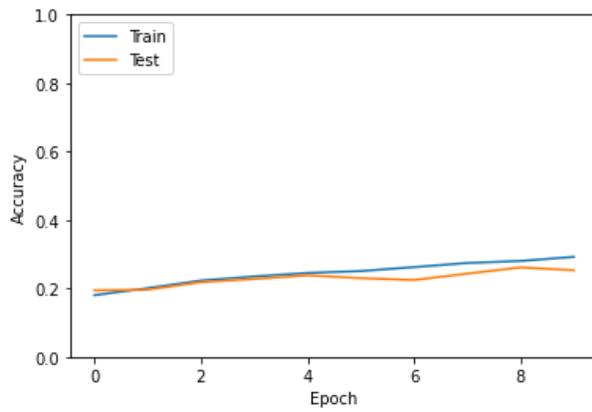
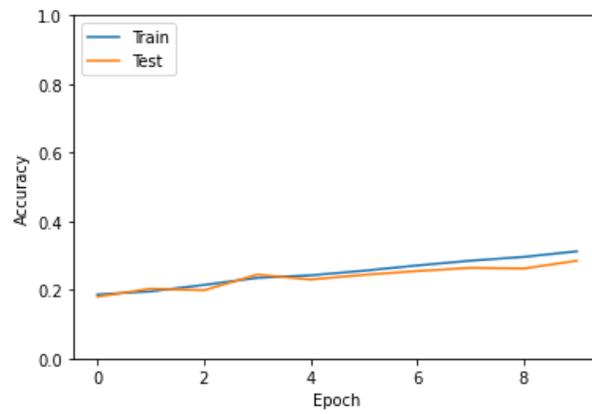
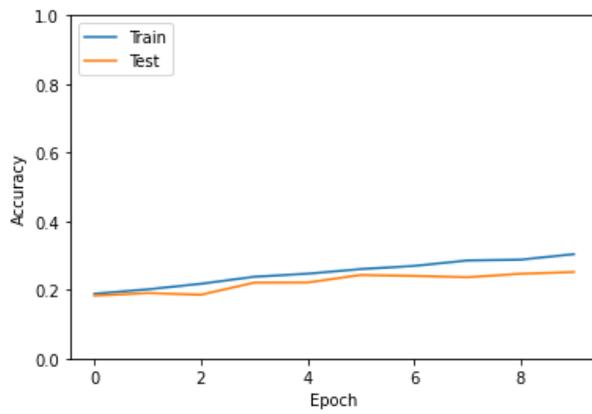
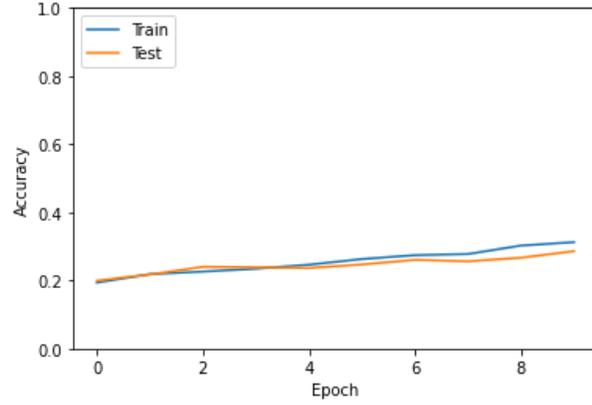
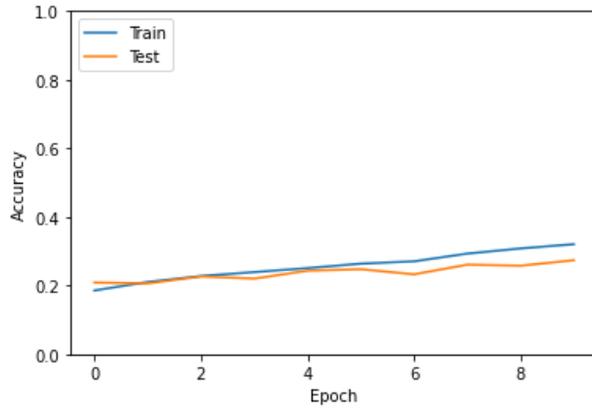
- Miura, H., & Worland, A. (1994). Genetic control of vernalization, day-length response, and earliness per se by homoeologous group-3 chromosomes in wheat. *Plant Breeding*, *113*(2), 160-169.
- Mouradov, A., Cremer, F., & Coupland, G. (2002). Control of Flowering Time: Interacting Pathways as a Basis for Diversity. *The Plant cell*, *14*(suppl 1), S111-S130.
- Rawson, H. M. (1970). Spikelet number, its control and relation to yield per ear in wheat (Vol. 23, pp. 1-15): Australian Journal of Biological Sciences.
- Ren, W.-L., Wen, Y.-J., Dunwell, J. M., & Zhang, Y.-M. (2018). pKWmEB: integration of Kruskal-Wallis test with empirical Bayes under polygenic background control for multi-locus genome-wide association study. *Heredity*, *120*(3), 208-218.
- Sakuma, S., Salomon, B., & Komatsuda, T. (2011). The domestication syndrome genes responsible for the major changes in plant form in the Triticeae crops. *Plant Cell Physiol*, *52*(5), 738-749.
- Scarth, R., & Law, C. N. (1983). The location of the photoperiod gene, Ppd2 and an additional genetic factor for ear-emergence time on chromosome 2B of wheat. *Heredity*, *51*(3), 607-619.
- Schneider, C., Rasband, W., & Eliceiri, K. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nature methods*, *9*(7), 671.
- Snape, J. W., Butterworth, K., Whitechurch, E., & Worland, A. J. (2001). Waiting for fine times: genetics of flowering time in wheat. *Euphytica*, *119*(1), 185-190.
- Stelmakh, A. F. (1998). Genetic systems regulating flowering response in wheat. *Euphytica*, *100*(1), 359-369.

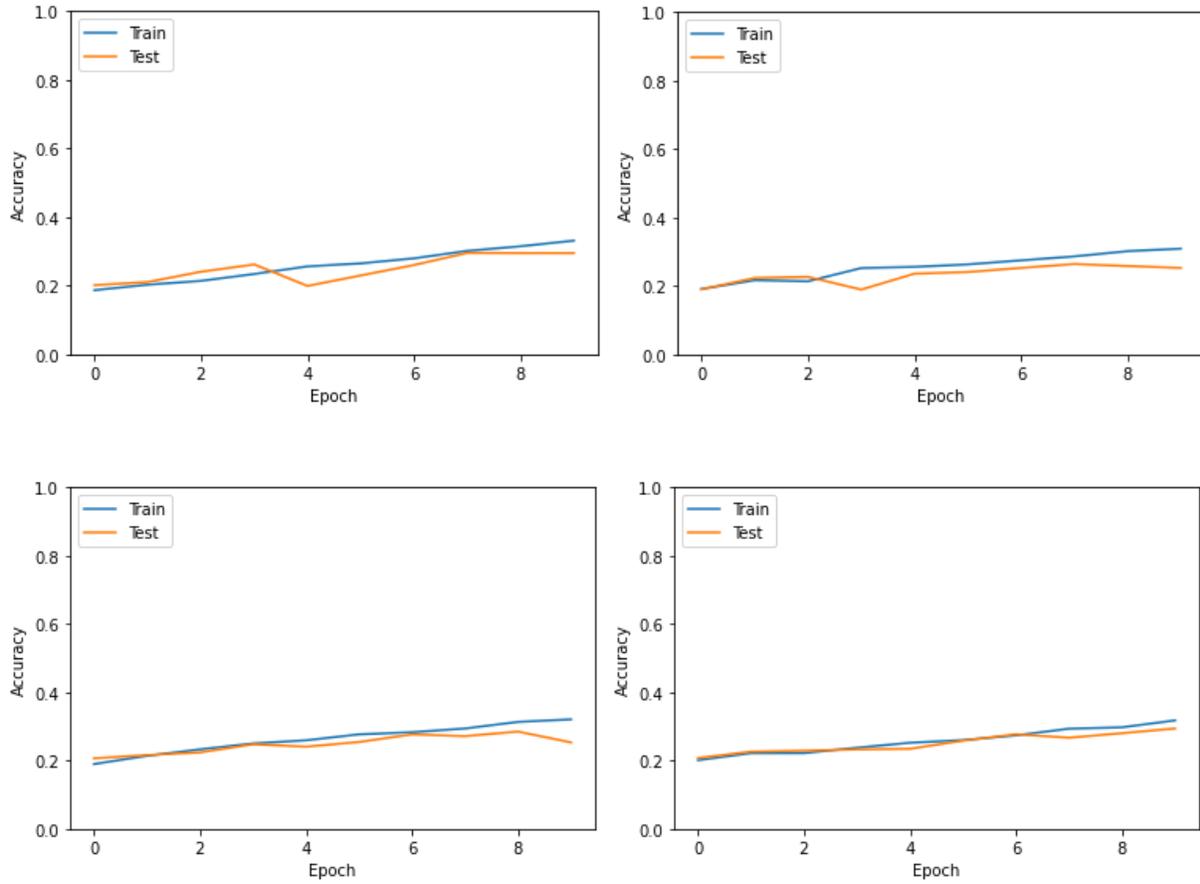
- Tamba, C. L., Ni, Y.-L., & Zhang, Y.-M. (2017). Iterative sure independence screening EM-Bayesian LASSO algorithm for multi-locus genome-wide association studies. *PLoS computational biology*, *13*(1), e1005357-e1005357.
- Tamba, C. L., & Zhang, Y.-M. (2018). A fast mrMLM algorithm for multi-locus genome-wide association studies. *bioRxiv*, 341784.
- Wang, S.-B., Feng, J.-Y., Ren, W.-L., Huang, B., Zhou, L., Wen, Y.-J., et al. (2016). Improving power and accuracy of genome-wide association studies via a multi-locus mixed linear model methodology. *Scientific reports*, *6*(1), 19444-19444.
- Wen, Y.-J., Zhang, H., Ni, Y.-L., Huang, B., Zhang, J., Feng, J.-Y., et al. (2017). Methodological implementation of mixed linear models in multi-locus genome-wide association studies. *Briefings in bioinformatics*, *18*(5), 906-906.
- Winn, Z. J., Larkin, D. L., Murry, J. T., Moon, D. E., & Mason, R. E. (2021). Phenotyping Anther Extrusion of Wheat Using Image Analysis. *Agronomy (Basel)*, *11*(6), 1244.
- Worland, A. J. (1996). The influence of flowering time genes on environmental adaptability in European wheats. *Euphytica*, *89*(1), 49-57.
- Wurchum, T., Leiser, W., Langer, S., Tucker, M., & Longin, C. (2018). Phenotypic and genetic analysis of spike and kernel characteristics in wheat reveals long-term genetic trends of grain yield components. (Vol. 131, pp. 2071-2084): *Theoretical and Applied Genetics*.
- Xu, Y., Wang, R., Tong, Y., Zhao, H., Xie, Q., Liu, D., et al. (2013). Mapping QTLs for yield and nitrogen-related traits in wheat: influence of nitrogen and phosphorus fertilization on QTL expression. *Theoretical and applied genetics*, *127*(1), 59-72.

- Yan, L. (2004). The Wheat VRN2 Gene Is a Flowering Repressor Down-Regulated by Vernalization. *Science (American Association for the Advancement of Science)*, 303(5664), 1640-1644.
- Yan, L., Fu, D., Li, C., Blechl, A., Tranquilli, G., Bonafede, M., et al. (2006). The wheat and barley vernalization gene VRN3 is an orthologue of FT. *Proceedings of the National Academy of Sciences - PNAS*, 103(51), 19581-19586.
- Zhai, H., Feng, Z., Li, J., Liu, X., Xiao, S., Ni, Z., et al. (2016). QTL Analysis of Spike Morphological Traits and Plant Height in Winter Wheat ( *Triticum aestivum* L.) Using a High-Density SNP and SSR-Based Linkage Map. *Frontiers in plant science*, 7, 1617-1617.
- Zhang, J., Feng, J. Y., Ni, Y. L., Wen, Y. J., Niu, Y., Tamba, C. L., et al. (2017). pLARmEB: integration of least angle regression with empirical Bayes for multilocus genome-wide association studies. *Heredity*, 118(6), 517-524.
- Zhang, Y.-W., Tamba, C. L., Wen, Y.-J., Li, P., Ren, W.-L., Ni, Y.-L., et al. (2020). mrMLM v4.0.2: An R Platform for Multi-locus Genome-wide Association Studies. *Genomics, proteomics & bioinformatics*, 18(4), 481-487.
- Zhu, T., Wang, L., Rimbart, H., Rodriguez, J. C., Deal, K. R., De Oliveira, R., et al. (2021). Optical maps refine the bread wheat *Triticum aestivum* cv. Chinese Spring genome assembly. *The Plant journal : for cell and molecular biology*, 107(1), 303-314.

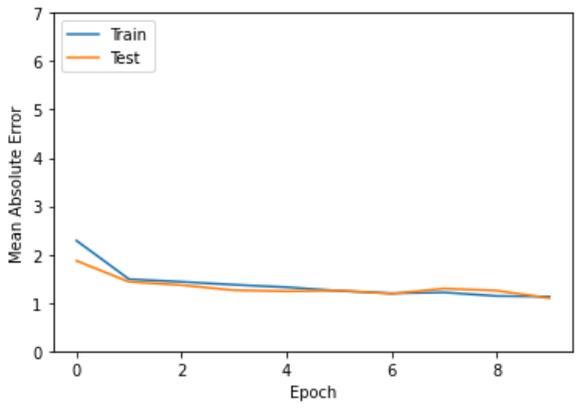
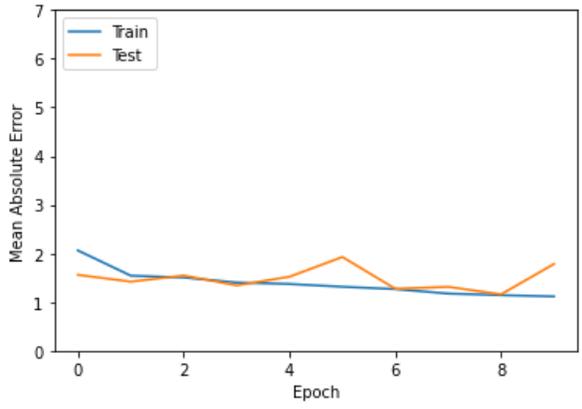
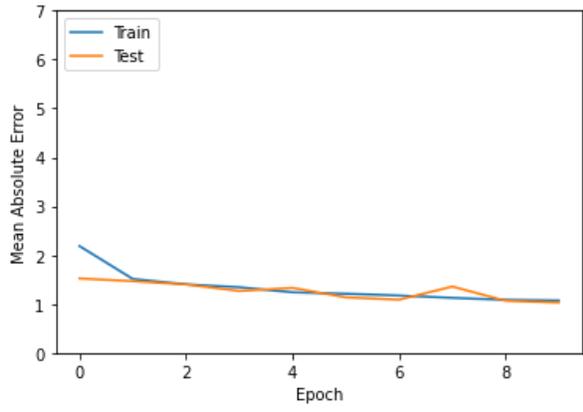
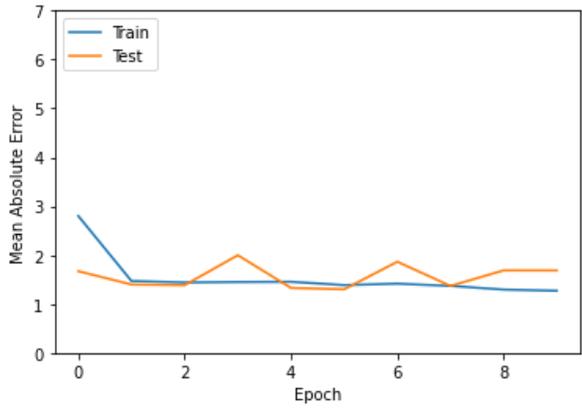
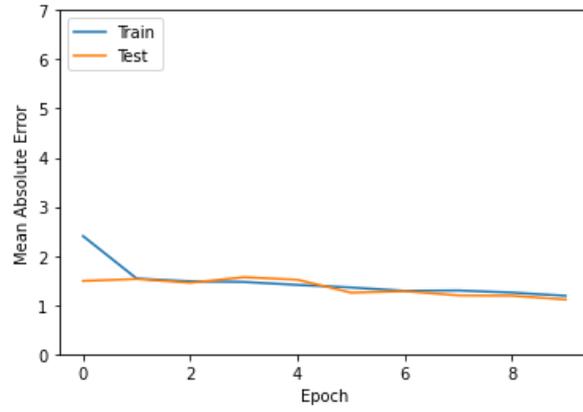
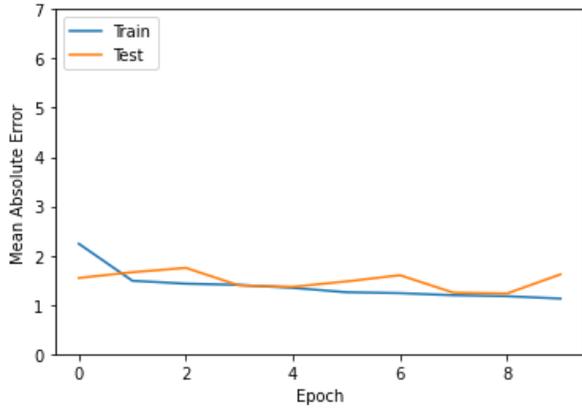
# APPENDIX

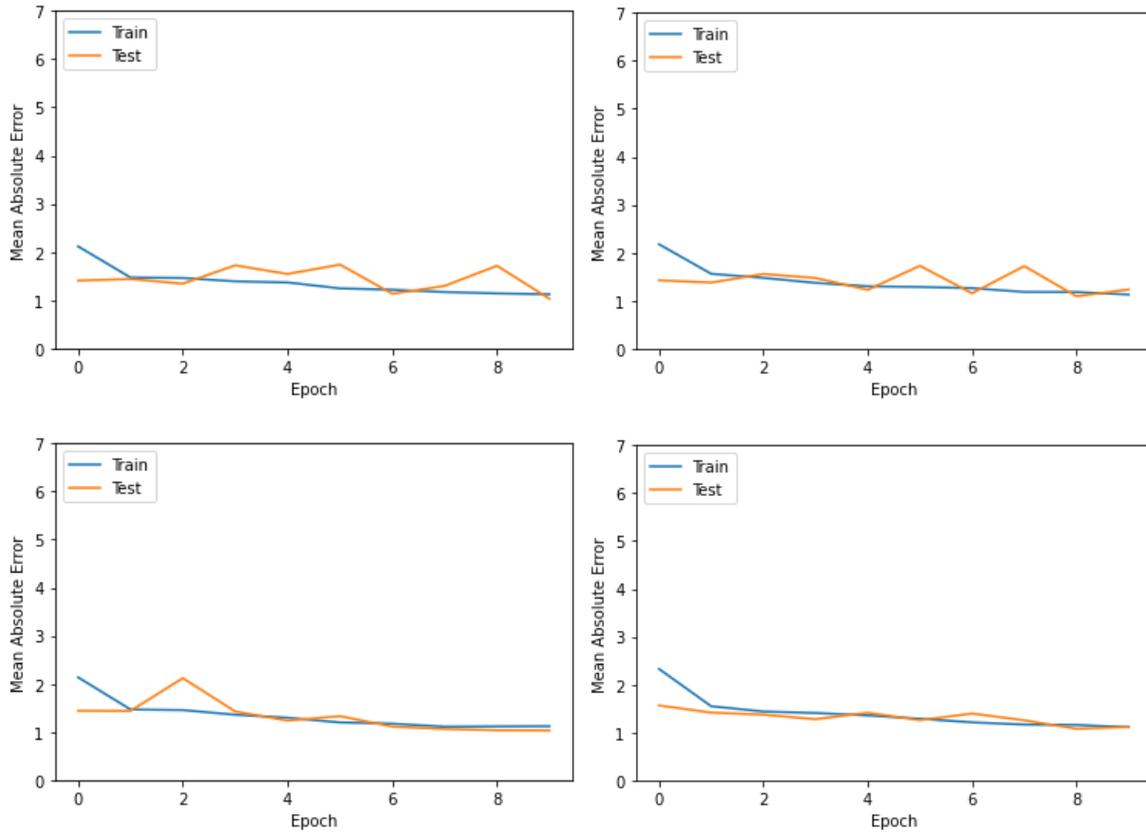
## Chapter II Supplemental



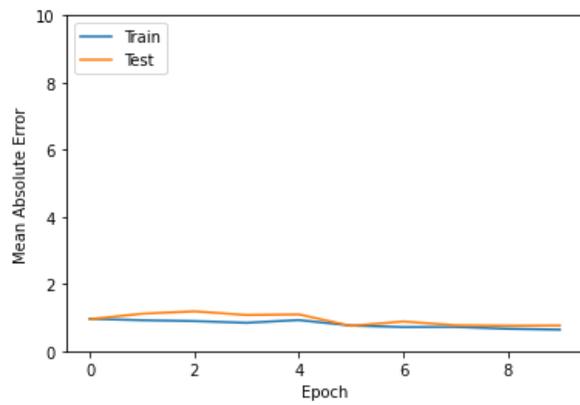
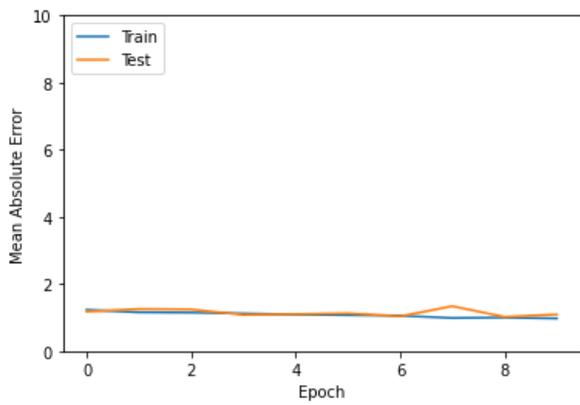
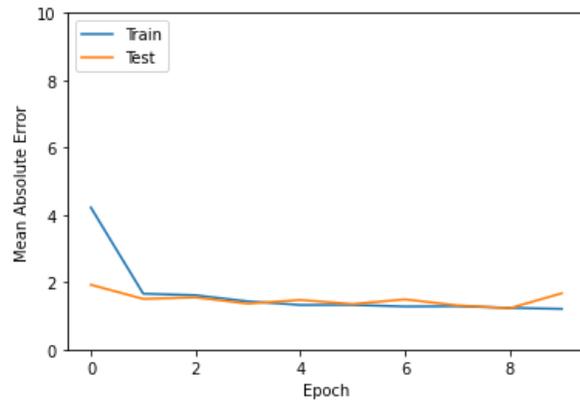
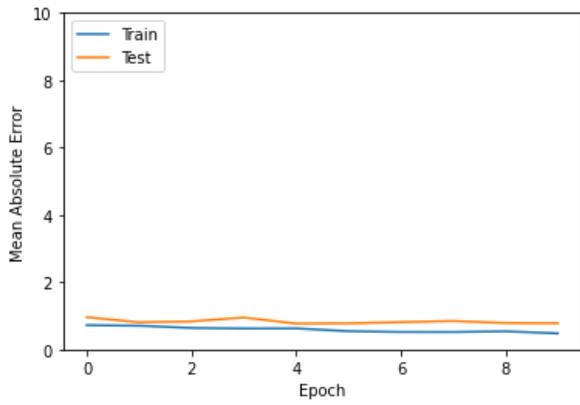
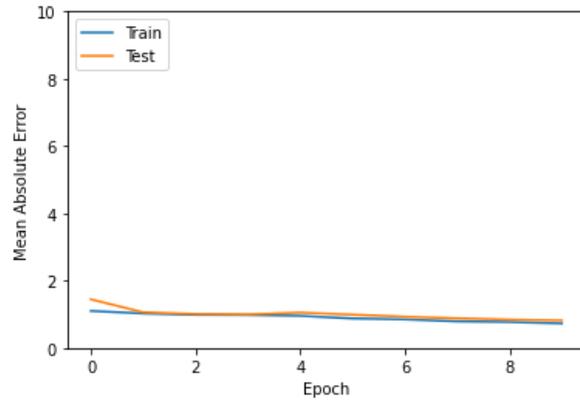
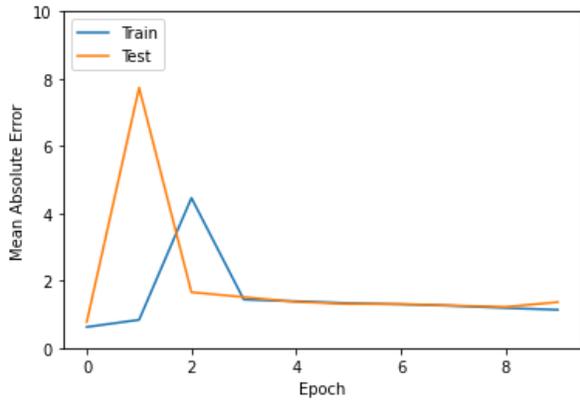
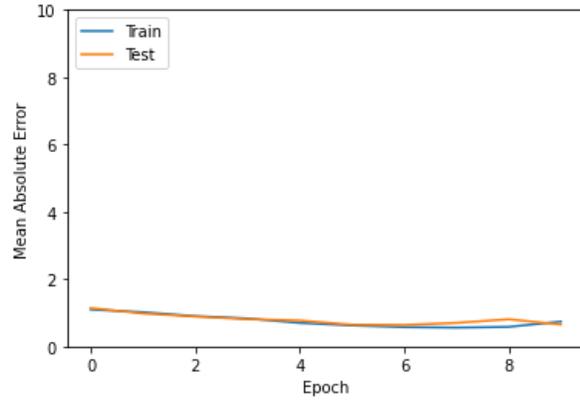
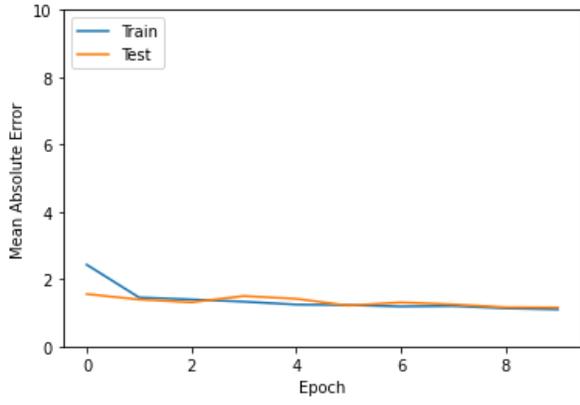


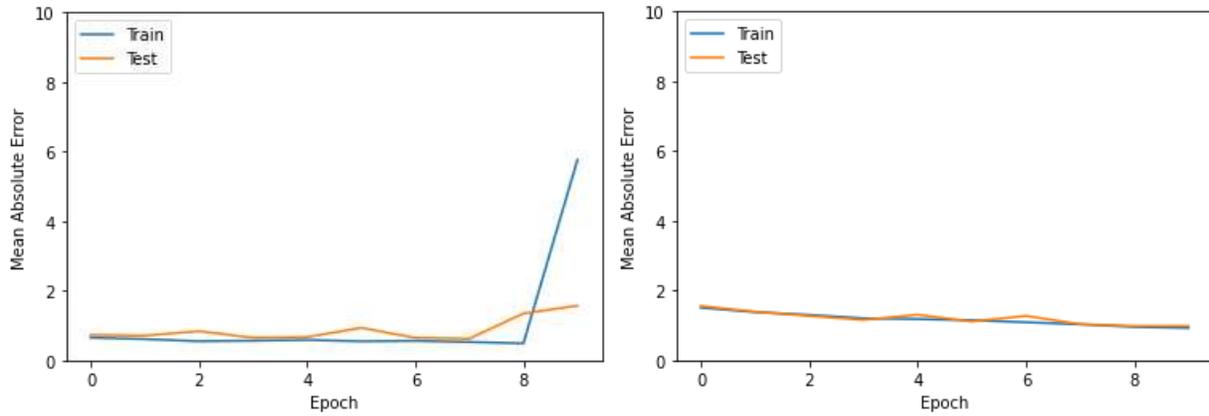
**Supplementary Figure 1.** Accuracy performance across all 10 iterations for 10 epochs of the basic five-layer classification CNN where the blue line is the performance of the training dataset and the orange line shows the performance of the test dataset. Lower accuracies indicate a lower estimation error from the true value.



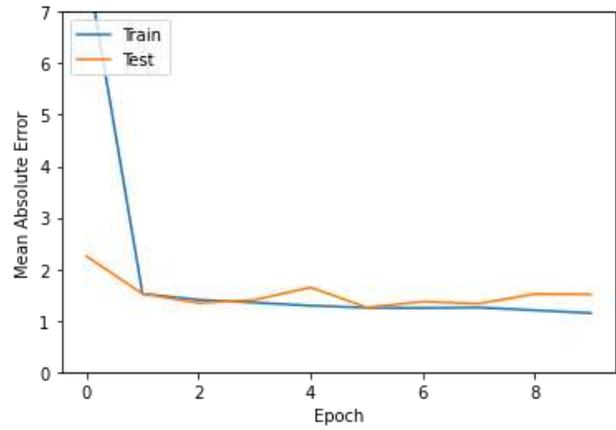
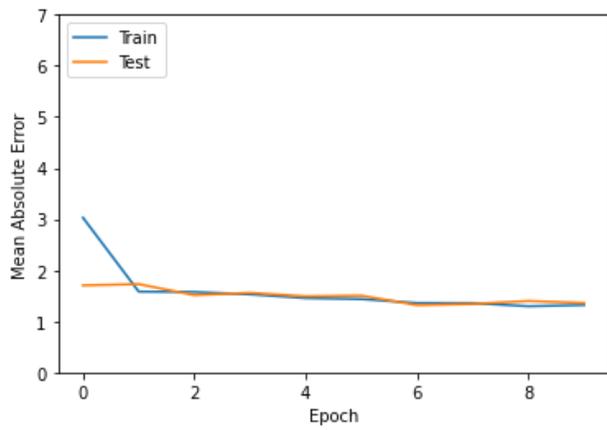
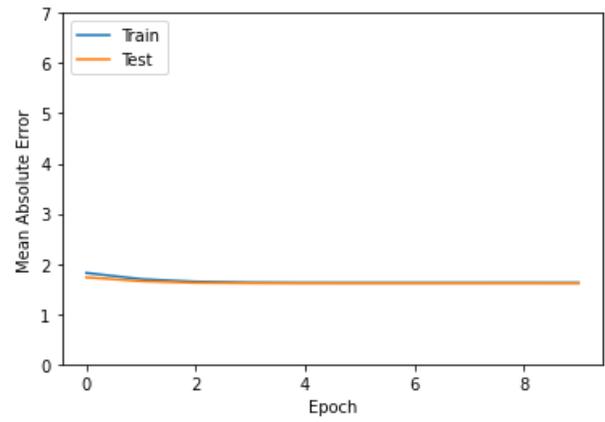
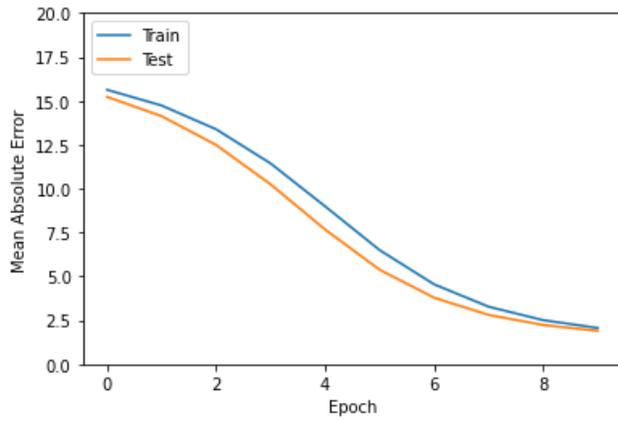
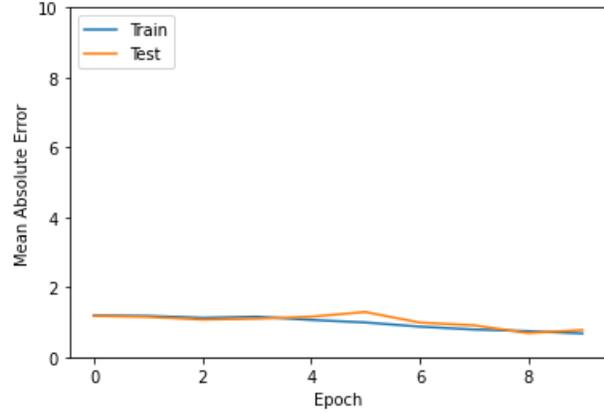
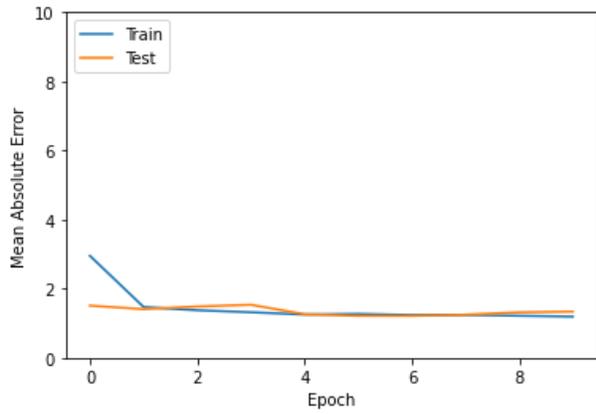


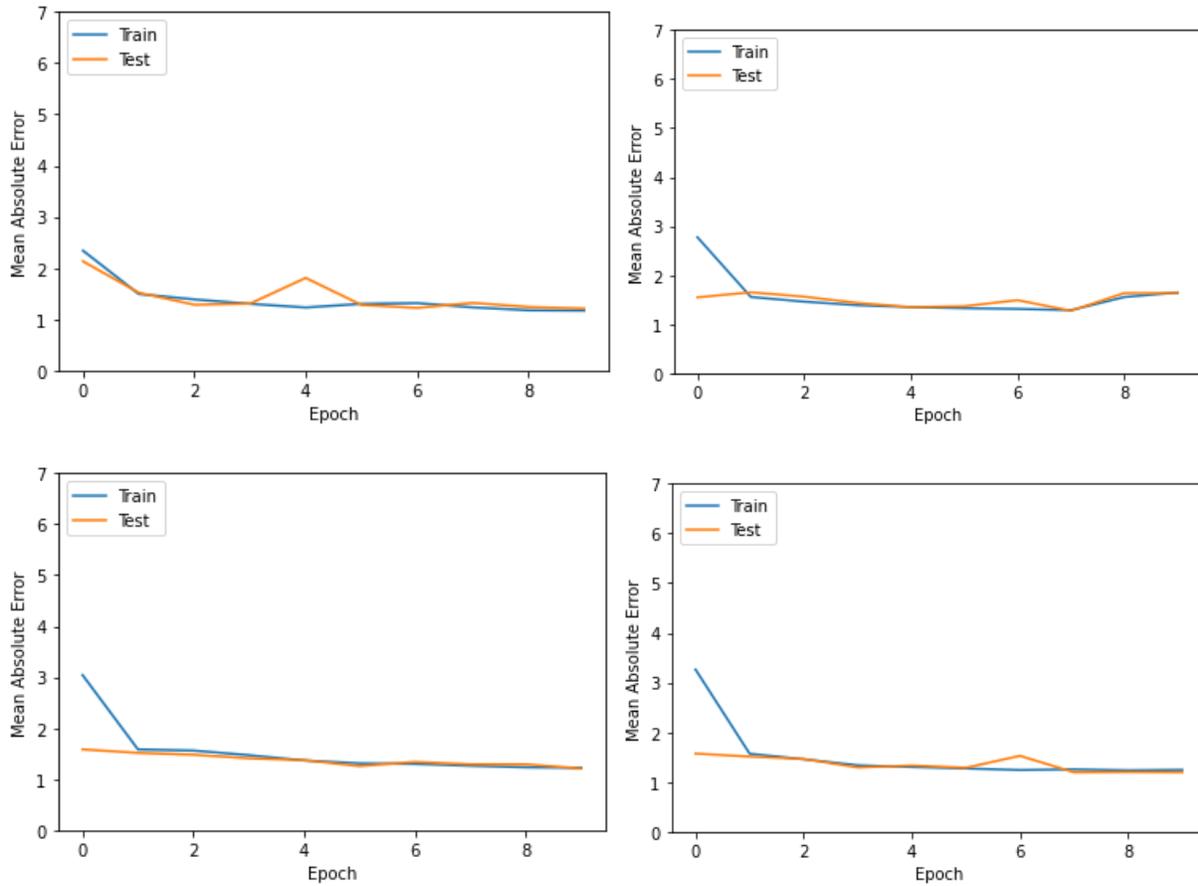
**Supplementary Figure 2.** MAE performance across all 10 iterations for 10 epochs of the basic five-layer regression CNN where the blue line is the performance of the training dataset and the orange line shows the performance of the test dataset. Lower MAEs indicate a lower estimation error from the true value.





**Supplementary Figure 3.** MAE performance across all 10 iterations for 10 epochs of the regression VGG16 pre-trained model where the blue line is the performance of the training dataset and the orange line shows the performance of the test dataset. Lower MAEs indicate a lower estimation error from the true value.





**Supplementary Figure 4.** MAE performance across all 10 iterations for 10 epochs of the regression VGG19 pre-trained model where the blue line is the performance of the training dataset and the orange line shows the performance of the test dataset. Lower MAEs indicate a lower estimation error from the true value.

### Chapter III Supplemental

**Supplemental Table 1.** MrMLM results for all six models for 2019, 2020, and combined analysis.

Trait	Method	RS#	Chr.	Position	QTN effect	LOD score	r2 (%)
<u>2019</u>							
HD	mrMLM	S1B_62023056	1B	62023056	-0.5999	4.9811	6.2503
		S1B_242040871	1B	242040871	0.2558	4.0442	2.0554
		S7A_675198728	7A	675198728	0.293	4.3895	2.1549
	FASTmrMLM	S6A_567013715	6A	567013715	-0.2517	4.1522	1.91
		S6B_497933658	6B	497933658	0.5917	4.7815	3.9831
	pLARmEB	S6D_9137410	6D	9137410	0.644	6.0808	2.2858
TKW	mrMLM	S5A_38118471	5A	38118471	-0.7461	5.1006	2.9353
		S7A_3322031	7A	3322031	-1.7441	4.1028	6.9544
	FASTmrEMMA	S5A_41964608	5A	41964608	-1.4187	4.6255	2.5812
		S5A_698528417	5A	698528417	-1.7742	5.3768	3.6293
	pLARmEB	S5A_698528417	5A	698528417	-0.7821	4.735	2.8643
	pKWmEB	S3B_249632650	3B	249632650	0.9115	5.5598	2.2991
		S3B_115807544	3B	115807544	-1.1107	4.216	2.3243
		S4A_194433192	4A	194433192	0.5334	4.2096	2.258
	ISIS EM-BLASSO	S7B_14167976	7B	14167976	-1.2652	4.7648	3.2796
		S4B_657290599	4B	657290599	0.6291	4.1385	1.8995
KNS	mrMLM	S5A_698528417	5A	698528417	-0.826	4.8652	3.1952
		S2D_55057832	2D	55057832	3.4813	4.5764	5.5962
	pLARmEB	S1A_335036539	1A	335036539	1.853	4.0422	2.0961
pKWmEB	S1B_53645659	1B	53645659	-3.0624	5.581	3.5863	

		S1A_335036539	1A	335036539	1.5911	4.5077	2.9358
	ISIS EM- BLASSO	S1B_53645659	1B	53645659	-2.8466	4.9941	3.1148
SPS	mrMLM	S2B_403704287	2B	403704287	0.3214	4.7337	3.4241
		S5B_590471022	5B	590471022	-0.3162	4.4739	3.7282
		S7A_672854561	7A	672854561	-0.455	7.7904	6.4215
	FASTmrMLM	S7A_672854561	7A	672854561	-0.4291	8.3764	5.7184
	FASTmrEMMA	S1A_531926715	1A	531926715	0.5229	4.483	2.1076
		S2B_403704287	2B	403704287	0.5183	4.0659	1.9143
		S6D_464495312	6D	464495312	-0.4964	5.1558	2.2607
		S7A_672854561	7A	672854561	-0.8078	6.9994	5.0206
	pLARmEB	S1A_531721876	1A	531721876	0.2558	4.6836	1.9815
		S2B_403704287	2B	403704287	0.2958	6.1403	2.8651
		S4A_722318181	4A	722318181	-0.2867	4.9304	2.4995
		S5B_590471022	5B	590471022	-0.2754	5.4851	2.7939
		S7A_672854561	7A	672854561	-0.4528	11.5502	6.2831
	pKWmEB	S1D_58005554	1D	58005554	0.3222	5.7392	3.0413
		S1A_531721876	1A	531721876	0.2553	6.0553	2.5136
		S2B_403704287	2B	403704287	0.2432	4.5255	3.3019
		S7A_672854561	7A	672854561	-0.4523	12.1018	6.742
	ISIS EM- BLASSO	S2B_403704273	2B	403704273	0.2616	4.3352	2.2579
		S6D_464495312	6D	464495312	-0.2438	4.9205	2.1812
		S7A_673030388	7A	673030388	-0.4208	9.399	5.6197
W	pLARmEB	S7A_696981995	7A	696981995	-0.2022	4.527	2.6518
	pKWmEB	S3B_829382536	3B	829382536	-0.1669	4.5216	2.4518

		S7A_696981995	7A	696981995	-0.2053	5.5167	4.3557
		S7A_51647371	7A	51647371	-0.1146	4.2112	1.6267
	ISIS EM- BLASSO	S2A_697000089	2A	697000089	-0.2825	5.1383	4.1738
		S6B_675985723	6B	675985723	0.1518	4.0805	1.6893
		S7A_696981995	7A	696981995	-0.1961	4.2207	2.4942
L	mrMLM	S2B_403704287	2B	403704287	1.979	4.4784	2.8316
		S2B_655107436	2B	655107436	1.7289	4.1767	2.4546
		S4A_733713873	4A	733713873	-2.3384	4.844	4.5088
		S4B_619080071	4B	619080071	-1.8975	5.2939	2.9904
		S5B_20131242	5B	20131242	-2.103	4.436	2.5879
		S7A_681913871	7A	681913871	2.01	4.5466	3.0406
		S7A_710959180	7A	710959180	-2.2521	4.3923	2.7494
	FASTmrMLM	S4A_733713873	4A	733713873	-1.8627	4.7584	2.921
	pLARmEB	S4A_733713873	4A	733713873	-1.8796	5.0939	2.9742
	pKWmEB	S5B_433023852	5B	433023852	-1.9149	4.8848	2.3481
		S7A_16975061	7A	16975061	1.5023	4.1663	2.8909
		S7B_459605010	7B	459605010	2.5726	5.3457	3.2813
	ISIS EM- BLASSO	S3A_118997578	3A	118997578	1.7419	4.0343	1.0809
		S5B_433023852	5B	433023852	-1.9087	4.3811	1.8845
		S7B_459605010	7B	459605010	2.3006	4.2884	2.0276
SA	mrMLM	S5A_698528417	5A	698528417	-145.8025	29.6291	21.748
		S7A_644620616	7A	644620616	-42.9281	5.917	2.1144
	FASTmrMLM	S5A_698528417	5A	698528417	-141.1526	27.0107	20.3829
		S5B_6538709	5B	6538709	-50.5694	4.188	1.6683

		S7A_644620616	7A	644620616	-44.128	5.6553	2.2343
FASTmrEMMA		S5A_698528417	5A	698528417	-292.045	26.9098	21.4833
pLARmEB		S5A_698528417	5A	698528417	-151.7601	33.0397	23.5615
		S7A_644620616	7A	644620616	-37.8072	4.4733	1.6401
pKWmEB		S4A_631950295	4A	631950295	-37.7737	4.2174	3.0676
		S5A_698528417	5A	698528417	-112.755	26.3209	24.7161
		S5B_419518823	5B	419518823	-35.2717	4.2263	1.2937
		S7A_644620616	7A	644620616	-40.2479	5.1144	2.4666
ISIS EM- BLASSO		S5A_698528417	5A	698528417	-146.0814	31.5905	21.8312
		S7A_616750641	7A	616750641	57.9699	4.3581	2.0286
		S7A_644620616	7A	644620616	-48.318	6.7193	2.6787

---

2020

HD	mrMLM	S3B_610036538	3B	610036538	1.2657	5.6884	5.2968
		S5A_585018041	5A	585018041	-1.154	6.1012	4.3979
		S5B_531530727	5B	531530727	0.979	4.992	3.1597
		S7B_486028824	7B	486028824	1.236	5.9785	4.3462
	FASTmrMLM	S3B_610036538	3B	610036538	0.9245	4.0431	2.826
		S5A_585018041	5A	585018041	-0.8756	4.284	2.5322
	FASTmrEMMA	S1B_563444512	1B	563444512	-1.7785	4.0043	1.7564
		S3B_610036538	3B	610036538	2.149	5.0819	3.4937
		S5A_585018041	5A	585018041	-1.8442	4.702	2.6656
		S5B_531530727	5B	531530727	1.6529	4.307	2.1215
	pLARmEB	S2D_29715360	2D	29715360	-0.765	4.7041	1.6197
		S3B_610036538	3B	610036538	1.1129	6.3932	3.4103

		S3D_596661381	3D	596661381	0.7155	4.5629	1.245
		S5A_585018041	5A	585018041	-1.0936	6.5929	3.2895
		S6D_15344114	6D	15344114	1.6866	4.2289	1.9268
		S7B_2327298	7B	2327298	-0.6844	4.3991	1.2411
		S7B_486028824	7B	486028824	1.1765	7.1098	3.2794
	pKWmEB	S3B_610036538	3B	610036538	0.7514	4.6138	4.605
		S5A_577963056	5A	577963056	1.2668	5.8836	5.7357
		S6D_15344114	6D	15344114	2.201	7.0225	5.709
		S6A_13672088	6A	13672088	0.9448	4.3535	2.9192
	ISIS EM- BLASSO	S2A_38307973	2A	38307973	0.8002	4.8534	2.1266
		S3B_820145958	3B	820145958	-1.0199	4.7112	1.3567
		S3D_596661381	3D	596661381	0.7505	4.5841	1.6448
		S5A_585018041	5A	585018041	-1.2109	9.1722	4.8423
		S5A_693604905	5A	693604905	-1.3413	4.3153	1.2824
		S5B_531530727	5B	531530727	0.7529	4.552	1.8685
		S6A_13672088	6A	13672088	0.9166	4.0057	1.6797
		S7A_82951751	7A	82951751	0.7254	4.2622	1.0895
		S7B_486028824	7B	486028824	1.1137	6.3719	3.5291
TKW	mrMLM	S2B_527875976	2B	527875976	-4.2051	6.5459	45.2997
		S2A_739309428	2A	739309428	0.9275	4.3202	1.9182
		S5A_48763086	5A	48763086	-1.0823	4.2633	1.4872
		S6B_680795066	6B	680795066	1.2299	5.458	1.6489
	FASTmrMLM	S6B_680795066	6B	680795066	1.1881	5.3914	2.0603
	FASTmrEMMA	S3B_52721463	3B	52721463	1.716	5.2346	2.7052

		S4A_639027818	4A	639027818	1.8074	4.0281	1.5417
	pLARmEB	S2A_733652274	2A	733652274	-0.5407	4.0853	1.0817
		S2A_739309428	2A	739309428	0.7944	4.558	1.8837
		S3B_52721463	3B	52721463	0.8013	5.0941	2.5744
		S4A_639027818	4A	639027818	1.0873	6.4054	2.4906
		S6B_102070407	6B	102070407	-0.7654	4.7997	1.8223
		S6B_662504589	6B	662504589	-1.2999	4.8669	2.6898
		S7A_76114512	7A	76114512	-0.6449	4.0908	1.4398
	pKWmEB	S2A_759456102	2A	759456102	-0.5276	4.3101	2.1319
		S3B_52721463	3B	52721463	0.7915	5.3538	4.5207
		S4A_639027818	4A	639027818	0.9596	5.428	3.542
		S4B_650666864	4B	650666864	0.4909	4.7003	2.632
		S5A_546396965	5A	546396965	-0.5078	4.4008	1.3913
		S6B_82524150	6B	82524150	1.0401	4.387	4.159
		S6B_198500379	6B	198500379	-0.6213	5.1158	3.0067
	ISIS EM- BLASSO	S2A_739309428	2A	739309428	0.7121	4.1595	1.5137
		S3A_303372474	3A	303372474	-0.5236	4.2562	1.0025
		S5A_698528417	5A	698528417	-0.7774	4.7981	2.1832
		S6B_532043730	6B	532043730	0.8309	6.2902	2.4333
		S6B_680795066	6B	680795066	1.2682	6.8103	2.3471
KNS	mrMLM	S1A_293734546	1A	293734546	-3.3099	5.8999	6.28
		S1D_420960639	1D	420960639	1.5375	4.458	3.4465
		S3A_187004480	3A	187004480	2.2659	5.0378	3.2131
		S3B_820862429	3B	820862429	2.1093	5.2923	2.6913

	S5A_692194752	5A	692194752	2.1497	5.1458	3.3164
	S5A_51560475	5A	51560475	1.9073	4.6835	2.5373
	S6D_4032694	6D	4032694	1.6488	6.2038	3.6727
	S7A_12454707	7A	12454707	-1.3864	5.6129	3.003
	S7D_73368696	7D	73368696	-2.3328	6.395	3.7409
FASTmrMLM	S1D_420960639	1D	420960639	1.1228	4.7245	1.8379
	S3A_187004480	3A	187004480	2.1514	6.324	2.8964
	S5B_559044987	5B	559044987	1.202	4.6714	2.2274
	S6A_499197587	6A	499197587	-1.218	4.6999	2.753
	S6A_609285176	6A	609285176	1.0197	4.544	1.7465
	S6A_614663595	6A	614663595	1.1635	5.0279	2.0358
	S6D_354074	6D	354074	-1.55	5.7543	2.6406
	S7D_73368696	7D	73368696	-1.7408	4.9567	2.0831
pLARmEB	S1D_428666044	1D	428666044	-1.5889	4.6205	1.7261
	S3A_187004480	3A	187004480	2.0383	5.6668	2.1801
	S5B_559044987	5B	559044987	1.1764	4.5379	1.789
	S6A_499197587	6A	499197587	-1.331	6.2124	2.7566
	S6A_610127009	6A	610127009	1.2459	5.7029	2.3938
	S6A_614663910	6A	614663910	1.4562	4.5061	1.4668
pKWmEB	S2D_591603483	2D	591603483	0.8207	4.2219	1.6588
	S5A_692194752	5A	692194752	1.6376	5.0066	3.653
	S6A_499197587	6A	499197587	-1.112	5.7268	5.1321
	S6A_580586410	6A	580586410	1.1732	5.6962	4.0344
	S6D_354074	6D	354074	-1.4502	5.562	3.3529
	S6A_609285176	6A	609285176	1.0174	4.5011	4.0916

		S6A_614523454	6A	614523454	1.3499	4.7935	3.0708
		S7B_14167976	7B	14167976	2.1123	4.5619	3.2382
ISIS EM-							
BLASSO		S1A_293734546	1A	293734546	-2.7976	6.34	4.4866
		S1D_420960639	1D	420960639	1.0486	4.5079	1.603
		S2B_769071732	2B	769071732	1.5785	5.5065	2.219
		S2D_399428380	2D	399428380	1.2774	7.0707	1.859
		S3A_70711463	3A	70711463	-1.5628	5.2213	2.1752
		S3A_187004480	3A	187004480	1.9382	5.9654	2.3509
		S3B_562731470	3B	562731470	1.3751	6.2545	2.2283
		S3B_578814680	3B	578814680	-1.0495	4.8333	1.8044
		S3B_820862429	3B	820862429	1.5844	4.4684	1.5184
		S5A_51560475	5A	51560475	1.3117	4.7051	1.2001
		S5A_692194752	5A	692194752	1.3307	4.1585	1.2707
		S5B_244411156	5B	244411156	-0.9989	5.6223	1.8489
		S6A_499197587	6A	499197587	-1.4407	8.3435	3.8515
		S6A_614663910	6A	614663910	1.0954	4.0389	0.9899
		S6B_118986455	6B	118986455	1.2938	4.0941	0.9239
		S6D_354074	6D	354074	-1.4724	5.5945	2.3827
SPS	mrMLM	S3A_642554084	3A	642554084	-0.1998	4.0483	2.0937
		S7B_11292130	7B	11292130	-0.3209	4.0414	2.8591
		S7A_672148737	7A	672148737	-0.3252	4.3991	4.9229
	FASTmrMLM	S7A_672148737	7A	672148737	-0.2696	5.0492	3.383
	FASTmrEMMA	S7A_672148737	7A	672148737	-0.5228	4.4937	2.9645
	pKWmEB	S7A_672148737	7A	672148737	-0.2327	4.4889	5.566

	ISIS EM- BLASSO	S7A_672045448	7A	672045448	-0.2298	4.1605	2.4571
		S7B_11292130	7B	11292130	-0.3082	5.0252	2.6369
W	mrMLM	S1A_535482663	1A	535482663	0.1849	5.6117	2.814
		S2B_107791597	2B	107791597	0.6238	4.0768	33.5524
		S3D_524111395	3D	524111395	0.2509	5.7894	1.7516
		S5A_621281176	5A	621281176	0.3907	5.906	5.8619
		S5B_581759220	5B	581759220	0.1757	5.1998	2.6632
		S5B_8087033	5B	8087033	0.1905	4.3308	1.5936
	FASTmrMLM	S1A_535482663	1A	535482663	0.1335	4.3686	2.1616
		S3D_524111395	3D	524111395	0.1778	4.1311	1.2959
		S5A_621281176	5A	621281176	0.3474	5.6603	6.8304
		S5B_8087033	5B	8087033	0.1885	5.1592	2.2992
	FASTmrEMMA	S5B_572818242	5B	572818242	0.244	4.0409	2.2289
	pLARmEB	S1A_535482663	1A	535482663	0.1274	4.2121	1.5755
		S5A_621281176	5A	621281176	0.3138	4.8701	4.4621
	pKWmEB	S1A_535482663	1A	535482663	0.1191	4.8379	2.8288
		S5A_621281176	5A	621281176	0.3195	5.4774	8.7047
		S5B_8087033	5B	8087033	0.1307	4.0092	2.4998
		S6A_337042184	6A	337042184	0.1261	4.0925	2.9341
	ISIS EM- BLASSO	S1A_535482663	1A	535482663	0.1398	4.7176	2.3707
		S3D_524111395	3D	524111395	0.1801	4.2835	1.3298
		S5A_621281176	5A	621281176	0.3619	7.2939	7.4112
		S6A_12344581	6A	12344581	-0.1596	4.8662	2.7016
L	mrMLM	S2A_708760295	2A	708760295	3.7123	4.4185	6.6197

		S5B_572342929	5B	572342929	1.6995	5.0338	3.2914
		S5A_577963056	5A	577963056	-2.5704	4.6216	4.144
		S6B_670729335	6B	670729335	-2.8647	5.0136	3.1493
		S6A_271232270	6A	271232270	2.7042	4.8281	3.4669
FASTmrMLM		S2A_708760295	2A	708760295	2.7848	4.3048	3.7251
		S6B_332505436	6B	332505436	1.9096	7.1988	4.2489
		S6B_670729335	6B	670729335	-2.244	4.4541	1.9324
		S7B_646206252	7B	646206252	2.1419	5.2081	4.2525
FASTmrEMMA		S6B_670729335	6B	670729335	-4.6376	4.1831	2.0634
pLARmEB		S2A_708760295	2A	708760295	2.811	4.4881	3.7535
		S2B_5017296	2B	5017296	-2.3691	5.7409	2.7006
		S5B_584851722	5B	584851722	-1.781	5.6741	3.3476
		S6B_332505436	6B	332505436	1.8814	6.9981	4.0787
		S7B_646206252	7B	646206252	1.9852	4.3579	3.6127
pKWmEB		S2A_744704521	2A	744704521	-1.1498	4.446	2.5124
		S3B_407018291	3B	407018291	-1.3243	5.0934	2.1181
		S6B_454837017	6B	454837017	1.3592	5.2157	4.7747
		S7B_482072315	7B	482072315	-1.7097	4.2495	3.6343
ISIS EM- BLASSO		S2A_708760295	2A	708760295	2.5813	4.0028	3.2004
		S2B_5017296	2B	5017296	-1.9319	4.4632	1.8159
		S3B_533976494	3B	533976494	1.4585	5.1494	2.0026
		S6B_226747335	6B	226747335	1.4059	4.8691	2.2041
		S6B_670729335	6B	670729335	-2.3762	5.1777	2.1669
SA	mrMLM	S2B_11178813	2B	11178813	-88.1999	4.1941	2.7909

		S5A_698528417	5A	698528417	-157.6733	32.0164	24.7734
		S6A_429672320	6A	429672320	-44.4955	4.4776	2.2113
FASTmrMLM		S5A_698528417	5A	698528417	-146.5784	28.9123	21.4097
		S6A_429672320	6A	429672320	-41.7427	4.8355	1.9461
		S6B_351590979	6B	351590979	-48.0378	5.1842	2.441
FASTmrEMMA		S5A_698528417	5A	698528417	-285.2209	24.0239	19.9601
		S6A_429672320	6A	429672320	-79.369	4.2386	1.7488
pLARmEB		S5A_698528417	5A	698528417	-148.3453	31.7035	21.9289
		S5B_390903560	5B	390903560	45.8989	4.0595	1.28
		S6A_429672320	6A	429672320	-40.6648	4.9004	1.8469
		S6B_158085055	6B	158085055	-40.5402	4.4033	1.7444
		S7A_634332221	7A	634332221	39.7617	4.5225	1.4244
pKWmEB		S4A_7176912	4A	7176912	36.0962	4.3402	2.2782
		S5A_698528417	5A	698528417	-116.3646	26.1496	27.789
ISIS EM- BLASSO		S5A_698528417	5A	698528417	-142.7534	25.8412	20.3069

---

Combined analysis

HD	mrMLM	S1B_580449038	1B	580449038	0.994	5.3386	0.2911
		S2B_717098094	2B	717098094	4.1229	6.5765	51.6939
		S6B_497933658	6B	497933658	1.3893	7.8669	2.4099
		S6A_305876358	6A	305876358	-0.6308	5.3118	0.8653
		S7B_596398733	7B	596398733	-2.8699	10.3944	23.3079
FASTmrMLM		S2B_31382110	2B	31382110	-0.3702	4.6182	1.7641
		S6B_497933658	6B	497933658	1.381	9.0144	8.4792
pLARmEB		S1A_124476146	1A	124476146	0.5366	5.078	1.8432

		S1B_580449038	1B	580449038	0.8893	4.9865	0.4887
		S2A_738040009	2A	738040009	0.9781	5.3699	1.9178
		S4B_219107827	4B	219107827	0.4136	4.6024	1.2648
		S5B_422697466	5B	422697466	-0.5869	4.1661	1.2503
		S5D_549874452	5D	549874452	0.4069	4.488	1.1505
		S6B_497933658	6B	497933658	1.4045	9.7122	5.1676
		S7B_83669403	7B	83669403	0.6847	4.2906	1.1679
	pKWmEB	S1A_11066883	1A	11066883	-0.3691	4.3723	2.045
		S2A_738040009	2A	738040009	0.7892	5.8355	3.667
		S3A_9082226	3A	9082226	0.4205	4.3958	2.0548
		S6B_497933658	6B	497933658	1.203	7.3503	10.2859
	ISIS EM- BLASSO	S1B_580449038	1B	580449038	0.8934	4.6925	0.8189
		S2A_738040009	2A	738040009	0.9393	5.5175	2.9359
		S3A_9082226	3A	9082226	0.5016	4.3964	1.849
		S5D_549874452	5D	549874452	0.5135	6.8105	3.0413
		S6B_497933658	6B	497933658	1.429	10.5641	8.8805
		S7A_116544032	7A	116544032	0.7488	5.4599	2.4387
		S7B_83669403	7B	83669403	0.7621	4.6159	2.4014
TKW	mrMLM	S5A_698528417	5A	698528417	-1.0413	5.6996	4.8436
		S5A_38118471	5A	38118471	-0.9612	5.282	4.6153
	FASTmrMLM	S5A_698528417	5A	698528417	-0.7561	4.1805	2.5539
	pLARmEB	S5A_698528417	5A	698528417	-0.7381	4.3351	2.4335
	pKWmEB	S4A_194433192	4A	194433192	0.5997	4.207	3.0019
		S5A_698528417	5A	698528417	-0.7247	5.5298	5.5402

		S5A_38118471	5A	38118471	-0.6439	4.5059	4.1348
	ISIS EM- BLASSO	S5A_38118471	5A	38118471	-0.8366	6.6258	3.4965
		S5A_698528417	5A	698528417	-0.911	5.892	3.7067
		S7B_14167976	7B	14167976	-1.2796	4.0168	1.8924
KNS	pKWmEB	S1A_335036539	1A	335036539	1.4849	4.2329	3.1173
		S1B_93946603	1B	93946603	1.1839	4.0514	2.3498
	ISIS EM- BLASSO	S1B_93946603	1B	93946603	1.2784	4.1294	1.7089
		S5A_596027539	5A	596027539	1.9547	4.6195	2.6064
SPS	mrMLM	S1A_531721876	1A	531721876	0.3375	4.406	3.3537
		S2B_403704287	2B	403704287	0.3161	4.2838	3.1912
		S5B_590471022	5B	590471022	-0.3322	4.7057	3.9814
		S6D_464495312	6D	464495312	-0.3048	5.0121	3.3027
		S7A_672854561	7A	672854561	-0.4507	7.8421	6.0929
	FASTmrMLM	S5B_590471022	5B	590471022	-0.2618	4.2993	2.5196
		S6D_464495312	6D	464495312	-0.2259	4.2508	1.8471
		S7A_672854561	7A	672854561	-0.4092	8.2858	5.1176
	FASTmrEMMA	S6D_464495312	6D	464495312	-0.5284	5.0936	2.5275
		S7A_672854561	7A	672854561	-0.8126	6.9802	5.0001
	pLARmEB	S1A_531721876	1A	531721876	0.2913	5.2363	2.5306
		S2B_403704287	2B	403704287	0.2739	4.808	2.4265
		S4A_722318181	4A	722318181	-0.2927	5.443	2.5745
		S5B_590471022	5B	590471022	-0.2748	4.8311	2.7592
		S6D_464495312	6D	464495312	-0.2564	5.3622	2.3657
		S7A_672854561	7A	672854561	-0.429	9.1713	5.5934

	pKWmEB	S2B_403704287	2B	403704287	0.2198	4.4812	2.7467
		S7A_672854561	7A	672854561	-0.4362	10.7186	6.333
	ISIS EM- BLASSO	S1A_531721876	1A	531721876	0.2756	4.2402	2.2785
		S5B_590471022	5B	590471022	-0.2693	4.0392	2.6661
		S6D_464495312	6D	464495312	-0.2519	5.1507	2.297
		S7A_672854561	7A	672854561	-0.4527	10.1841	6.2647
W	mrMLM	S1A_21719710	1A	21719710	0.1458	4.4106	2.6642
		S3D_260727860	3D	260727860	0.2248	4.7129	4.1162
		S6A_89170732	6A	89170732	-0.1722	5.0107	3.6907
	FASTmrMLM	S6A_89170732	6A	89170732	-0.1265	4.1086	1.9913
	pLARmEB	S3D_260727860	3D	260727860	0.1732	4.2908	2.4422
	pKWmEB	S7A_696981995	7A	696981995	-0.1984	5.024	4.1558
	ISIS EM- BLASSO	S2A_697000089	2A	697000089	-0.2603	4.4115	3.4247
		S3D_260727860	3D	260727860	0.1962	5.7957	3.1351
		S6A_89170732	6A	89170732	-0.1382	4.9192	2.375
L	mrMLM	S4A_733713873	4A	733713873	-2.1206	4.219	3.8171
	pLARmEB	S2D_399428380	2D	399428380	-1.8788	4.5229	1.8819
		S4A_733713873	4A	733713873	-1.7411	4.2532	2.5734
	pKWmEB	S4A_733713873	4A	733713873	-1.7729	4.8435	4.4664
		S5B_433023852	5B	433023852	-1.7426	4.745	2.5037
		S7B_459605010	7B	459605010	2.097	4.1829	3.1116
	ISIS EM- BLASSO	S2B_17144240	2B	17144240	1.7	4.848	1.8379
		S2D_32455414	2D	32455414	1.3809	4.1418	1.0839

		S2D_399428380	2D	399428380	-2.0728	5.8494	2.2904
		S4A_733713873	4A	733713873	-1.8096	4.7114	2.7796
		S5B_45705998	5B	45705998	-1.8157	4.0594	1.6219
		S5B_455689900	5B	455689900	-1.3308	4.5096	1.4912
		S7A_466127003	7A	466127003	-1.1669	4.0086	1.1584
		S7B_459605010	7B	459605010	2.2121	4.1829	1.9052
SA	mrMLM	S5A_698528417	5A	698528417	-145.6928	29.7843	21.9901
		S7A_644620616	7A	644620616	-42.5442	5.5646	2.09
	FASTmrMLM	S5A_698528417	5A	698528417	-138.4548	26.3815	19.8594
		S7A_644620616	7A	644620616	-42.8039	5.2469	2.1156
	FASTmrEMMA	S5A_698528417	5A	698528417	-284.3217	25.2849	20.6124
	pLARmEB	S2B_455867951	2B	455867951	-49.3791	4.5016	2.6909
		S3B_562274249	3B	562274249	-41.1567	4.176	1.3272
		S5A_698528417	5A	698528417	-150.2117	32.4273	22.7134
		S7A_644620616	7A	644620616	-41.0424	5.2674	1.89
	pKWmEB	S4A_575936480	4A	575936480	-43.222	4.7996	1.6573
		S5A_698528417	5A	698528417	-111.7406	27.9114	24.5734
		S7A_153804344	7A	153804344	-62.888	4.6109	2.775
		S7A_644620616	7A	644620616	-37.6489	4.6794	2.4553
	ISIS EM-BLASSO	S5A_698528417	5A	698528417	-143.115	29.1987	21.2188
		S7A_644620616	7A	644620616	-47.0522	6.4016	2.5564