ABSTRACT

DEWITT, NOAH DANIEL. Exploring and Predicting Genetic Causes of Yield Variation in Winter Wheat. (Under the direction of Gina Brown-Guedira).

For most economically important plant species, yield is the primary focus of crop improvement, as increasing it increases the amount of food that can be obtained from a production system by human activity. Growers of soft red winter wheat (*Triticum aestivum* L.) receive a return on investment in proportion to their per-acre yield, given that harvested grain meets quality standards. Yield is also genetically complex, integrating information on interactions between genotype and environment that occur over the course of a growing season. Yield relates to the total accumulated and reallocated resources obtained by plants over the course of their lives, and any factors that affect any aspect of plant growth and development will also affect yield. These complex interactions between genetic variation and environmental variation are difficult to untangle. By understanding important parts of those interactions, breeders and geneticists may better develop cultivars that maximize performance across a range of possible environmental conditions.

Much previous work has been done to understand and breed for yield. In **CHAPTER 1**, I present some background on wheat, its history, and major genes that have been characterized that influence wheat phenology. I include discussion on various ways that geneticists and breeders can predict future phenotypes, including methods that take advantage of the characterization of major genes and methods that are blind to it, and discuss the yield component framework that is a major focus of the research presented here.

In the following chapters, I present work that explores some parts of the genetic basis of yield variation in our germplasm. In **CHAPTER 2**, I attempt to break down the genetic character of traits that measure part of the variation in life histories, heading date and plant height. I find that the genetic base of both traits is fairly simple in this population, and find that models that explicitly account for major genes are more predictive than models that weigh all sites equally. In **CHAPTER 3**, I collect per-spike yield and its component traits. I try to extend the work of chapter 2 to understand how genetic variants for heading date and plant height affect spike yield through the effects of those phenotypes on components of yield, and to understand QTL that alter yield and not phenology. I develop models that rely

on an understanding of the causal structure of the data to first map QTL then understand the mechanisms through which they alter spike yield. I find a set of yield component QTL, some known and some novel, that have large effects on the yield component space as a whole. QTL effects on total spike yield vary from environment to environment due to both altered QTL effects on trait values, and altered relationships between traits.

In **CHAPTERS 4 and 5**, I focus in on a single yield component QTL, associated with the presence of awns, as a "model QTL" to understand the full sequence of events through which DNA sequence variation translates into environment-specific yield difference. I identify the gene underlying *B1* as a C2H2 transcriptional repressor, and find promoter haplotypes associated with increased transcription of *B1* early on in spike development that suppresses awn elongation. I find that this has no effect on phenology, but increases spike yield in some years and environments and decreases it in others. This pattern results from differential increases in grain number relative to differential decreases in grain size that sum to produce either positive or negative effects on spike yield. In **CHAPTER 5**, I use historical data to associate a large number of per-environment yield differences associated with *B1* with climate variables, and suggest that the effect of *B1* on yield may relate to heat stress during grainfill, and suggest physiological explanations for those interactions. The resulting map of intermediate effects, from sequence variation to yield variation, may provide a model for characterizing the varietal GxE that results from the accumulation of many such loci.

Exploring and Predicting Genetic Causes of Yield Variation in Winter Wheat

by
Noah Daniel DeWitt

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Crop Science

Raleigh, North Carolina
2022

APPROVED BY:

J. Paul Murphy

James B. Holland

Christian Maltecca

Ralph Dewey

Gina Brown-Guedira
Chair of Advisory Committee

DEDICATION

To my grandparents, Bernard and Doreen Booth.

BIOGRAPHY

The author was born in 1995 and raised in Brooksville, a small town on the Nature Coast of Central Florida. He became interested in agriculture through local food systems, and in plant breeding as a way to improve those systems as the Citrus Greening (HLB) epidemic changed the local landscape from citrus groves to cattle farms. After an internship in 2011 he worked from 2013 to 2017 under Dr. Stuart McDaniel at the University of Florida, where he was involved in population genetics work on the moss *Ceratodon purpureus*, including comparative genomics of lines collected across a latitudinal gradient and optimizing protocols for transitioning from previous transformation pipelines to one involving CRISPR/Cas9. Starting in 2015, he worked in the Citrus and Stone Fruit Breeding Program under Dr. José X. Chaparro, characterizing the population structure of wild citrus relatives with resistance to HLB. In 2017 he earned a B.S. in Plant Genetics with minors in Bioinformatics and Spanish from the University of Florida, and moved to Raleigh in 2018 to begin research with Dr. Brown-Guedira on mapping, cloning, and utilizing QTL for yield component traits in wheat. Still at the Eastern Regional Genotyping lab under Dr. Brown-Guedira, in 2020 he began working as a research assistant managing the group's genotype database, coordinating SNP calling for Eastern US wheat genomic selection programs. After graduation, he plans to continue to pursue public-sector breeding research with the USDA in Raleigh as a support scientist working on Uniform Nursery coordination and hard wheat breeding for end-use quality traits.

ACKNOWLEDGEMENTS

I need to thank my committee members for most of what I learned at NC State, in and out of the classroom. Thank you to Gina Brown-Guedira, for giving me a chance to come to Raleigh to learn about wheat and supporting me the whole way. I want to thank Paul Murphy, for showing me theoretical and practical aspects of plant breeding, Jim Holland for teaching me about linear models and when not to take my own results too seriously, Christian Maltecca for helping me with analyses and giving me resources to implement research ideas, and Ralph Dewey for helping guide me through my graduate work.

This manuscript is a result of the labor of a whole lab group. I want to thank Mohammed Guedira for sharing his knowledge and passion for the field, and for his friendship over the years. Brian Ward helped me at many points with questions I had about genotyping and modeling, and showed a lot of patience while doing it, while Kim Howell and Jared Smith both have done a huge amount of work to support these projects and train me in lab work. I also want to thank former lab member Kou Vang, and former undergraduate students Destiny Tyson, Rakin Rouf, Anisa Guedira, and Tori Rosen, for helping me with all the tedious data collection (especially counting spikelets per spike) necessary for the following chapters. Nearly all the data presented in this dissertation was collected from plots planted and managed by Justin Page, Charlie Glover, and Myron Fountain.

Some work in this manuscript, results from germplasm and data generated by Sun-Grains breeders and the wheat community – including Jeanette Lyerly, Jerry Johnson, Mohamed Mergoum, Stephen Harrison, Carl Griffey, Nicholas Santantonio, and Rick Boyles. A good deal of this work and other manuscripts originated from help, conversations, and friendships with other WheatCAP graduate students in my cohort, especially Dylan Larkin, Amanda Peters, Priscilla Glenn, Saarah Kuzay, and Andrew Katz.

I also want to also thank other students who made attending NC State and living in Raleigh a pleasant experience, especially Josh Whelan, Jessica Brown, Sydney Graham, Anna Rogers, Lais Bastos-Martins, Zachary Winn, Shannon Bradley, Eric Butoto, and Nicole Choquette, as well as the support from Michael Bennett, Maaz Ali, and Alex Meyer.

And of course, Laura and Daniel DeWitt, who have been nothing but supportive every step of the way.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Review of Literature

1.1 A brief history of Wheat Domestication and Breeding

1.1.1 The wheat family

Common wheat (*Triticum aestivum* L. (2n=6x=BBAADD)) is an allohexaploid crop composed of three sub-genomes, each inherited from a different progenitor species related to extant wild diploid species (Fig. 1.1). Early domesticated "emmer" tetraploid wheat (*Triticum turgidum* ssp. *diccocum* Koern. ex Schweinf. (2n=4x=BBAA)) was domesticated from "wild emmer" tetraploid wheat (*Triticum turgidum* ssp. *diccocoides* (Schrank) Schübl). Wild emmer wheat originated from a cross between the progenitor species *Triticum urartu* Thumanian ex Gandilyan (2n=2x=AA), related to cultivated diploid wheat *Triticum monococcum* L. (2n=2x=AA) and its progenitor *Triticum boeoticum* Boiss. (2n=2x=AA), and a relative of the wild species *Aegilops speltoides* Tausch. (2n=2x=SS, contributing the B subgenome) (Faris 2014). Diploid *T. monoccocum* has been cultivated as "einkorn" wheat for approximately 11,000 years, and is still grown in some parts of Turkey and as a specialty crop in the US and Western Europe (Charmet 2011). Emmer wheat is found in the wild between the Palestinian region and the southern Caucuses, and was likely domesticated in this area around 10,000 years ago. Cultivation of emmer wheat from heterogenous stands by early neolithic communities led to selection for mutations that facilitated the harvest and processing of wheat – *BRITTLE RACHIS 1* (*Br1*) genes consisting of alleles in homeologous genes on the short arms of the group three chromosomes prevents premature seed shattering, allowing for harvest of whole heads (Watanabe and Ikebata 2000; Li and Gill 2006; Nalam et al. 2006; Zhao et al. 2019). Today, modern tetraploid varieties of durum wheat (*Triticum turgidum* ssp. *durum* L. (2n=4x=BBAA)) are grown, and globally used principally in the production of pasta and semolina due to a hard kernel and weak gluten.

The D genome of common wheat was contributed by the wild grass *Aegilops tauschii* Coss. (2n=2x=DD), found principally in the Caucuses and Central Asia. Because of the more recent hybridization between *Ae. tauschii* and tetraploid wheat, occurring approximately 8,000 years ago, sequence polymorphism diversity is much lower in the D subgenome relative to the A and B subgenomes (Charmet 2011; Faris 2014). Additional mutations in major genes occurring after this hybridization generated domestication phenotypes that facilitated use. The domestication gene *Tg* (*TENACIOUS GLUME*) on 2DS facilitates threshing of the grain (Sood et al. 2009). The transition from longer, spelt type heads (*Triticum*

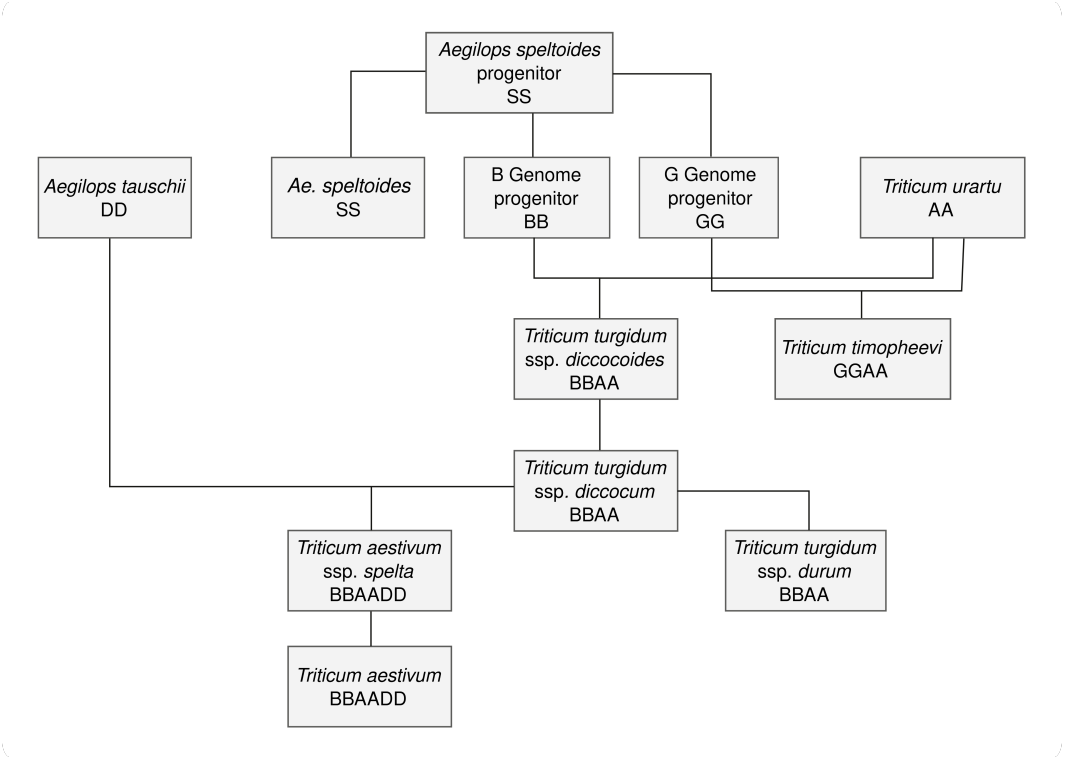

Figure 1.1: **Overview of part of the wheat family.** Adapted with modifications from Tadesse et al. (2019). All *Triticum* and *Aegilops* species descend from a common, 7-chromosome diploid itself descended from a 7-chromosome pair diploid that was also a progenitor of rye (*Secale cereale* L.) and barley (*Hordeum vulgare* L.).

aestivum ssp. *spelta* L.) resulted from one or more mutations in the AP2 class gene *Q* on the long arm of chromosome 5A, which generated varieties with heads that were more compact and easier to thresh (Kato et al. 1998; Faris and Gill 2002; Simons et al. 2006). Spelt wheat continues to be grown as a specialty crop for artisan bread flour. Also on the long arm of chromosome 5A, but in a chromosomal region originating from a translocation between the group 5 and group 4 chromosomes, a mutation in the *Tipped 1* (*B1*) gene suppressed awn elongation (*B1* for "bearded 1", with bearded an alternate way to refer to awned varieties). Selection on awns suppression may relate to ease of harvest and reduction of the portion of chaff in harvested wheat, or potential yield benefits in northern climates. Awn suppression was only fixed in some parts of the world, perhaps because awns protect from herbivory, retain water which can increase susceptibility to pre-harvest sprouting, or because they may improve yields in some climates (Cao et al. 2016; King and Richards 1984). Depending on the market class, hexaploid common wheat is used in the production of bread, pastries,

noodles, seitan, and most other wheat products, as well as in animal feed.

In modern breeding, the large secondary and tertiary genetic pools of wheat have been invaluable for introgression of chromosome segments carrying disease resistance genes. Cereal rye (*Secale cereale* L., 2*n*=2x=RR), also a member of the *Triticeae* tribe, has contributed the entire short arm of its group one chromosome (1R) to common wheat's A and B subgenomes. These t1RS:1AL and t1RS:1BL translocation chromosomes have both been associated with heightened yield and resistances to fungal diseases such as leaf rust (*Puccinia triticina*; *Lr26*), stem rust (*Puccinia graminis*; *Sr31*), stripe rust (*Puccinia striiformis*; *Yr9*) and powdery mildew (*Blumeria graminis*; *Pm8*), as well as the greenbug aphid (*Schizaphis graminum*) (McKendry et al. 1996; Anderson et al. 2003; Anugrahwati et al. 2008).

The secondary gene pools of diploid and tetraploid relatives *T. monococcum* and *T. turgidum* ssp. *diccocoides* (einkorn and emmer wheat, respectively) have proven to be valuable resources for discovery and introgression of novel resistance genes. The primary utility of introgression from einkorn has been in the identification of stem rust genes, principally *Sr21*, *Sr22*, and *Sr35* (Kerber and Dyck 1973; McIntosh et al. 1984). Emmer wheat has contributed stem rust genes, including *Sr2* and *Sr9* (McFadden 1930; Knott and Anderson 1956), leaf rust genes inlcuding *Lr53* and *Lr64* (Marais et al. 2005b; Kolmer et al. 2010), and stripe rust genes including *Yr15* and *Yr36* (Gerechter-Amitai et al. 1989; Uauy et al. 2005). Emmer wheats have also served as an important source of resistance to another fungal pathogen, powdery mildew; *Pm4a*, *Pm5*, and *Pm16* have been introgressed from *T. dicoccoides* into common wheat (Dhaliwal et al. 1993).

The tetraploid relative *Triticum timopheevii* Zhuk. (2n=4x=AAGG) has also proved to be a useful source of resistance genes for commercially important pathogens (Brown-Guedira 1996). Powdery mildew resistance genes *Pm6*, *Pm27*, and *Pm37* derive from *T. timopheevii* (Helmsjrgensen and Jensen 1973; Peusha et al. 2004; Perugini et al. 2008). A novel allele of *Pm1* conferring powdery mildew resistance, designated *MlAG12*, was also brought into North Carolina germplasm by Maxwell et al. (2009). Stem rust genes *Sr36* and *Sr37* derived from *T. timopheevii* have been mapped to chromosomes 2B and 4B, respectively (McIntosh and Gyarfas 1971; Gyarfas 1978). Novel resistance to Fusarium head blight (FHB, caused by *Fusarium graminearum*) has also been mapped to chromosomal segments derived from a *T. timopheevii* parent (Malihipour et al. 2017).

The D-genome donor *Ae. tauschii* has proven invaluable as a source of resistance genes for common wheat pathogens. Examples of pathogens for which resistance genes have been introgressed from *Ae. tauschii* include the UG99 rust variety (Olson et al. 2013), powdery mildew (Murphy et al. 1998, 1999; Miranda et al. 2006, 2007), leaf rust (Leonova et al. 2007), and Hessian fly (Ma et al. 1993). Introgression of resistance genes from the relative of B genome donor *Ae. speltoides* has led to a number of resistance genes for rust, including *Lr28*, *Lr35*, *Lr36*, *Lr47*, and *Lr51* for leaf rust (Wells et al. 1973; Dvořák 1977; Friebe et al. 1996; Dubcovsky et al. 1998), *Sr32*, *Sr39*, and *Sr47* for stem rust (Kerber and Dyck 1990; Friebe et al. 1996; Faris et al. 2008), and *Pm12* and *Pm53* for powdery mildew (Miller et al. 1988; Petersen et al. 2015).

Other *Aegilops* species have contributed major resistance genes. The major leaf rust gene *Lr9* is a translocation originating from *Aegilops umbellulata* Zhuk. ($2n$=2x=UU, Sears (1956)), while powdery mildew resistance gene *Pm13* was derived from *Aegilops longissima* Schweinf. Muschl. ($2n$=2x=S^lS^l, Ceoloni (1984)). *Aegilops ventricosa* Tausch ($2n$=4x=$D^vD^vN^vN^v$) donated a chromosomal segment to wheat chromosome 7D, initially identified for eyespot (*Pseudocercosporella herpotrichoides*) resistance, but which has provided resistance to rust species in the form of *Lr37*, *Yr17*, and *Sr38* (Jahier et al. 1989). Leaf rust gene *Lr54* and stripe rust gene *Yr37* originate from *Aegilops kotschyi* Boiss. ($2n$=4x=$U^kU^kS^kS^k$, Marais et al. (2005a)). Work in this area continues, and further resistance genes not yet heavily used in commercial cultivars have been introduced from *Aegilops sharonensis* Eig (2n=2x=$S^{sh}S^{sh}$), *Aegilops searsii* M. Feldman & M. Kislev (2n-2x=S^sS^s), *Aegilops peregrina* (Hack.) (2n=4x=$U^pU^pS^pS^p$), *Aegilops geniculata* Roth (2n=4x=$U^gU^gM^gM^g$), *Aegilops triuncialis* L. (2n=4x=$U^tU^tC^tC^t$), and *Aegilops neglecta* Req. ex Bertol. (2n=6x=$U^nU^nX^nX^nN^nN^n$) (Zhang et al. 2015). This work is likely to provide a necessary base of resistances to future virulent strands of rust due to the tendency of rust to rapidly evolve virulence to heavily utilized resistance genes.

1.1.2 Wheat breeding in the US

As a principally selfing species, wheat was traditionally cultivated in stands of genetically heterogenous inbred lines. Occasional outcrossing allows for the generation of genetic diversity, but it can be assumed that most selection by pre-modern farmers involved altering the relative proportion of inbred lines within a landrace (Bonjean and Angus 2001). Today,

wheat is the major crop grown for human consumption in the United States and globally. Wheat spread to North America with the occupation of the continent by European colonizers beginning in the early 1600s, with soft wheats from Western and Northern Europe constituting much of the early germplasm base (Brigham 1910). Variety development prior to the 1900s was limited, but a number of early introductions became popular and were cultivated widely. The red variety Red Lammas, grown in England as far back as the 17th century, was brought to Virginia as Red May and cultivated for over two hundred years, with 400,000 acres still planted in 1924 (Ball 1930). The red variety Mediterranean (also called Lancaster) became dominant in the East after being imported into the country in 1819, because of its supposed resistance to Hessian Fly. Purplestraw was being grown prior to the 1800s, and was still grown on 116,00 acres in 1924. By 1875 over 25 million acres of wheat was being grown in the United States, with 49 million acres grown in 1901, greater than the 2021 acreage of 47 million.

Development of the pure-line method in wheat on a larger scale occurred in Svalöf, Sweden with the hiring of Hjalmar Nilsson by the Swedish Seed Association in 1889, who implemented a "repeated pedigree" method inspired by the Vilmorin program in France, but scientific breeding programs took longer to establish in the US (Åkerberg 1986). Single-head selections from landraces, including some off-types that may have constituted outcrossing events, led to early North American pure-line varieties. The farmer David Fife made a single-head selection of a hard winter wheat out of a stand of spring-planted wheat of Polish provenance in Ontario that became known as Red Fife, and that contributes to the modern US hard wheat germplasm base, along with its direct descendant Marquis. Abraham Fultz of Pennsylvania selected the soft, awnless variety Fultz from a field of Mediterranean (synonymous with Lancaster) in Pennsylvania in 1862; Fultz was grown on nearly 5 million acres by 1919 (Ball 1930).

Ultimately, the limits of cultivar improvement from single-head selections was reached. Botanist Cyrus G. Pringle began developing varieties from crosses in Vermont starting in the 1870s, and W.M. Hayes, based at the University of Minnesota experimental station, pioneered a system by which commercial varieties were tested and grown to provide both variety recommendations and pure seed to farmers in the state (Ball 1930). Beyond variety testing, Hayes developed a system for the development of pure lines from an initial hybridization event using the "centagener" system in which selections were made from a spaced plot of 100 individual progeny from a single plant. By the early 1900s farmer-breeders

were developing varieties from crosses – S.M. Schindle crossed Fultz with Mediterranean to produce Fulcaster, which enjoyed success as both a variety and a parent. Development of varieties from crossing of inbred lines would be more rigorously taken up by land-grant universities and by private companies such as the Coker Pedigreed Seed Company starting in the early 1900s. The importance of these older cultivars is apparent even going into the modern period; cultivar Knox, an important parent in cultivars released in the 1980s developed at Purdue in 1953, is primarily a product of crosses involving Fultz and Purplestraw (Murphy et al. 1986). Fultz contributes heavily to Arthur, another cultivar produced by Purdue in 1968 that contributed heavily to the modern soft wheat germplasm base (Cox et al. 1986). In general, Mediterranean and its descendants have been important in constructing the soft wheat germplasm base, though the SRW germplasm base of the East has been less genetically homogeneous than wheat in other regions.

The development of steel roller mills in the 1870s led to the cultivation of hard wheats, including those of Eastern European provenance, in the Great Plains for bread flour. Variation for kernel hardness is principally controlled by the *Pin-D1* locus on 5DS (formerly the *Hardness Ha* locus), consisting of two mutations (*Pina-D1* and *Pinb-D1*) in adjacent friabilin subunits (Morris 2002). These genes are missing altogether on the A and B subgenomes, giving durum wheats a hard kernel, and the introduction of wild-type *pin-d1* with the addition of the D genome in hexaploid wheat confers a softer kernel texture through the action of the frialibin protein, which transports proteins to granule membrane surfaces and alters the distribution of starches within the endosperm. Different alleles for *Pin-D1* involve mutations in these two subunits (*Pina-D1b* and *Pinb-D1b*) which interfere with friabilin function and re-confer hardness. The meal produced from the milling of harder wheats are courser, and the altered distribution of starch granules within grain meal particles are preferred for making flour for bread. Botanist Mark A. Carleton noticed the suitability of hard Eastern European lines like Turkey Red (also referred to as Kharkof, Turkey, and Crimean) brought by Mennonite immigrants into Kansas, and traveled to the Russian Empire to collect, test, and distribute additional varieties (Carleton 1900). This germplasm base, although not totally uniform due to differing selections from what were genetically heterogenous landraces, became responsible for the entirety of wheat acreage in the Great Plains. The influence of these early varieties can be seen in the pedigree of the variety Thatcher, which dominated acreage in the Plains and Canada for years starting in the 1930s, and was derived from a cross between Kanred (a selection from Turkey Red)/Marquis and Marquis/Iumillo. Lines derived from Turkey Red became the major genetic contributor

to the hard red winter germplasm base, and a significant contributor to the soft wheat germplasm base (Murphy et al. 1986). The genetic homogeneity in wheats grown in the Great Plains which resulted from the early success of a small number of related varieties decreased dramatically over the middle part of the 20th century as a result of dedicated breeding work, but with a core germplasm base still closely related to Turkey Red (Cox et al. 1986).

The spread of different wheat classes in different regions of the US led to the formalization of the market class system by the United States Grain Standards Act of 1916 (Slaughter et al. 1992). The distribution of regional milling and processing networks have reinforced the geographic boundaries of market classes, so that different parts of the country specialize in the production of different kinds of wheat. Winter wheat is divided into four classes based off red versus white kernel color and hard versus soft kernel texture. Soft wheat is principally grown east of the Great Plains, where acreage is dominated by soft red wheat, in part because environmental conditions are less suited for bread wheat production. As Mark Carleton wrote in 1920, "The annual wheat production [of the southern wheat district] is comparatively small. In the greater portion of the district the combination of great rainfall with mild temperature is not conducive to great success in wheat growing. The soil is also generally not of the best for such purpose. Rust is always very bad, because of the constantly damp, warm climate" (Carleton 1920). These soft wheats are used for other wheat products such as cookies, cakes, and biscuits, as well as for animal feed. A minuscule amount of hard wheat acreage is grown in the East for specialty products and to reduce shipping costs for millers located in the Eastern US, and soft white wheats are grown on the Great Lakes coast of Michigan and New York for cake and pastry flours. Hard red wheat for bread is principally grown in the Great Plains, where a brief, dry grain fill period in the summer is conducive to maintaining high quality. Specialty hard white and club wheat are grown in the Pacific Northwest, principally for export to East Asia. In the Dakotas, Montana, and the Canadian Prairie, where temperatures are often too cold for winter wheat to successfully overwinter, spring hard red wheat for bread flour and durum wheat for pasta and semolina are often grown.

A principal difference distinguishing older cultivars from modern ones is the reduced height of most modern wheat. In most genotypes, stature is shortened by one of the two principal "green revolution" dwarfing genes – *Rht-B1b* (*Rht1*) or *Rht-D1b* (*Rht2*). Mutations in DELLA protein-coding genes on chromosomes 4B (*Rht-B1b*) and 4D (*Rht-D1b*) generate

premature stop codons in the N-terminal regions of both genes (Peng et al. 1999). DELLA proteins repress giberellic acid (GA) response, and Peng et al. (1999) suggest that mutations in the N-terminal domain prevents proper degradation of DELLA proteins, leading to constitutive repression of GA signaling, reduced cell elongation, and shortened stature. Both of these variants were initially identified in Japan, where wheat farmers have preferred to cultivate dwarf varieties since at least the sixteenth century (Borojevic and Borojevic 2005). Wheat breeder Gonjiro Inazuka, working at the Iwate Prefectural Agircultural Experimental Station in 1917, crossed a semi-dwarf Japanese landrace carrying these alleles to the American variety Fultz, and in 1924 crossed the resultant progeny to Turkey Red to develop a semi-dwarf cultivar Norin-10 with both dwarfing alleles (Inazuka 1971). These alleles were introduced into the U.S. by Washington State breeder Orville Vogel, and were soon adopted by breeders domestically and globally (Lumpkin 2015). Semidwarf varieties are less likely to lodge, particularly with modern applications of nitrogen fertilizer, and partition a larger portion of biomass to the head as opposed to the stalk. North Carolina State cultivar Blueboy (Brevor/Norin-10//Anderson/Coker 55–9), Cornell cultivar Yorkstar (Genesee*3/3/Yorkwin//Norin-10/Brevor), and Coker cultivar Coker 68-15 utilized these alleles and became common as parents in Eastern soft wheat breeding programs after their release in the 1960s (Guedira et al. 2010). Blueboy particularly became an important parent for cultivars released in the 1980s (Murphy et al. 1986), with important varieties Saluda (Virginia Tech) derived from Coker 68-15, and Hunter from Coker 68-15 and Blueboy. Knox and Blueboy also contributed to notable Virginia Tech cultivar Massey (1981). Florida-302, released from UF in 1984, became another important vector for *Rht* genes with Vogel and Norin germplasm in its pedigree. By the 1980s, a majority of released cultivars in the Eastern U.S. had at least one of the two *Rht1* dwarfing alleles, and today nearly all varieties do.

One weakness of *Rht1* cultivars is that they are GA-irresponsive, and reduced cell elongation reduces not only stature but coleoptile length and grain size. These varieties have smaller grain and need to be planted closer to the surface to emerge properly (Rebetzke and Richards 2000). As a result, a shortened stature produced by a combination of other genes may lead to higher yields. The dwarfing gene *Rht8*, the most frequently used outside of *Rht1*, also has its origins in Japan. Italian breeder Nazareno Strampelli utilized the Japanese *Rht8* semidwarf variety Akakomugi to develop popular Italian semidwarf varieties in the 1910s (Worland et al. 1998; Borojevic and Borojevic 2005). Although a predictive marker for *Rht8* is not available, it is known to be found in soft red winter wheat germplasm. Other genes conferring a semidwarf stature that are GA-responsive have been identified, including

a commonly mapped *Rht* gene in the centromeric region of chromosome 6A (Guedira, unpublished results). With increased weed pressure from *e.g.* herbicide resistant Italian Ryegrass, and advances in straw strength since the 1960s, breeding for non-*Rht1* lines with a somewhat taller stature may prove advantageous in the future.

In modern public-sector wheat breeding, programs cooperate on the federal and regional levels to develop competitive varieties. The establishment of regional late-stage nurseries by the USDA allows for the cooperative testing of both public and private lines across a diverse set of potential target environments, and for direct comparisons in those environments useful for making decision on variety release (see Boyles et al. (2019)). Email and excel-based data sharing have facilitated coordination of late-stage trials between collaborators across public and private institutions in the eastern US. Breeders in the Southeast United States collaborate through the SunGRAINs cooperative, allowing for sharing of royalties, sharing of germplasm, and regional testing of lines at multiple stages that provides advantages of a larger-scale program while allowing individual flexibility. Cooperative trials have evolved into a common genomic selection program across the SunGRAINs programs that has developed into a model for modernizing public-sector breeding. The success of the wheat community in maintaining robust public breeding programs in the face of for-profit encroachment results in part from this success in inter-program cooperation. Research by land-grant institutions into the improvement of wheat genetics, agronomy, and their interaction has driven large increases in U.S. wheat yield. From 1899 to 1904, growers in the United States averaged 13.7 bu/acre (Brigham 1910), but the application of empirical plant breeding promised to change that – in 1899 W.M. Hayes remarked of the new varieties he was testing, including farmer selections "Hayne's Blue Stem" and "Power's Fife", that "23 bu/acre is substantially higher than yields observed in the state [of Minnesota]" (Hayes 1899). Average wheat yields reported across the United States are now in the range of 40-50 bu/acre, while yields under proper management are much higher (USDA Census, 2017). Analyses of Uniform trial data collected over the past 15 years suggest that even today breeders are improving yields at a rate of 0.56 bu/acre per year (unpublished results), suggesting that growth in yields may continue and even increase in pace with implementation of new technologies.

1.2 Wheat phenology

1.2.1 Wheat growth and development

Common wheat is primarily cultivated under two major production systems. Wheat planted in the spring matures and fills grain during the spring and summer. Lines may instead be planted in the fall, where wheat germinates and develops into small plants, and lays mostly dormant until winter dormancy release in the spring. This division is typically regional, and relates to climate differences between regions, but also has effects on downstream end use which can incentivize growth of different classes in the same geographic region. Typically, fall-planted winter wheats are grown in temperate climates where sufficient chilling hours are accumulated during the winter months to promote spring flowering, but temperatures are not so low as to cause major injury to the crop. Spring production systems dominate in both Northern production systems, especially in the Dakotas and the Canadian Prairie, and in subtropical and equatorial regions of the world. Wheat varieties must be developed for specific production systems and generally in specific regions, so that they respond to environmental cues and mature at the appropriate times.

The transition of fall-planted wheats from dormancy to reproductive growth results from the expression of the ortholog of the *Flowering Time (FT)* protein, *VRN3* in wheat, and its movement through the phloem from the leaves to the shoot apical meristem (SAM) (Gauley and Boden 2019). When sufficient FT signal reaches the SAM, plants irreversibly commit to reproductive growth and the SAM begins developing into the adult wheat spike. Transcription of *Vrn3* is initiated only through activation of a set of upstream genes which integrate signals from night length (the photoperiod pathway), accumulated chilling hours (the vernalization pathway), and the age of the plant in units of thermal time (the earliness per-se pathway). In winter wheat, the SAM is maintained underground to protect the gametic tissue from cold stress, but after dormancy release the elongating tiller carries the growing spike upwards. The growing spike is visible as a widening of the leaf sheath just below the apex of the tiller, before emerging at "heading" as the flag leaf folds back and the peduncle (the uppermost, exposed internode) elongates. Anthesis typically occurs a few days after heading, and mature plant height is reached shortly after with the termination of the extension of the peduncle.

Maturity is an important secondary trait for wheat breeders, as non-optimal heading

dates for a given environment may have negative yield effects. The ability of plants to integrate information on and respond to their environments appropriately is necessary to produce stable yield over many years and locations (Lewontin 1957; Allard and Bradshaw 1964). Genetic variability for these responses changes the developmental stage-specific environments of plants through a back-and-forth process, as plants construct and are constructed by their local micro-environments (Levins and Lewontin 1985). Later maturity is generally associated with an increase in total spikelets per spike and grain number, as a longer development time reduces the number of spikelet primordia aborted in the developing SAM (Boden et al. 2015). Earlier maturity tends to reduce grain size by shortening the grain fill period during which carbohydrates are allocated to developing grain. A too-early heading date may also reduce yield by exposing developing spikes to below-freezing temperatures that may reduce grain number by rendering florets infertile. Heading too late risks reducing yield by exposing plants to heat or drought stress during grain fill, reducing grain size by preventing the movement of carbohydrates into developing grain. Heat stress during grain fill lowers grain size by increasing root and shoot temperatures, decreasing sucrose levels, and deactivating the starch synthase enzyme that assimilates carbon, which is particularly heat-sensitive in wheat (Bhullar and Jenner 1985; Hawker and Jenner 1993; Guedira and Paulsen 2002). Understanding the genetic basis of variation for flowering time within breeding populations, and selecting for appropriate heading dates, is therefore a priority for wheat breeders and geneticists.

1.2.2 Variation in major genes alter wheat phenology

The division of production systems into winter and spring planting is generally, but not always, associated with allelic variation at the vernalization genes. The major heading date genes of wheat involve allelic series at the A, B, and D homeologues of the major integrator of vernalization signal, *Vernalization1* (*Vrn-1*) (Yan et al. 2004; Fu et al. 2005; Li et al. 2013), and the major integrator of photoperiod signal, *Photopeiord1* (*Ppd1*) (Beales et al. 2007; Nishida et al. 2013; Guedira et al. 2016), both of which act upstream of *Vrn3* in the wheat flowering time pathway. A series of large-effect alleles in important genes in the wheat flowering time pathway condition major additive genetic variation for heading date and have facilitated the adaptation of wheat to different growing conditions in different parts of the world. Typically, fall-planted wheats have the wild-type, sensitive vernalization alleles also found in wild relatives, but facultative winter wheats with spring alleles at vernalization

loci can overwinter using signal from night length to determine dormancy release in the spring.

In *Drosophila*, a large number of genes with obvious physical effects and a small number of chromosomes allowed genetic maps to be constructed early on (Sturtevant 1913). With relatively few of these physical markers and a much larger number of chromosomes, mapping studies in wheat first became possible with the advent of cytology and the construction of detailed libraries of monosomic and nullisomic lines (Sears 1944, 1953). The direct observation of nullisomics may permit the association between a gene and a chromosome by observing qualitative changes in the nullisomic lines lacking a whole chromosome or chromosome arm. Additionally, F_2 populations can be generated from a wild type and monosomics for different chromosomes to study altered ratios of inheritance for a character, or varieties can be compared through a more detailed transfer of a chromosome from a different variety into a monosomic for that chromosome, creating substitution lines differing for that whole chromosome. Chromosome 5A (referred to as chromosome IX by Sears and as chromosome "C" in prior work) was the first chromosome for which genetic maps were constructed, in part due to the co-location of major dominant alleles on the long arm. Winge (1924) used nullisomics to co-locate the "grannenhemungsfaktor" (awn-inhibition factor, *B1*) and the "ahrenverkurzender faktor" (ear-shortening factor, *Q*) on chromosome "C" (5A).

Major genes controlling the quantitative trait of heading date have sufficiently large effect to be mapped within this paradigm. The spring allele *Vrn-A1* was first identified by Åkerman and J. MacKey (1949) as S_k. Aneuploids and nullisomics were used to place a major gene controlling spring type on chromosome IX (5A) alongside *B1* and *Q* (Unrau 1950; Sears 1954). By 1968 S_k had been identified by Mackay as being *Vrn-A1* according to Pugsley (1971), who was able to assign genotypes at *Vrn-A1* to a number of Australian cultivars. Using populations derived from 5A substitution lines, Law et al. (1976) were able to map *Vrn-A1* to at least 50 cM distal to the centromere on the long arm. Using chromosome substitution lines, Snape et al. (1985) were able to map *Vrn1*, *B1*, and *Q* together on the long arm of 5A, though they were ordered incorrectly. With the help of restriction fragment length polymorphism (RFLP) markers, Kato et al. (1998) ordered the major genes correctly; from proximal to distal, *Vrn1*, *Q*, then *B1*.

As variation in the response to chilling hour accumulation is in large part generated by variation in the *Vrn1* genes, variation in plant response to day length is primarily generated

by variation in the *Photoperiod1* (*Ppd1*) genes on all three homeologues (Beales et al. 2007; Nishida et al. 2013; Guedira et al. 2016). Law et al. (1978) state that two genes controlling photoperiod response were identified by J.R. Welsh and presented in 1973, mapped to 2D and 2B as *Ppd1* and *Ppd2* respectively. The earliness allele with the largest effect is associated with *Ppd1*, now known as *Ppd-D1a*, which was also brought into European and American germplasm from Akakomugi and is typically in close linkage with the dwarfing allele of *Rht8*. A smaller-effect mutation affecting *Ppd1* on chromosome 2A known as *Ppd3* (now known as *Ppd-A1*) which segregates in US germplasm may have its origins in germplasm from India (Law and Worland 1997).

Advances in sequencing and molecular biology allowed for the cloning of some of these major genes. The gene underlying *Vrn1* was identified by Yan et al. (2003) as a MADS-box *AP1* gene orthologous to meristem identity genes *AP1*, *CAL*, and *FUL* in *Arabadopsis*. Alleles generating a spring growth habit through constitutive up-regulation were identified at the *Vernalization1* (*Vrn1*) loci on the three group 5 homeologues (Yan et al. 2004; Fu et al. 2005; Li et al. 2013), while additional alleles in vernalization-sensitive (winter) lines reduce the degree of vernalization sensitivity and promote earlier flowering, especially in warmer climates (Fu et al. 2005; Díaz et al. 2012; Guedira et al. 2014, 2016; Kippes et al. 2018). The cloning of a CCT-domain containing gene (CONSTANS (CO), CO-LIKE, and TIMING OF CAB1) as underlying *Ppd-H1* in Barley by Turner et al. (2005), and the colinearity of the region on 2H with the mapped *Ppd-D1* locus on the group 2 chromosomes in wheat, lead to the identification of orthologous CCT proteins as underlying *Ppd-1* on all three wheat homoeologues (Beales et al. 2007; Wilhelm et al. 2009; Díaz et al. 2012). Sequence variation at or near these three genes leads to constitutive expression of *Ppd1* and the photoperiod-insensitive, earlier flowering habit associated with certain alleles of *Ppd-A1*, *Ppd-B1*, and *Ppd-D1* (Beales et al. 2007; Nishida et al. 2013; Guedira et al. 2016).

A common approach to attempt to identify genes underlying quantitative trait variation begins with a marker-trait association and a position, moves to identifying the associated gene, and then works to understand the mechanisms underlying the translation of sequence variation to trait variation through the action of that gene. Here, characterization of these major genes started from an observation of varieties with varying responses to environmental cues and moved forward to characterization of segregating genes, then positioning on chromosomes and identification of the associated gene. At the same time, these genes are commonly detected in marker-trait association studies. Prediction models

relying on assumptions of large numbers of small-effect alleles with additive effects do not capture the effects on yield of these major genes. As a start, fitting markers for these genes as fixed effects may improve genomic-selection models for related traits (Sarinelli et al. 2019). As these major genes alter plant response to environmental cues, explicitly modeling interactions between major alleles at these loci and environmental variables may improve future predictive models, and help us understand the contribution of variation in growth habit to yield variation.

1.3 Prediction of phenotypes from genotypes

Wheat geneticists have been able to associate qualitative differences in phenotypes with discrete loci. But typically, the traits of interest for wheat breeders are quantitative, and under the control of many thousands of causal variants. For the improvement of these traits, prediction of phenotypes can occur with or without the use of genetic information, or even information on the physical locations of variants.

1.3.1 Covariance between relatives and prediction

Individuals with a shared ancestry often have more similar phenotypes than two individuals randomly sampled from a given population (Fisher 1918; Wright 1921). A major part of the covariance between relatives results from those two individuals being more likely to share DNA sequence at any given locus than two random individuals, although relatives may also have more similar environments ($\sigma^2_P(x,y) = \sigma^2_G(x,y) + \sigma^2_E(x,y) + 2\sigma_{G,E}(x,y)$). Breeders and geneticists have used this insight to develop a series of tools that allow for the prediction of phenotype for individuals with little or no observed phenotype – covariance between relatives allows for the prediction of individuals based on the previous observation of identical or related genotypes. In most plants, the ability to generate clones, inbred lines, or hybrids between inbred lines allows for the replication of the same genotype in multiple environments. Properly designed replicated experiments permit the partitioning of environmental convariances from genetic covariances, and can eliminate the covariance between genotypes and environments. Breeders rely on the observation of identical genotypes to predict the performance of the same genotype in future environments, assuming that some

population of sampled environments is representative of the target environments. At the same time, estimation of additive genetic variation (σ^2_A), dominance genetic variation (σ^2_D), epistatic genetic variation (σ^2_I) and environmental variation (σ^2_E) gives breeders a sense of how predictive those values are in a separate set of environments. Depending on the genetic structure of the population, predictive ability as a function of the ratios of estimated variance components analagous to r^2 may take the form of broad-sense heritability for clones (broadly, $H = \frac{\sigma^2_A+\sigma^2_D+\sigma^2_I}{\sigma^2_A+\sigma^2_D+\sigma^2_I+\sigma^2_E}$) or inbreds ($H = \frac{\sigma^2_A+\sigma^2_I}{\sigma^2_A+\sigma^2_I+\sigma^2_E}$). The relative ratios of genetic variance components in a given population of a given generation also determines the ability to predict future phenotypes; in selfing populations, the ability of a phenotype to predict its progeny is lower in earlier generations where σ^2_D is proportionally higher than σ^2_A relative to later generations (Falconer and MacKay 1995). Comstock and Robinson (1952) developed a series of experimental designs at NC State, termed NC Designs I, II, and III, that allowed for the estimation of additive and dominance variances through variance components associated with different levels of relationships among the individuals in those mating designs.

In other commercially important species, like cattle (*Bos taurus* L.) or pine (*Pinus taeda* L.), replication of identical genotypes is impossible or cost-prohibitive. In this case, prediction of a genotype's value for some trait is based on the observations of relatives. In the simplest case, many progeny produced from the mating of one individual to others can be phenotyped. All of these half-sib individuals share a common kinship with the common parent, allowing for an estimation of the genetic merit of the common individual, even if that individual is not observed. The predictive ability in these cases relate to the inherited additive effects of alleles and the narrow-sense heritability, $h^2 = \frac{\sigma^2_A}{\sigma^2_A+\sigma^2_D+\sigma^2_I+\sigma^2_E}$. In crops such as maize (*Zea mays* L.) where hybrids dominate commercial production, replicated testing of offspring from crosses between different pairings of a set of inbred lines is used to predict both the overall genetic merit of lines as a hybrid parent (general combining ability; GCA) and the genetic merit in combination with specific individuals (specific combining ability; SCA) (Lynch and Walsh 1998).

Predictions for all individuals in a given population can be obtained at once using a linear model ($\mathbf{y} = \mathbf{Xb} + \mathbf{e}$, for a vector of phenotypes $\mathbf{y}$, vector of genotype effects $\mathbf{b}$, vector of residuals $\mathbf{e}$, and design matrix relating phenotypes to genotypes $\mathbf{X}$, with number of rows equal to the length of $\mathbf{y}$ and number of columns equal to the length of $\mathbf{b}$) (Isik et al. 2017). In the simplest case, least squares estimates calculate a mean for each genotype across all

replications by solving $\hat{\mathbf{b}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$, where $\mathbf{X}'\mathbf{y}$ sums values across genotypes and $(\mathbf{X}'\mathbf{X})^{-1}$ weights each sum by the number of observations, so that the least squares estimate of any individuals equals the sum of observations over the number of individuals. In a mixed linear model, genotype values may be fit as a random effect where estimates are shrunk closer to zero inversely proportional to the information available to make that estimate. Here, $\mathbf{y} = \mathbf{Z}\mathbf{u} + \mathbf{e}$, where $\mathbf{u}$ is a vector of coefficients for random effects and $\mathbf{Z}$ is a design matrix relating phenotypes to random effects. In the simple case where all phenotypes are centered and no fixed effects are fit, solving for genotype values takes the form $\mathbf{y} = \mathbf{Z}\mathbf{u} + \mathbf{e}$ and can be solved similarly to before as:

$$\hat{\mathbf{u}} = (\mathbf{Z}'\mathbf{Z} + \lambda\mathbf{I})^{-1}\mathbf{Z}'(\mathbf{y} - \bar{y})$$

This formulation is equivalent to the least squares estimator except that prior to taking the inverse of the diagonal matrix of number of observations of each genotypes to derive the weights ($\mathbf{Z}'\mathbf{Z}$), a single scalar (λ, the ridge factor) is added to each element of the diagonal, which has the effect of disproportionally placing more weight on genotypes for which more observations were recorded (*e.g.*, comparing two genotypes with equal phenotypes but with one versus two reps, $n * \frac{1}{1} = 2n * \frac{1}{2}$, but $n * (\frac{1}{1} + \lambda) < 2n * (\frac{1}{2} + \lambda)$ for $\lambda > 0$). The estimates obtained from this rely on λ, which is computed as a form of inverted heritability as σ^2_e / σ^2_u, where σ^2_u is the variance associated with the random effects and may be constructed from different types of genetic variance components depending on the genetics of the system of interest. A decrease in heritability, representing an increase in σ^2_e relative to σ^2_u, increases λ, and further shrinks the estimates of genotypes for which less information is available relative to those for which more information is available.

When phenotypes are not centered and additional fixed effects are included in the model, coefficients are estimated for both fixed effects as $\mathbf{b}$ and random effects $\mathbf{u}$ in the form $\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{u} + \mathbf{e}$ (Bernardo 2010). When the number of random effects becomes large, the inversion of large matrices may become computationally difficult. Henderson (1976) presented a more computationally efficient method for jointly obtaining values for fixed effects, non-genetic random effects, and random effects for genotypes. A simpler and more generalized form that combines random effects and jointly solves for $\hat{\mathbf{b}}$ and $\hat{\mathbf{u}}$ can be written as:

$$\begin{pmatrix} \mathbf{X}^T\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}^T\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}^T\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}^T\mathbf{R}^{-1}\mathbf{Z}+\mathbf{G}^{-1} \end{pmatrix} \begin{pmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} \end{pmatrix} = \begin{pmatrix} \mathbf{X}^T\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}^T\mathbf{R}^{-1}\mathbf{y} \end{pmatrix}$$

For fixed effect design matrix $\mathbf{X}$, random effect design matrix $\mathbf{Z}$, and covariance matrices for random effects $\mathbf{G}$ and for residual effects $\mathbf{R}$. In the simple case of identical and independent variances of residual effects, $\mathbf{R} = \mathbf{I}$, and since the inverse of the diagonal matrix $\mathbf{I}^{-1} = \mathbf{I}$, Henderson's equations simplify nicely to:

$$\begin{pmatrix} \mathbf{X}^T\mathbf{X} & \mathbf{X}^T\mathbf{Z} \\ \mathbf{Z}^T\mathbf{X} & \mathbf{Z}^T\mathbf{Z}+\lambda\mathbf{I} \end{pmatrix} \begin{pmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} \end{pmatrix} = \begin{pmatrix} \mathbf{X}^T\mathbf{y} \\ \mathbf{Z}^T\mathbf{y} \end{pmatrix}$$

If variance components σ_e^2 and σ_u^2 used to calculate λ are known, $\hat{\mathbf{b}}$ and $\hat{\mathbf{u}}$ are BLUEs (best linear unbiased estimates) and BLUPs (best linear unbiased predictions) respectively (Henderson 1975). For real-world data, where variances are typically estimated iteratively at the same time as parameters, BLUPs and BLUEs are really estimated BLUPs and BLUEs – unbiased but not necessarily optimal (Kackar and Harville 1981). Standard errors for n individual BLUPs can be computed from the n diagonal elements of the bottom-right sub-matrix of the inverse of the whole left-hand side matrix, and can be used to calculate reliability ($\hat{r}^2 = 1 - \frac{(Std.Error)^2}{\hat{\sigma}_u^2}$), a direct measure of the predictive ability of the BLUP analogous to the heritability (Isik et al. 2017).

If pedigree information is available, this same method can be used to improve the accuracy of estimates of a genotype by incorporating information on the phenotypes of relatives in proportion to their kinship to said genotype. When a kinship matrix $\mathbf{A}$ is constructed such that each individual element $A_{i,j}$ equals twice the coefficient of coancestry between i and j, covariances between individuals are modeled as their kinship coefficient times the additive genetic variance ($\mathbf{G} = \sigma_A^2\mathbf{A}$) (Lynch and Walsh 1998). A linear mixed model equation incorporating pedigree information is solved as above except that instead of assuming zero covariance between individuals by using the identity matrix $\mathbf{I}$, the inverted kinship matrix $\mathbf{A^{-1}}$ is used. Solutions for genotype effects then account for expectations of phenotypic covariances between individuals in proportion to their coefficient of coancestry:

$$\hat{\mathbf{u}} = (\mathbf{Z}'\mathbf{Z}+\lambda\mathbf{A^{-1}})^{-1}\mathbf{Z}'(\mathbf{y}-\bar{y})$$

This method is equivalent to the computation of selection indices for each individual (Robinson 1991), except that it occurs simultaneously, so that an individual's phenotype is the sum of the phenotypes of its relatives, weighted by its relationships to those relatives and itself and the amount of information available for its relatives and itself. If no information is available on an individual, this method produces a prediction of that individual's value based solely on the phenotypes collected on its relatives.

Considering one form of the breeder's equation $R = \frac{ir\sigma_A^2}{L}$ (where yearly response to selection R equals the product of selection intensity i, accuracy of selection r, and additive variation σ_A^2 over generation interval L), incorporation of information on relatives when predicting lines may improve total response to selection. For starters, incorporation of relatives' phenotypes may increase the accuracy of selection r. Perhaps more importantly, prediction of unobserved individuals or early-stage individuals with little phenotypic data available may be useful in reducing breeding cycle time L, allowing for recycling of individuals likely to be high-yielding as parents prior to evaluation, or prior to the point where they would normally be recycled as parents. A common problem in pedigree-based wheat breeding programs is the availability of different phenotypes with different difficulties of evaluation at different stages in the program; if some of these phenotypes are correlated, a multivariate model may provide greater prediction of individuals with missing information for one phenotype based both on relatives' phenotypes and phenotypes for correlated traits (as a practical example, prediction of more evaluation-intensive disease traits from less intensive disease traits in Larkin et al. (2021) and Gaire et al. (2022)).

1.3.2 Use of markers to track chromosomal segments

An alternate approach to predicting phenotypes of unobserved individuals is to use information on sequence variation as measured by molecular markers. This approach is particularly useful if a small number of large-effect genes cause a substantial amount of the overall trait variation within a population of interest. Even before the development of molecular markers, geneticists sought to identify major genes affecting quantitative traits: indicators of the segregation of alleles with large effects include increases in heritability after selection (if the allele is initially at a low frequency) (Latter 1965), preservation of genetic variance with repeated cycles of crossing individuals from the two tails of the phenotypic distribution (Wright 1952), and various tests of departure from normality or

differing variances of sibling sets (Hammond and James 1970; Fain 1978). When a major gene for a quantitative trait is tightly linked to a gene with some qualitative effect on appearance, that physical marker can serve as a marker for the QTL in lieu of molecular information (as originally proposed by, Sax 1923). In the case of wheat, some major genes identified in this manner may be tracked without molecular information, for example by not vernalizing populations segregating for an allele at a *Vrn-1* locus. For most practical applications, however, molecular markers in linkage with the causal variant or that test the causal variant directly are necessary to exploit information on these major QTL for marker-assisted selection (MAS).

Initial studies on molecular markers used isozymes, which allow for the measuring of sequence variation within exons via the differential amino acid sequence of transcribed proteins. The measurement of changes in frequency of isozyme alleles over the course of selection for a trait may allow for an association of changing allele frequency with a change in phenotype, implying that variants in the chromosomal chunk associated with the isozyme variation had a causal effect on the trait value (Stuber and Moll 1972). A more powerful tool relies on the limited total number of cross-over events that occur per generation on a single chromosome, which generates a correlation between markers as some function of their physical proximity on the chromosome. Working at NC State, Lewontin and Kojima (1960) termed this association "linkage disequilbrium", and defined it as deviation from linkage equilibrium D, which is decays over n generations or random mating as a function of recombination frequency c as $D_n = (1-c)^n D_0$ for any two loci. The linkage of isozymes and nearby causal polymorphisms allows for direct association of a chromosomal segment with a difference in mean phenotype within a population of full-sib F_2 (*e.g.* Edwards et al. 1987) or inbred lines (*e.g.* Stuber et al. 1987), and the tracking of such a quantitative trait locus (QTL) and its placement onto a genetic map. Within a population of full-sib inbreds produced via single seed descent or double haploid induction, for a given locus there should be two alleles in approximately equal frequencies, and there should be no correlation between the allele at a locus and alleles at loci on other chromosomes. The size of the difference in mean phenotypes between the two classes of genotypes at a single locus can then be tested to associate sequence variation near that locus with a causal change in phenotype. In the case of testing markers, because of recombination between a tested marker M and QTL Q, the allele effect will be underestimated in proportion to the genetic distance between M and Q (for the true allele effect a of a QTL Q estimated via marker M linked to Q with recombination fraction $c_{M,Q}$, $\hat{a}_{QQ} = \frac{\mu_{MM} - \mu_{mm}}{2} = a_{QQ}(1-2c_{M,Q})$, so that $\hat{a}_{QQ} < a_{QQ}$ when

$c_{M,Q} > 0$) (Edwards et al. 1987; Lynch and Walsh 1998). In double haploid populations c is the per-generation recombination frequency, while in inbred populations its value approaches twice the genetic distance with continued inbreeding.

A slightly different approach with improved power uses the conditional expectation of the QTL genotype given observed trait distributions (Lander and Botstein 1989). The maximum likelihood approach (somewhat tautologically) identifies the most likely parameters by identifying the maximum value of the likelihood function $\ell(\boldsymbol{\theta} \mid \mathbf{x})$, for a vector of observed values $\mathbf{x}$ and unknown parameters $\boldsymbol{\theta}$ which generate some probability distribution. In practice, it is generally easier to work at the scale of the log of the likelihood $L(\boldsymbol{\theta} \mid \mathbf{x}) = ln[\ell(\boldsymbol{\theta} \mid \mathbf{x})]$. In the case of QTL mapping in an inbred or double haploid population, the likelihood function for an individual trait value z at a locus linked to marker M at distance c is the sum of the likelihoods for the recombinant and non-recombinant cases:

$$\ell(z \mid MM) = (1-c)^2 \phi(z, \mu_{QQ}, \sigma^2) + c^2 \phi(z, \mu_{qq}, \sigma^2)$$
$$\ell(z \mid mm) = c^2 \phi(z, \mu_{QQ}, \sigma^2) + (1-c)^2 \phi(z, \mu_{qq}, \sigma^2)$$

Where likelihood for the total set of n F_∞ individuals is the product of each individual's likelihood given their individual trait values and marker genotypes $\ell(\mathbf{z}) = \prod_{i=1}^{n} \ell(z_i \mid M_i)$. The likelihood of odds (LOD) scores used to generate graphs of QTL likelihood over chromosome intervals is generated from the log 10 ratio (due to a history of use in human genetics, Morton (1955)) of the QTL likelihood at locus l versus the likelihood of a unimodal normal distribution given vector of trait values $\mathbf{z}$ ($\ell_{H0}(\mathbf{z})$), for $LOD_l = log_{10}[\frac{max\,\ell(\mathbf{z},l)}{\ell_{H0}(\mathbf{z})}]$.

Maximum likelihood methods are computationally intensive – in simple cases the maximum value of the function can be solved via the derivative of the likelihood function, but generally iterative methods such as restricted maximum likelihood are used – and regression-based approaches such as the Haley-Knott approximation return nearly equivalent results, much faster (Haley and Knott 1992). Relying on the equivalency of maximum likelihood and regression when assumptions about the normality and independence of errors are satisfied, the LOD scores for markers can be computed as the ratio of residual variances associated with the model distinguishing the two classes at a locus with a null model ($LOD = \frac{n}{2} log_{10}(\frac{RSS_0}{RSS_1})$, where n equals the number of genotypes) (Broman et al. 2003). In the null case, the residual sum of squares equals the maximum likelihood estimator

of phenotypic variance, $RSS_0 = n * \sigma^2 = \Sigma_i (y_i - \overline{y})^2$. Under the alternative hypothesis, fit with the model $y_i \sim g_i + \varepsilon_i$ for QTL genotype g in individual i, the residual sum of squares becomes the squared deviations of each individual phenotype from the mean of its genotype class $RSS_1 = \Sigma_i (y_i - \hat{\mu}_{gi})^2$. In the case of loci in linkage with the QTL, LOD scores so computed are extremely similar to the true likelihood ratio (Haley and Knott 1992). In practice, genotype probabilities can be calculated for an array of loci between flanking markers based on the genotypes of their flanking markers and the recombination fraction separating them prior to QTL mapping, and LOD scores calculated for each simulated locus in the array to compute likelihood curves (Broman et al. 2003).

To determine significance thresholds, the same model is run across permutations where trait values are randomly re-assigned to new genotypes, observing how often a certain LOD value is observed given a population of runs where there is no causal relationship between genotype and phenotypes (Churchill and Doerge 1994). QTL models that repeat the tests incorporating information on already mapped QTL may improve power to map additional QTL.

Early QTL mapping studies relied on isozymes, which were difficult to assay and limited in number. The discovery of RFLP markers allowed for the construction of chromosome-scale linkage maps (Botstein et al. 1980). SSRs (simple sequence repeats) further facilitated linkage map construction, relying on the fact that repeating sections or DNA are more prone to "slippage" of the replication machinery, generating size polymorphisms at these loci (Levinson and Gutman 1987). Assaying of SSRs may be particularly powerful for studies of population structure, as there are generally more than two alleles at a given locus. The development of next generation sequencing technology, particularly arrays ("SNP chips") and reduced-representation sequencing, have made it much more practical to perform QTL mapping studies by greatly increasing the number of variants assayed for a given amount of money and labor. SNP arrays pair DNA probes that capture DNA segments known to harbor SNPs in a population of interest, which are then sequenced using next generation sequencing; the development of the 9K SNP chip in wheat greatly facilitated mapping studies by allowing for the assaying of many SNPs at a low cost (Cavanagh et al. 2013). Genotyping-by-sequencing (GBS) relies on repeatably fragmenting the genome into fixed segments using pairs of restriction enzymes, size-selecting fragments within a given size range, and then sequencing and calling SNPs within those segments (Poland et al. 2012). SNP chips may be preferred in some instances as they deliver less missing data per

SNP, but in general GBS is preferred, as it can typically be performed at a lower cost per sample and is able to identify novel SNPs within a given population. Use of both of these technologies are facilitated by the Wheat reference genome (Appels et al. 2018), which allows for the placement of SNPs onto a physical map alongside gene models and other features of interest.

The principal limitations of QTL mapping for forward prediction are the false positives and over-estimation of QTL effect sizes produced via the Beavis effect as a result of variation in effect estimates and the setting of a hard cutoff for QTL detection (see Beavis 1998; Schön et al. 2004). Another major limitation of QTL mapping for forward prediction is that an association is related to a chromosome interval, and not the causal SNP or structural variant. Two approaches are used to address this: one can develop flanking markers for a polymorphism that are easy to assay, develop a marker associated with the haplotype within which the causal variant resides, or identify and develop a marker for the causal variant directly. Fine-mapping and positional cloning are the principal methods used to identify causal variants underlying QTL, and involve the generation of partially inbred lines heterozygous for different intervals within the broader QTL interval (for a detailed description of modern methodology, see Kuzay et al. 2019). Testing of near-isogenic lines derived from plants heterozygous for different regions allows for the association of a sub-segment with the trait of interest, to determine if the QTL is found within that interval. Once the region is narrowed down to a smaller interval, a new set of recombination events may be generated by selfing to further narrow down the region. Restriction of the QTL region to an area containing only a handful of genes allows for direct investigation of possible causal polymorphisms and their validation through transformation of new lines with the candidate variant.

In a genome-wide association study (GWAS),a diverse population is tested in the same manner as a marker-regression QTL analysis. The idea of studying associations between markers and quantitative traits in diverse populations is not new – for example, Hickman (1889) attempted to associate the *B1* locus for awn status with yield in wheat variety trials – but the development of marker assays covering the whole genome suggested the possibility that the overall additive genetic bases of a trait could be fully characterized. While many studies have identified useful major-effect genes, the goal of characterizing the genetic bases of important traits has generally not been realized (McClellan and King 2010). In part, the divergence of allele frequencies from 0.5 and LD between markers not on the

same chromosome resulting from population structure may reduce power or generate confounding results (Korte and Farlow 2013). Markers with large effects at low frequency, perhaps mutants under strong negative selection, may contribute an outsized effect to total variance due to the squaring of the effect size, but may be difficult to detect due to their low frequency(Gibson 2012). Their detection is further complicated by the need to filter out SNP markers with low minor allele frequency to produce clean data, as sequencing errors may produce false SNPs at low frequencies. In systems where replicated common garden experiments are not possible, covariance between genotype and environment may make causal inference impossible, especially in humans where phenotypic covariance between parents and offspring is produced by social structures (Turkheimer 2012). In the presence of population structure, markers whose frequency varies between sub-populations may have a significant effect if those sub-populations differ on average for the considered phenotype. In the cases where obvious sub-populations are apparent, estimating coefficients for the first few eigenvectors of the covariance matrix between genotypes estimated from marker or pedigree data as covariates may correct for population structure. The full covariance matrix may also be used in a mixed-model framework as noted previously, or the two may fit together (see Lipka et al. (2012)). Despite corrections for population structure, the ability to make causal inference or predict out-of-sample from marker-trait associations is limited, and GWAS results primarily have utility as a precursor to separate validation experiments. These two approaches may be complementary; GWAS studies can begin to narrow down the interval of a QTL, or diverse haplotypes associated with one allele may be used to rule out possible causal polymorphisms within separate fine-mapping populations.

Conversion of a linked or causal polymorphism into an assay (for example, with kompetitive allele-specific PCR, He et al. (2014)), allows for the prediction of unobserved lines and marker-assisted selection. This may have utility in crops like wheat where many alleles have large effects on some phenotypes. This is particularly useful for disease resistance, phenology, and quality traits. For biotrophic pathogens such as rust, identification of major resistance genes that induce a qualitative hypersensitive response allows for the prediction of resistance from markers given an understanding of the composition of rust types in the target population of environments. For necrotrophic pathogens such as Fusarium head blight where typical classes of qualitative resistance genes are not effective, genes underlying QTL (such as *Fhb1*) may still be identified and found useful in enriching germplasm for resistance. Assays underlying variation in genes that control time to flowering, and for major height genes *Rht-B1* and *Rht-D1*, allow for prediction of approximate phenotypes

without field evaluations. Markers for major polymorphisms found within the glutenin genes allows for prediction of general protein quality (Payne and Lawrence 1983), and assaying a polymorphism in *Pin-D1* distinguishes hard and soft wheats. Markers for *MYB* genes on all three chromosomes distinguish varieties into red and white market classes (Himi et al. 2011). For traits like yield that are controlled by a huge number of underlying causal polymorphisms, genotyping of one or a small set of markers is not useful in predicting phenotypes, and effects for most markers linked to causal polymorphisms are too small to be estimated well.

1.3.3 Genomic prediction as a unifying framework

Both approaches to prediction of unobserved individuals converge in the mixed model framework for prediction of phenotypes using large numbers of genome-wide markers (referred to by some as the $p > n$ case, where number of predictors p is greater than the number of individuals n). In this case, the phenotype of unobserved individuals may be predicted given phenotypic data collected on related individuals and data on hundreds to thousands of genome-wide, generally neutral markers collected on both sets of individuals (Whittaker et al. 2000; Meuwissen et al. 2001). From the marker framework, a penalized estimate of marker effects may be computed by marching down each chromosome, and taking the difference in phenotype between the two classes distinguished by each marker with a shrinkage factor λ:

$$\hat{\mathbf{u}} = (\mathbf{M}'\mathbf{M} + \lambda\mathbf{I})^{-1}\mathbf{M}'(\mathbf{y} - \bar{y})$$

This formula is equivalent to that given for estimates of breeding values assuming zero covariance between relatives given in section 1.3.1, except that the numeric marker matrix **M** is used instead of the design matrix for individuals **Z**. An additional feature of the inclusion of the λ penalty is that it allows estimations for all markers without generating singularities, even when the total number of markers exceeds the number of individuals. The values of unobserved individuals are computed from the multiplication of the vector of marker effects by the matrix of marker values for each individual ($\hat{Y}_i = \hat{\mu} + \sum_{j=1}^{n} x_{ij}\hat{u}_j$, where x_{ij} is the allele dosage of marker j in individual i, and $\hat{u}_j$ is the estimated coefficient of the marker effect), to calculate a sum of effects given the markers of each individual. This

method is referred to as ridge regression, as the multiplication of each estimate by the "ridge" (λ) penalty allows for fitting of many more marker effects than there are individuals.

In contrast, phenotypes of unobserved lines may be predicted using the same set of markers by using them to estimate relatedness between lines, and using the same model as in prediction of phenotypes using pedigree information (GBLUP). The mixed model equations are solved equivalently, except that we replace the inverse of the kinship matrix containing estimates derived from pedigree information $\mathbf{A^{-1}}$ with a matrix containing realized relationships estimated from marker data $\mathbf{G^{-1}}$ (Bernardo 1994). In the case of a biparental population where allele frequencies in the total population are 0.5, the relationships can be calculated by multiplying the matrices and taking the proportion of markers shared between two genotypes ($\mathbf{G} = \frac{\mathbf{MM'}}{n}$ for n markers in (-1,0,1) coded marker matrix **M**). In a diverse population where allele frequency varies, VanRaden (2008) developed a modified method that lends more weight to rarer alleles in the estimation of kinship. First, a vector p of allele frequencies for each the marker matrix is calculated, and used to adjust **M** as $\mathbf{Z} = \mathbf{M}$ - **P**, where **P** is a matrix consisting of the vector p repeated rowwise to match the dimensions of **M**. Realized relationships are then calculated using **Z** similarly as before as:

$$\mathbf{G} = \frac{\mathbf{ZZ'}}{2\Sigma p(1-p)}$$

With values scaled by the twice the sum of the frequency of the major and minor alleles instead of the overall number of markers. While the prediction of individuals from observed relatives versus the summation of marker effects are conceptually different, the GBLUP and rr-BLUP methods use the same data as inputs into the same types of models to get equivalent results. They differ only in whether effects for the "long" side of the marker-genotype matrix are estimated by $\hat{\mathbf{u}}$ and then summed across genotypes, or if $\hat{\mathbf{u}}$ computes estimates for the genotypes directly by multiplying the matrices on their "short" side. As it is computationally simpler, GBLUP is typically used for practical prediction of breeding values (Endelman 2011), and can be extended to include individuals for which pedigree but not genotype data is available (Legarra et al. 2009).

Alternate parameterizations of the marker-based rr-BLUP model can accommodate different genetic architectures than the intesimal model assumed by rr-BLUP. Both Bayes B and Bayesian LASSO models accommodate markers with larger effects size by assuming

non-gaussian prior distributions of marker effects and estimating effects through the Gibbs sampler Markov chain Monte Carlo method (Pérez and De Los Campos 2014). Like ridge regression, both models take the form $\mathbf{y} \sim \mu + \mathbf{Zu} + \mathbf{e}$, where $\mathbf{Z}$ is a design matrix of markers and $\mathbf{u}$ is a vector of random marker effects. In the BayesB model, a certain proportion of markers given by the prior probability π $(u_i|\sigma_i^2, \pi)$ are assumed to have an effect size of 0, with the remainder having effects following a scaled-t distribution (Meuwissen et al. 2001). Bayesian LASSO assumes that marker effects are distributed according to a double exponential (laplace; λ) distribution that assumes a greater frequency of marker effects near zero and in the tails (Legarra et al. 2011). Other "bayesian alphabet" models include BayesA, which shares BayesB's scaled-t distribution but lacks a "spike" of markers with no effect, and BayesC, which swaps BayesB's scaled-t distribution for a gaussian distribution. As with rr-BLUP, predictions of genotypic values are obtained by summing marker effects. Both of these models may show a substantial improvement over GBLUP and rr-BLUP in populations and for traits where major genes are associated with a substantial portion of additive genetic variation. So-called "deep learning" methods that use more complicated models can be used for prediction as well, but produce similar or equivalent results to GBLUP models, in part because they likely use marker data to estimate relationships between relatives "under the hood" (Ubbens et al. 2021).

1.4 Yield components

1.4.1 The Yield Component Framework

For most wheat breeding programs, the primary focus is on improving grain yield. The nature of yield makes it challenging to characterize genetically. The final yield of a plot of wheat reflects the whole history of genetic and environmental processes created and experienced by plants within that plot, making it in some respects analogous to overall fitness. Because of this, causal variants that have some effect on any given phenotype will have at least some small effect on yield. Because of the overwhelming number of loci contributing to yield variation, genomic selection models and standard breeding practices that operate under the assumptions of the infinitesimal model generally work well for wheat yield (see Sarinelli et al. (2019)). At the same time, sequence variation affecting major developmental genes controlling both phenology and spike development in wheat can lead

to large yield effects, but often in an environment-specific manner (Addison et al. 2016; Lozada et al. 2019; Dreisigacker et al. 2021). While treating yield as a fully infinitesimal trait can and has worked well to move population means, an understanding of these major variants and their interaction with environmental conditions to generate yield variation may improve our understanding of yield variation and our ability to target varieties for specific environments.

Geneticists and breeders have made studying the bases of yield variation more tractable by studying simpler component traits that have a direct causal relationship with overall grain yield (Brinton and Uauy 2019). Yield component traits relate to mathematical components of yield that may differ in their genetic bases. Total yield is treated as the direct product of spike number per square meter and per-spike yield. Per-spike yield (SY) results from the multiplication of kernel number (KN) and kernel size (known as thousand grain weight, TGW), which can be treated respectively as products of kernels per spikelet (KpS) and spikelets per spike (SpS), and of kernel density (known as test weight, TW), kernel length (KL), and width (KW) squared. Variation in these traits have different environmental and genetic bases. As an example, spikelets per spike is determined early on in the development of the wheat spike prior to heading, influenced by genetic variants controlling the duration of spike development and the environmental conditions experience by plants prior to heading (Boden et al. 2015). On the other hand, kernel width is under the control of factors that include genetic variants that alter cell size in the grain, the amount of carbohydrates allocated to those grains, and the environmental conditions during grain fill that are generated by plants' previous decisions on developmental rate. As spikelets per spike is already fixed when kernel width is still plastic, it can have a causal effect on that subsequent trait – for example, if plants are source-limited, an increase in spikelets per spike and thus total kernel number will decrease the mean width of each individual kernel. Each of these traits have different bases at the intersection of complex sets of environmental and genetic processes that are separate physiologically into different tissues, and temporally into different developmental stages and climatic conditions. Study of the processes contributing to variation in these individual components may allow geneticists to better understand the overall relationship between the histories of plants and their environments, and yield. As one moves further "downstream" to more and more specific components of yield, it is presumed that these individuals components are under the influence of a smaller number of causal environmental and genetic effects.

The study of these components of yield originated with John E. Grafius, who considered primary yield components as a "geometry of breeding" (Grafius 1956, 1964). In this model, yield relates to the volume of a rectangular prism with scalar sides representing values for spikes per square meter, kernel number per spike, and kernel weight. If two sides are held constant, increasing the third will increase the volume of the prism and therefore yield, but an increase of one or more component associated with a decrease in another may not necessarily lead to an increase in the total volume. In many papers presenting on mapping of QTL for a yield component trait, it is presumed that increasing that single component will increase overall yield, and serves to justify the research into the identification of those QTL. This relies on an assumption that modern wheat cultivars are "sink" limited (*i.e.*, a plant accumulates more carbon than it is able to partition into developing grain, because a genotype has some genetically determined volume of grain proportional to the product of the yield components) (Borrás et al. 2004). However, typically a compensatory effect is observed where increasing grain number concomitantly decreases grain size (*e.g.* Knott and Talukdar 1971). This effect may be due to "source" limitation – there only being so much carbon to go around – or it may reflect pleiotropic relationships of underlying QTL generated by the genetic pathways they affect. Depending on the pleiotropic relationships, a yield component QTL may generate a positive yield effect in some environments, and a negative effect in others.

1.4.2 Exploring Yield Variation Effects through Yield Components

Understanding how major yield component QTL alter yield in an environment-specific manner may provide insights into breeding lines for given environments, even if straightforward use in MAS is not practical. Ultimately, characterization of underlying variants opens the door to genomic selection models that treat different sets of SNPs differently, and may be useful for predicting performance under certain conditions, or separating out the "causal" portion of the genome from sections with smaller or no yield effects. In order to do this, major yield component genes may be characterized. A gene underlying a major QTL on 7AL regulating the number of spikelets per spike independent of heading date was identified as WAPO-A1 (*WHEAT ABERRANT PANICLE ORGANIZER 1*), an ortholog of *Arabidopsis LFY* and *UFO* (Kuzay et al. 2019, 2022). A major QTL on 6A altering grain width and size was mapped to *TaGW2*, a negative regulator of grain weight (Su et al. 2011; Wang et al. 2018).

A principal goal of modern genetics research is to understand how individual genes contribute to differences in grain yield. Their use in the improvement of yield will require further research into the environment-specific mechanisms through which genetics and physiology interact as plants develop their reproductive organs, accumulate carbon, and allocate that carbon into developing grain. In this research, we try to characterize the genetic architecture of phenology and yield component traits within a single population, and attempt to understand how QTL generate spike yield variation through intermediate traits. We identify a single yield component, the *B1* awn suppressor QTL, and attempt to characterize it on a molecular level, and then understand how it contributes to yield differences in an environmentally specific manner. We hope this might give insight into methods for the prediction of yield with information on many QTL, especially when considering variants in major heading date genes *Vrn1* and *Ppd1*, and major yield component QTL like *WAPO-A1* and *TaGW2*.

References

Addison, C. K., Mason, R. E., Brown-Guedira, G., Guedira, M., Hao, Y., Miller, R. G., Subramanian, N., Lozada, D. N., Acuna, A., Arguello, M. N., Johnson, J. W., Ibrahim, A. M., Sutton, R., and Harrison, S. A. (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.

Åkerberg, E. (1986). Nilsson-Ehle and the development of plant breeding at Svalöf during the period 1900-1915. *Hereditas*, 105(1):1–5.

Åkerman, A. and J. MacKey (1949). Efforts to increase yield in spring wheat. II. Crosses between spring and winter wheats: description of Svalöf Ella spring wheat. *Sveriges Utsädesfor. tidskr.*, 59:105–117.

Allard, R. W. and Bradshaw, A. D. (1964). Implications of Genotype-Environmental Interactions in Applied Plant Breeding 1. *Crop Science*, 4(5):503–508.

Anderson, G. R., Papa, D., Peng, J., Tahir, M., and Lapitan, N. L. V. (2003). Genetic mapping of Dn7, a rye gene conferring resistance to the Russian wheat aphid in wheat. *Theoretical and Applied Genetics*, 107(7):1297–1303.

Anugrahwati, D. R., Shepherd, K. W., Verlin, D. C., Zhang, P., Ghader Mirzaghaderi, Walker, E., Francki, M. G., and Dundas, I. S. (2008). Isolation of wheat–rye 1RS recombinants that break the linkage between the stem rust resistance gene SrR and secalin. *Genome*, 51(5):341–349.

Appels, R., Eversole, K., Feuillet, C., Keller, B., Rogers, J., Stein, N., Pozniak, C. J., Choulet, F., Distelfeld, A., Poland, J., Ronen, G., Barad, O., Baruch, K., Keeble-Gagnère, G., Mascher, M., Ben-Zvi, G., Josselin, A. A., Himmelbach, A., Balfourier, F., Gutierrez-Gonzalez, J., Hayden, M., Koh, C. S., Muehlbauer, G., Pasam, R. K., Paux, E., Rigault, P., Tibbits, J., Tiwari, V., Spannagl, M., Lang, D., Gundlach, H., Haberer, G., Mayer, K. F., Ormanbekova, D., Prade, V., Wicker, T., Swarbreck, D., Rimbert, H., Felder, M., Guilhot, N., Kaithakottil, G., Keilwagen, J., Leroy, P., Lux, T., Twardziok, S., Venturini, L., Juhasz, A., Abrouk, M., Fischer, I., Uauy, C., Borrill, P., Ramirez-Gonzalez, R. H., Arnaud, D., Chalabi, S., Chalhoub, B., Cory, A., Datla, R., Davey, M. W., Jacobs, J., Robinson, S. J., Steuernagel, B., Van Ex, F., Wulff, B. B., Benhamed, M., Bendahmane, A., Concia, L., Latrasse, D., Alaux, M., Bartoš, J., Bellec, A., Berges, H., Doležel, J., Frenkel, Z., Gill, B., Korol, A., Letellier, T., Olsen, O. A., Šimková, H., Singh, K., Valárik, M., Van Der Vossen, E., Vautrin, S., Weining, S., Fahima, T., Glikson, V., Raats, D., Toegelová, H., Vrána, J., Sourdille, P., Darrier, B., Barabaschi, D., Cattivelli, L., Hernandez, P., Galvez, S., Budak, H., Jones, J. D., Witek, K., Yu, G., Small, I., Melonek, J., Zhou, R., Belova, T., Kanyuka, K., King, R., Nilsen, K., Walkowiak, S., Cuthbert, R., Knox, R., Wiebe, K., Xiang, D., Rohde, A., Golds, T., Čížkova, J., Akpinar, B. A., Biyiklioglu, S., Gao, L., N'Daiye, A., Číhalíková, J., Kubaláková, M., Šafář, J., Alfama, F., Adam-Blondon, A. F., Flores, R., Guerche, C., Loaec, M., Quesneville, H., Sharpe, A. G., Condie, J., Ens, J.,

Maclachlan, R., Tan, Y., Alberti, A., Aury, J. M., Barbe, V., Couloux, A., Cruaud, C., Labadie, K., Mangenot, S., Wincker, P., Kaur, G., Luo, M., Sehgal, S., Chhuneja, P., Gupta, O. P., Jindal, S., Kaur, P., Malik, P., Sharma, P., Yadav, B., Singh, N. K., Khurana, J. P., Chaudhary, C., Khurana, P., Kumar, V., Mahato, A., Mathur, S., Sevanthi, A., Sharma, N., Tomar, R. S., Holušová, K., Plíhal, O., Clark, M. D., Heavens, D., Kettleborough, G., Wright, J., Balcárková, B., Hu, Y., Ravin, N., Skryabin, K., Beletsky, A., Kadnikov, V., Mardanov, A., Nesterov, M., Rakitin, A., Sergeeva, E., Kanamori, H., Katagiri, S., Kobayashi, F., Nasuda, S., Tanaka, T., Wu, J., Cattonaro, F., Jiumeng, M., Kugler, K., Pfeifer, M., Sandve, S., Xun, X., Zhan, B., Batley, J., Bayer, P. E., Edwards, D., Hayashi, S., Tulpová, Z., Visendi, P., Cui, L., Du, X., Feng, K., Nie, X., Tong, W., and Wang, L. (2018). Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science*, 361(6403).

Ball, C. R. (1930). The History of American Wheat Improvement. *Agricultural History*, 4(2):48–71.

Beales, J., Turner, A., Griffiths, S., Snape, J. W., and 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.

Beavis, W. D. (1998). QTL Analyses: Power, Precision, and Accuracy. In Paterson, A., editor, *Molecular dissection of complex traits*, pages 145–162. CRC Press.

Bernardo, R. (1994). Prediction of Maize Single-Cross Performance Using RFLPs and Information from Related Hybrids. *Crop Science*, 34(1):20–25.

Bernardo, R. (2010). *Breeding for Quantitative Traits in Plants*. Stemma Press, Woodburry, Minnesota, 2 edition.

Bhullar, S. and Jenner, C. (1985). Differential Responses to High Temperatures of Starch and Nitrogen Accumulation in the Grain of Four Cultivars of Wheat. *Functional Plant Biology*, 12(4):363.

Boden, S. A., Cavanagh, C., Cullis, B. R., Ramm, K., Greenwood, J., Jean Finnegan, E., Trevaskis, B., and Swain, S. M. (2015). Ppd-1 is a key regulator of inflorescence architecture and paired spikelet development in wheat. *Nature Plants*, 1.

Bonjean, A. P. and Angus, W. J. (2001). *The World Wheat Book: A History of Wheat Breeding*.

Borojevic, K. and Borojevic, K. (2005). The Transfer and History of "Reduced Height Genes" (Rht) in Wheat from Japan to Europe. *Journal of Heredity*, 96(4):455–459.

Borrás, L., Slafer, G. A., and Otegui, M. E. (2004). Seed dry weight response to source-sink manipulations in wheat, maize and soybean: A quantitative reappraisal. *Field Crops Research*, 86(2-3):131–146.

Botstein, D., White, R., Skolnick, M., and Davis, R. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics*, 32(3):314–331.

Boyles, R. E., Marshall, D. S., and Bockelman, H. E. (2019). Yield Data from the Uniform Southern Soft Red Winter Wheat Nursery Emphasize Importance of Selection Location and Environment for Cultivar Development. *Crop Science*, 59(5):1887.

Brigham, A. P. (1910). The Development of Wheat Culture in North America. *The Geographical Journal*, 35(1).

Brinton, J. and Uauy, C. (2019). A reductionist approach to dissecting grain weight and yield in wheat. *Journal of Integrative Plant Biology*, 61(3):337–358.

Broman, K. W., Wu, H., Sen, S., and Churchill, G. A. (2003). R/qtl: QTL mapping in experimental crosses. *Bioinformatics*, 19(7):889–890.

Brown-Guedira, G. L. (1996). Evaluation of a Collection of Wild Timopheevi Wheat for Resistance to Disease and Arthropod Pests. *Plant Disease*, 80(8):928.

Cao, L., Hayashi, K., Tokui, M., Mori, M., Miura, H., and Onishi, K. (2016). Detection of QTLs for traits associated with pre-harvest sprouting resistance in bread wheat (<i>Triticum aestivum</i> L.). *Breeding Science*, 66(2):260–270.

Carleton, M. A. (1900). *Russian Cereals Adapted For Cultivation in the United States*. US. Department of Agriculture, Division of Botany.

Carleton, M. A. (1920). *The Small Grains*. The Macmillan Company, New York.

Cavanagh, C. R., Chao, S., Wang, S., Huang, B. E., Stephen, S., Kiani, S., Forrest, K., Saintenac, C., Brown-Guedira, G. L., Akhunova, A., See, D., Bai, G., Pumphrey, M., Tomar, L., Wong, D., Kong, S., Reynolds, M., da Silva, M. L., Bockelman, H., Talbert, L., Anderson, J. A., Dreisigacker, S., Baenziger, S., Carter, A., Korzun, V., Morrell, P. L., Dubcovsky, J., Morell, M. K., Sorrells, M. E., Hayden, M. J., and Akhunov, E. (2013). Genome-wide comparative diversity uncovers multiple targets of selection for improvement in hexaploid wheat landraces and cultivars. *Proceedings of the National Academy of Sciences*, 110(20):8057–8062.

Ceoloni, C. (1984). Transfer of a mildew resistance gene from Triticum longissimum to common wheat by induced homoeologous recombination. *Genetica Agrari*, 38(3):326–327.

Charmet, G. (2011). Wheat domestication: Lessons for the future. *Comptes Rendus Biologies*, 334(3):212–220.

Churchill, G. A. and Doerge, R. W. (1994). Empirical threshold values for quantitative trait mapping. *Genetics*, 138(3):963–971.

Comstock, R. E. and Robinson, H. (1952). Estimation of average dominance of genes. *Heterosis*, 2:494–516.

Cox, T. S., Murphy, J. P., and Rodgers, D. M. (1986). Changes in genetic diversity in the red winter wheat regions of the United States. *Proceedings of the National Academy of Sciences*, 83(15):5583–5586.

Dhaliwal, H., Singh, H., Gill, K., and Randhawa, H. (1993). *Evaluation and cataloguing of wheat germplasm for disease resistance and quality*. John Wiley and Sons.

Díaz, A., Zikhali, M., Turner, A. S., Isaac, P., and 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).

Dreisigacker, S., Burgueño, J., Pacheco, A., Molero, G., Sukumaran, S., Rivera-Amado, C., Reynolds, M., and Griffiths, S. (2021). Effect of flowering time-related genes on biomass, harvest index, and grain yield in CIMMYT elite spring bread wheat. *Biology*, 10(9).

Dubcovsky, J., Lukaszewski, A. J., Echaide, M., Antonelli, E. F., and Porter, D. R. (1998). Molecular Characterization of Two Triticum speltoides Interstitial Translocations Carrying Leaf Rust and Greenbug Resistance Genes. *Crop Science*, 38(6):1655–1660.

Dvořák, J. (1977). Transfer of leaf rust resistance from Aegilops speltoides to Triticum aestivum. *Canadian Journal of Genetics and Cytology*, 19(1):133–141.

Edwards, M. D., Stuber, C. W., and Wendel, J. F. (1987). Molecular-Marker-Facilitated Investigations of Quantitative-Trait Loci in Maize. I. Numbers, Genomic Distribution and Types of Gene Action. *Genetics*, 116(1):113–125.

Endelman, J. B. (2011). Ridge Regression and Other Kernels for Genomic Selection with R Package rrBLUP. *The Plant Genome*, 4(3):250–255.

Fain, P. R. (1978). Characteristics of simple sibship variance tests for the detection of major loci and application to height, weight and spatial performance. *Annals of Human Genetics*, 42(1):109–120.

Falconer, D. and MacKay, T. (1995). *Introduction to Quantitative Genetics*. Longman, 4 edition.

Faris, J. D. (2014). Wheat Domestication: Key to Agricultural Revolutions Past and Future. In *Genomics of Plant Genetic Resources*, pages 439–464. Springer Netherlands, Dordrecht.

Faris, J. D. and Gill, B. S. (2002). Genomic targeting and high-resolution mapping of the domestication gene Q in wheat. *Genome*, 45(4):706–718.

Faris, J. D., Xu, S. S., Cai, X., Friesen, T. L., and Jin, Y. (2008). Molecular and cytogenetic characterization of a durum wheat–Aegilops speltoides chromosome translocation conferring resistance to stem rust. *Chromosome Research*, 16(8):1097–1105.

Fisher, R. (1918). The Correlation between Relatives on the Supposition of Mendelian Inheritance. *Transactions of the Royal Society of Edinburgh*.

Friebe, B., Jiang, J., Raupp, W. J., McIntosh, R. A., and Gill, B. S. (1996). Characterization of wheat-alien translocations conferring resistance to diseases and pests: current status. *Euphytica*, 91(1):59–87.

Fu, D., Szűcs, P., Yan, L., Helguera, M., Skinner, J. S., Von Zitzewitz, J., Hayes, P. M., and Dubcovsky, J. (2005). Large deletions within the first intron in VRN-1 are associated with spring growth habit in barley and wheat. *Molecular Genetics and Genomics*, 273(1):54–65.

Gaire, R., Arruda, M. P., Mohammadi, M., Brown-Guedira, G., Kolb, F. L., and Rutkoski, J. (2022). Multi-trait genomic selection can increase selection accuracy for deoxynivalenol accumulation resulting from fusarium head blight in wheat. *The Plant Genome*.

Gauley, A. and Boden, S. A. (2019). Genetic pathways controlling inflorescence architecture and development in wheat and barley. *Journal of Integrative Plant Biology*, 61(3):296–309.

Gerechter-Amitai, Z. K., van Silfhout, C. H., Grama, A., and Kleitman, F. (1989). Yr15 — a new gene for resistance to Puccinia striiformis in Triticum dicoccoides sel. G-25. *Euphytica*, 43(1-2):187–190.

Gibson, G. (2012). Rare and common variants: twenty arguments. *Nature Reviews Genetics*, 13(2):135–145.

Grafius, J. E. (1956). Components of Yield in Oats: A Geometrical Interpretation ¹. *Agronomy Journal*, 48(9):419–423.

Grafius, J. E. (1964). A Geometry for Plant Breeding. *Crop Science*, 4(3):241–246.

Guedira, M., Brown-Guedira, G., Van Sanford, D., Sneller, C., Souza, E., and Marshall, D. (2010). Distribution of Rht Genes in Modern and Historic Winter Wheat Cultivars from the Eastern and Central USA. *Crop Science*, 50(5):1811–1822.

Guedira, M., Maloney, P., Xiong, M., Petersen, S., Murphy, J. P., Marshall, D., Johnson, J., Harrison, S., and Brown-Guedira, G. (2014). Vernalization duration requirement in soft winter wheat is associated with variation at the VRN-B1 locus. *Crop Science*, 54(5):1960–1971.

Guedira, M. and Paulsen, G. M. (2002). Accumulation of starch in wheat grain under different shoot/root temperatures during maturation. *Functional Plant Biology*, 29(4):495.

Guedira, M., Xiong, M., Hao, Y. F., Johnson, J., Harrison, S., Marshall, D., and Brown-Guedira, G. (2016). Heading date QTL in winter wheat (Triticum aestivum L.) coincide with major developmental genes VERNALIZATION1 and PHOTOPERIOD1. *PLoS ONE*, 11(5).

Gyarfas, J. (1978). *Transference of disease resistance from Triticum timopheevii to Triticum aestivum*. PhD thesis, University of Sydney, Sydney, NSW, Australi.

Haley, C. S. and Knott, S. A. (1992). A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity*, 69(4):315–324.

Hammond, K. and James, J. (1970). Genes of Large Effect and the Shape of the Distribution of a Quantitative Character. *Australian Journal of Biological Sciences*, 23(4):867.

Hawker, J. and Jenner, C. (1993). High Temperature Affects the Activity of Enzymes in the Committed Pathway of Starch Synthesis in Developing Wheat Endosperm. *Functional Plant Biology*, 20(2):197.

Hayes, W. M. (1899). Wheat: Variety, Breeding, Cultivation. Technical report, University of Minnesota Agricultural Experiment Station.

He, C., Holme, J., and Anthony, J. (2014). SNP Genotyping: The KASP Assay.

Helmsjrgensen, J. and Jensen, C. J. (1973). Gene Pm6 for resistance to powdery mildew in wheat. *Euphytica*, 22(2):423–423.

Henderson, C. (1976). A Simple Method for Computing the Inverse of a Numerator Relationship Matrix Used in Prediction of Breeding Values. *Biometrics Biometrics*, 32(1):69–83.

Henderson, C. R. (1975). Best Linear Unbiased Estimation and Prediction under a Selection Model. Technical Report 2.

Hickman, J. F. (1889). Experiments in wheat seeding; Comparative test of varieties of wheat. Technical report, Ohio Agricultural Experiment Station.

Himi, E., Maekawa, M., Miura, H., and Noda, K. (2011). Development of PCR markers for Tamyb10 related to R-1, red grain color gene in wheat. *Theoretical and Applied Genetics*, 122(8):1561–1576.

Inazuka, G. (1971). Norin 10, a Japanese semi-dwarf wheat variety. *Wheat Information Service*, 32.

Isik, F., Holland, J., and Maltecca, C. (2017). *Genetic Data Analysis for Plant and Animal Breeding*. Springer International Publishing, Cham.

Jahier, J., Tanguy, A. M., and Doussinault, G. (1989). Analysis of the level of eyespot resistance due to genes transferred to wheat from Aegilops ventricosa. *Euphytica*, 44(1-2):55–59.

Kackar, R. N. and Harville, D. A. (1981). Unbiasedness of two-stage estimation and prediction procedures for mixed linear models. *Communications in Statistics - Theory and Methods*, 10(13):1249–1261.

Kato, K., Miura, H., Akiyama, M., Kuroshima, M., and Sawada, S. (1998). RFLP mapping of the three major genes, Vrn1, Q and B1, on the long arm of chromosome 5A of wheat. Technical report.

Kerber, E. R. and Dyck, P. L. (1973). Inheritance of Stem Rust Resistance Transferred from Diploid Wheat (Triticum monococcum) to Tetraploid and Hexaploid Wheat and Chromosome Location of the Gene Involved. *Canadian Journal of Genetics and Cytology*, 15(3):397–409.

Kerber, E. R. and Dyck, P. L. (1990). Transfer to hexaploid wheat of linked genes for adult-plant leaf rust and seedling stem rust resistance from an amphiploid of Aegilops speltoides × Triticum monococcum. *Genome*, 33(4):530–537.

King, R. and Richards, R. (1984). Water uptake in relation to pre-harvest sprouting damage in wheat: ear characteristics. *Australian Journal of Agricultural Research*, 35(3):327.

Kippes, N., Guedira, M., Lin, L., Alvarez, M. A., Brown-Guedira, G. L., and Dubcovsky, J. (2018). Single nucleotide polymorphisms in a regulatory site of VRN-A1 first intron are associated with differences in vernalization requirement in winter wheat. *Molecular Genetics and Genomics*, 293(5):1231–1243.

Knott, D. and Anderson, R. (1956). The Inheritance of Rust Resistance.: I. the Inheritance of Stem Rust Resistance in Ten Varieties of Common Wheat. *Canadian Journal of Agricultural Science*.

Knott, D. R. and Talukdar, B. (1971). Increasing Seed Weight in Wheat and Its Effect on Yield, Yield Components, and Quality. *Crop Science*, 11(2):280–283.

Kolmer, J. A., Anderson, J. A., and Flor, J. M. (2010). Chromosome Location, Linkage with Simple Sequence Repeat Markers, and Leaf Rust Resistance Conditioned by Gene Lr63 in Wheat. *Crop Science*, 50(6):2392–2395.

Korte, A. and Farlow, A. (2013). The advantages and limitations of trait analysis with GWAS: a review. *Plant Methods*, 9(1):29.

Kuzay, S., Lin, H., Li, C., Chen, S., Woods, D. P., Zhang, J., Lan, T., von Korff, M., and Dubcovsky, J. (2022). WAPO-A1 is the causal gene of the 7AL QTL for spikelet number per spike in wheat. *PLOS Genetics*, 18(1):e1009747.

Kuzay, S., Xu, Y., Zhang, J., Katz, A., Pearce, S., Su, Z., Fraser, M., Anderson, J. A., Brown-Guedira, G., DeWitt, N., Peters Haugrud, A., Faris, J. D., Akhunov, E., Bai, G., and Dubcovsky, J. (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.

Lander, E. S. and Botstein, D. (1989). Mapping Mendelian Factors Underlying Quantitative Traits Using RFLP Linkage Maps. *Genetics*, 121(1):185–199.

Larkin, D. L., Mason, R. E., Moon, D. E., Holder, A. L., Ward, B. P., and Brown-Guedira, G. (2021). Predicting Fusarium Head Blight Resistance for Advanced Trials in a Soft Red Winter Wheat Breeding Program With Genomic Selection. *Frontiers in Plant Science*, 12.

Latter, B. (1965). The Response to Artificial Selection Due to Autosomal Genes of Large Effect I. Changes in Gene Frequency at an Additive Locus. *Australian Journal of Biological Sciences*, 18(3):585.

Law, C. N., Sutka, J., and Worland, A. J. (1978). A Genetic study of day-length response in wheat. *Heredity*, 41(2):185–191.

Law, C. N. and Worland, A. J. (1997). Genetic analysis of some flowering time and adaptive traits in wheat. *New Phytologist*, 137(1):19–28.

Law, C. N., Worland, A. J., and Giorgi, B. (1976). The genetic control of ear-emergence time by chromosomes 5A and 5D of wheat. *Heredity*, 36(1):49–58.

Legarra, A., Aguilar, I., and Misztal, I. (2009). A relationship matrix including full pedigree and genomic information. *Journal of Dairy Science*, 92(9):4656–4663.

Legarra, A., Robert-Granié, C., Croiseau, P., Guillaume, F., and Fritz, S. (2011). Improved Lasso for genomic selection. *Genetics Research*, 93(1):77–87.

Leonova, I. N., Laikova, L. I., Popova, O. M., Unger, O., Börner, A., and Röder, M. S. (2007). Detection of quantitative trait loci for leaf rust resistance in wheat—T. timopheevii/T. tauschii introgression lines. *Euphytica*, 155(1-2):79–86.

Levins, R. and Lewontin, R. C. (1985). Organism as Subject and Object. In *The Dialectical Biologist*, chapter 3, pages 85–106. Harvard University Press, Cambridge, Massachusetts.

Levinson, G. and Gutman, G. (1987). Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Molecular Biology and Evolution*.

Lewontin, R. C. (1957). The Adaptations of Populations to Varying Environments. *Cold Spring Harbor Symposia on Quantitative Biology*, 22(0):395–408.

Lewontin, R. C. and Kojima, K.-I. (1960). The Evolutionary Dynamics of Complex Polymorphisms. *Evolution*, 14(4):458–472.

Li, G., Yu, M., Fang, T., Cao, S., Carver, B. F., and Yan, L. (2013). Vernalization requirement duration in winter wheat is controlled by TaVRN-A1 at the protein level. *Plant Journal*, 76(5):742–753.

Li, W. and Gill, B. S. (2006). Multiple genetic pathways for seed shattering in the grasses. *Functional & Integrative Genomics*, 6(4):300–309.

Lipka, A. E., Tian, F., Wang, Q., Peiffer, J., Li, M., Bradbury, P. J., Gore, M. A., Buckler, E. S., and Zhang, Z. (2012). GAPIT: Genome association and prediction integrated tool. *Bioinformatics*, 28(18):2397–2399.

Lozada, D. N., Mason, R. E., Sarinelli, J. M., and Brown-Guedira, G. (2019). Accuracy of genomic selection for grain yield and agronomic traits in soft red winter wheat. *BMC Genetics*.

Lumpkin, T. A. (2015). How a Gene from Japan Revolutionized the World of Wheat: CIMMYT's Quest for Combining Genes to Mitigate Threats to Global Food Security. In Ogihara, Y., Takumi, S., and Handa, H., editors, *Advances in Wheat Genetics: From Genome to Field*, pages 13–20. Springer Japan, Tokyo.

Lynch, M. and Walsh, B. (1998). *Genetics and Analysis of Quantitative Traits*. Sinauer Associates, Inc.

Ma, Z.-Q., Gill, B. S., Sorrells, M. E., and Tanksley, S. D. (1993). RELP markers linked to two Hessian fly-resistance genes in wheat (Triticum aestivum L.) from Triticum tauschii (coss.) Schmal. *Theoretical and Applied Genetics*, 85-85(6-7):750–754.

Malihipour, A., Gilbert, J., Fedak, G., Brûlé-Babel, A., and Cao, W. (2017). Mapping the A Genome for QTL Conditioning Resistance to Fusarium Head Blight in a Wheat Population with Triticum timopheevii Background. *Plant Disease*, 101(1):11–19.

Marais, G. F., McCallum, B., Snyman, J. E., Pretorius, Z. A., and Marais, A. S. (2005a). Leaf rust and stripe rust resistance genes Lr54 and Yr37 transferred to wheat from Aegilops kotschyi. *Plant Breeding*, 124(6):538–541.

Marais, G. F., Pretorius, Z. A., Wellings, C. R., McCallum, B., and Marais, A. S. (2005b). Leaf rust and stripe rust resistance genes transferred to common wheat from Triticum dicoccoides. *Euphytica*, 143(1-2):115–123.

Maxwell, J. J., Lyerly, J. H., Cowger, C., Marshall, D., Brown-Guedira, G., and Murphy, J. P. (2009). MlAG12: a Triticum timopheevii-derived powdery mildew resistance gene in common wheat on chromosome 7AL. *Theoretical and Applied Genetics*, 119(8):1489–1495.

McClellan, J. and King, M.-C. (2010). Genetic Heterogeneity in Human Disease. *Cell*, 141(2):210–217.

McFadden, E. (1930). A successful transfer of emmer characters to vulgare wheat. *Journal of the American Society of Agronomy*, 22:1020–1034.

McIntosh, R., Dyck, P., The, T., and Cusick, J. (1984). Cytogenetical studies in wheat XIII. Sr35, a third gene from Triticum monococcum for resistance to Puccinia graminis tritici . *Zeitschrift fur Pflanzenzuchtung*.

McIntosh, R. and Gyarfas, J. (1971). Triticum timopheevi as a source of resistance to wheat stem rust. *Z Pflanzenzucht*.

McKendry, A. L., Tague, D. N., and Miskin, K. E. (1996). Effect of 1BL.1RS on Agronomic Performance of Soft Red Winter Wheat. *Crop Science*, 36(4):844–847.

Meuwissen, T. H. E., Hayes, B. J., and Goddard, M. E. (2001). Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. *Genetics*, 157(4):1819–1829.

Miller, T., Reader, S., Ainsworth, C., and Summers, R. (1988). The introduction of a major gene for resistance to powdery mildew of wheat, Erysiphe graminis f. sp. tritici, from Aegilops speltoides into wheat, Triticum aestivum. In *Proc. Conference Cereal Sect. EUCARPIA. Wageningen*, pages 179–183.

Miranda, L. M., Murphy, J. P., Marshall, D., Cowger, C., and Leath, S. (2007). Chromosomal location of Pm35, a novel Aegilops tauschii derived powdery mildew resistance gene introgressed into common wheat (Triticum aestivum L.). *Theoretical and Applied Genetics*, 114(8):1451–1456.

Miranda, L. M., Murphy, J. P., Marshall, D., and Leath, S. (2006). Pm34: a new powdery mildew resistance gene transferred from Aegilops tauschii Coss. to common wheat (Triticum aestivum L.). *Theoretical and Applied Genetics*, 113(8):1497–1504.

Morris, C. F. (2002). Plant Molecular Biology. *Plant Molecular Biology*, 48(5/6):633–647.

Morton, N. E. (1955). Sequential tests for the detection of linkage. *American Journal of Human Genetics*, 7(3):277–318.

Murphy, J., Leath, S., Huynh, D., Navarro, R., and Shi, A. (1999). Registration of NC97BGTD7 and NC97BGTD8 wheat germplasms resistant to powdery mildew. *Crop science*, 39(3):884–884.

Murphy, J. P., Cox, T. S., and Rodgers, D. M. (1986). Cluster Analysis of Red Winter Wheat Cultivars Based upon Coefficients of Parentage ¹. *Crop Science*, 26(4):672–676.

Murphy, J. P., Leath, S., Huynh, D., Navarro, R. A., and Shi, A. (1998). Registration of NC96BGTD1, NC96BGTD2, and NC96BGTD3 Wheat Germplasm Resistant to Powdery Mildew. *Crop Science*, 38(2):570–571.

Nalam, V. J., Vales, M. I., Watson, C. J. W., Kianian, S. F., and Riera-Lizarazu, O. (2006). Map-based analysis of genes affecting the brittle rachis character in tetraploid wheat (Triticum turgidum L.). *Theoretical and Applied Genetics*, 112(2):373–381.

Nishida, H., Yoshida, T., Kawakami, K., Fujita, M., Long, B., Akashi, Y., Laurie, D. A., and Kato, K. (2013). Structural variation in the 5 upstream region of photoperiod-insensitive alleles Ppd-A1a and Ppd-B1a identified in hexaploid wheat (Triticum aestivum L.), and their effect on heading time. *Molecular Breeding*, 31(1):27–37.

Olson, E. L., Rouse, M. N., Pumphrey, M. O., Bowden, R. L., Gill, B. S., and Poland, J. A. (2013). Introgression of stem rust resistance genes SrTA10187 and SrTA10171 from Aegilops tauschii to wheat. *Theoretical and Applied Genetics*, 126(10):2477–2484.

Payne, P. I. and Lawrence, G. J. (1983). Catalogue of alleles for the complex gene loci, Glu-A1, Glu-B1, and Glu-D1 which code for high-molecular-weight subunits of glutenin in hexaploid wheat. *Cereal Research Communications Cereal Research Communications*, 11(1):29–35.

Peng, J., Richards, D. E., Hartley, N. M., Murphy, G. P., Devos, K. M., Flintham, J. E., Beales, J., Fish, L. J., Worland, A. J., Pelica, F., Sudhakar, D., Christou, P., Snape, J. W., Gale, M. D., and Harberd, N. P. (1999). 'Green revolution' genes encode mutant gibberellin response modulators. *Nature*, 400(6741):256–261.

Pérez, P. and De Los Campos, G. (2014). Genome-Wide Regression and Prediction with the BGLR Statistical Package. *Genetics*, 198:483.

Perugini, L. D., Murphy, J. P., Marshall, D., and Brown-Guedira, G. (2008). Pm37, a new broadly effective powdery mildew resistance gene from Triticum timopheevii. *Theoretical and Applied Genetics*, 116(3):417–425.

Petersen, S., Lyerly, J. H., Worthington, M. L., Parks, W. R., Cowger, C., Marshall, D. S., Brown-Guedira, G., and Murphy, J. P. (2015). Mapping of powdery mildew resistance gene Pm53 introgressed from Aegilops speltoides into soft red winter wheat. *Theoretical and Applied Genetics*, 128(2):303–312.

Peusha, H., Enno, T., and Priilinn, O. (2004). Chromosomal Location of Powdery Mildew Resistance Genes and Cytogenetic Analysis of Meiosis in Common Wheat Cultivar Meri. *Hereditas*, 132(1):29–34.

Poland, J., Endelman, J., Dawson, J., Rutkoski, J., Wu, S., Manes, Y., Dreisigacker, S., Crossa, J., Sánchez-Villeda, H., Sorrells, M., and Jannink, J. L. (2012). Genomic selection in wheat breeding using genotyping-by-sequencing. *Plant Genome*, 5(3):103–113.

Pugsley, A. (1971). A genetic analysis of the spring-winter habit of growth in wheat. *Australian Journal of Agricultural Research*, 22(1).

Rebetzke, G. J. and Richards, R. A. (2000). Gibberellic acid-sensitive dwarfing genes reduce plant height to increase kernel number and grain yield of wheat. *Australian Journal of Agricultural Research*, 51(2):235–245.

Robinson, G. (1991). That BLUP Is a Good Thing: The Estimatino of Random Effects. *Statistical Science*, 6(1):15–32.

Sarinelli, J. M., Murphy, J. P., Tyagi, P., Holland, J. B., Johnson, J. W., Mergoum, M., Mason, R. E., Babar, A., Harrison, S., Sutton, R., Griffey, C. A., and Brown-Guedira, G. (2019).

Training population selection and use of fixed effects to optimize genomic predictions in a historical USA winter wheat panel. *Theoretical and Applied Genetics*, 132(4):1247–1261.

Sax, K. (1923). The Association of Size Differences with Seed-Coat Pattern and Pigmentation in Phaseolus vulgaris. *Genetics*, 8(6):552–560.

Schön, C. C., Friedrich Utz, H., Groh, S., Truberg, B., Openshaw, S., and Melchinger, A. E. (2004). Quantitative Trait Locus Mapping Based on Resampling in a Vast Maize Testcross Experiment and Its Relevance to Quantitative Genetics for Complex Traits. *Genetics*.

Sears, E. (1956). The transfer of leaf-rust resistance from Aegilops umbellulata to wheat. In *Genetics in plant breeding. Brook-haven Symposia in Biology*, pages 1–22.

Sears, E. R. (1944). Cytogenetic Studies with Polyploid Species of Wheat. II. Additional Chromosomal Aberrations in Triticum vulgare. *Genetics*, 29(3):232–246.

Sears, E. R. (1953). Nullisomic Analysis in Common Wheat. *The American Naturalist*, 87(835):245–252.

Sears, E. R. (1954). *The aneuploids of common wheat*. University of Missouri, College of Agriculture, Agricultural Experiment Station.

Simons, K. J., Fellers, J. P., Trick, H. N., Zhang, Z., Tai, Y. S., Gill, B. S., and Faris, J. D. (2006). Molecular characterization of the major wheat domestication gene Q. *Genetics*, 172(1):547–555.

Slaughter, D. C., Norris, K. H., and Hruschka, W. R. (1992). Quality and Classification of Hard Red Wheat. *Cereal Chemistry*, 69(4):428–432.

Snape, J. W., Law, C. N., Parker, B. B., and Worland, A. J. (1985). Genetical analysis of chromosome 5A of wheat and its influence on important agronomic characters. *Theoretical and Applied Genetics*, 71(3):518–526.

Sood, S., Kuraparthy, V., Bai, G., and Gill, B. S. (2009). The major threshability genes soft glume (sog) and tenacious glume (Tg), of diploid and polyploid wheat, trace their origin to independent mutations at non-orthologous loci. *Theoretical and Applied Genetics*, 119(2):341–351.

Stuber, C. W., Edwards, M. D., and Wendel, J. F. (1987). Molecular Marker-Facilitated Investigations of Quantitative Trait Loci in Maize. II. Factors Influencing Yield and its Component Traits 1. *Crop Science*, 27(4):639–648.

Stuber, C. W. and Moll, R. H. (1972). Frequency Changes of Isozyme Alleles in a Selection Experiment for Grain Yield in Maize (Zea mays L.) 1. *Crop Science*, 12(3):337–340.

Sturtevant, A. H. (1913). The linear arrangement of six sex-linked factors in Drosophila, as shown by their mode of association. *Journal of experimental zoology*, 14(1):43–59.

Su, Z., Hao, C., Wang, L., Dong, Y., and Zhang, X. (2011). Identification and development of a functional marker of TaGW2 associated with grain weight in bread wheat (Triticum aestivum L.). *Theoretical and Applied Genetics*, 122(1):211–223.

Tadesse, W., Sanchez-Garcia, M., Assefa, S. G., Amri, A., Bishaw, Z., Ogbonnaya, F. C., and Baum, M. (2019). Genetic Gains in Wheat Breeding and Its Role in Feeding the World. *Crop Breeding, Genetics and Genomics*.

Turkheimer, E. (2012). Genome Wide Association Studies of Behavior are Social Science. In *Philosophy of Behavioral Biology*, pages 43–64.

Turner, A., Beales, J., Faure, S., Dunford, R. P., and Laurie, D. A. (2005). The Pseudo-Response Regulator Ppd-H1 Provides Adaptation to Photoperiod in Barley. *Science*, 310(5750):1031–1034.

Uauy, C., Brevis, J. C., Chen, X., Khan, I., Jackson, L., Chicaiza, O., Distelfeld, A., Fahima, T., and Dubcovsky, J. (2005). High-temperature adult-plant (HTAP) stripe rust resistance gene Yr36 from Triticum turgidum ssp. dicoccoides is closely linked to the grain protein content locus Gpc-B1. *Theoretical and Applied Genetics*, 112(1):97–105.

Ubbens, J., Parkin, I., Eynck, C., Stavness, I., and Sharpe, A. G. (2021). Deep neural networks for genomic prediction do not estimate marker effects. *The Plant Genome*, 14(3).

Unrau, J. (1950). The use of monosomes and nullisomes in cytogenetic studies of common wheat. *Scientific Agriculture*, 30(2):66–89.

VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. *Journal of Dairy Science*, 91(11):4414–4423.

Wang, W., Simmonds, J., Pan, Q., Davidson, D., He, F., Battal, A., Akhunova, A., Trick, H. N., Uauy, C., and Akhunov, E. (2018). Gene editing and mutagenesis reveal inter-cultivar differences and additivity in the contribution of TaGW2 homoeologues to grain size and weight in wheat. *Theoretical and Applied Genetics*, 131(11):2463–2475.

Watanabe, N. and Ikebata, N. (2000). The effects of homoeologous group 3 chromosomes on grain colour dependent seed dormancy and brittle rachis in tetraploid wheat. *Euphytica*, 115(3):215–220.

Wells, D. G., Sze-Chung Wong, R., Lay, C. L., Gardner, W. S., and Buchenau, G. W. (1973). Registration of C.I. 15092 and C.I. 15093 Wheat Germplasm 1 (Reg. No. 34 and 35). *Crop Science*, 13(6):776–776.

Whittaker, J. C., Thompson, R., and Denham, M. C. (2000). Marker-assisted selection using ridge regression. *Genetical Research*, 75(2):249–252.

Wilhelm, E. P., Turner, A. S., and Laurie, D. A. (2009). Photoperiod insensitive Ppd-A1a mutations in tetraploid wheat (Triticum durum Desf.). *Theoretical and Applied Genetics*, 118(2):285–294.

Winge, V. (1924). Zytologische Untersuchungen Über Speltoide und Andere Mutanten-Ähnliche Aberranten Beim Weizen. *Hereditas*, 5(3):241–286.

Worland, A. J., Korzun, V., Röder, M. S., Ganal, M. W., and Law, C. N. (1998). Genetic analysis of the dwarfing gene Rht8 in wheat. Part II. The distribution and adaptive significance of allelic variants at the Rht8 locus of wheat as revealed by microsatellite screening. *Theoretical and Applied Genetics*, 96(8):1110–1120.

Wright, S. (1921). Wright, Sewall. "Systems of mating. I. The biometric relations between parent and offspring. *Genetics*, 6(2):111–123.

Wright, S. (1952). The genetics of quantitative variability. *The genetics of quantitative variability.*

Yan, L., Helguera, M., Kato, K., Fukuyama, S., Sherman, J., and Dubcovsky, J. (2004). Allelic variation at the VRN-1 promoter region in polyploid wheat. *Theoretical and Applied Genetics*, 109(8):1677–1686.

Yan, L., Loukoianov, A., Tranquilli, G., Helguera, M., Fahima, T., and Dubcovsky, J. (2003). Positional cloning of the wheat vernalization gene VRN1. *Proceedings of the National Academy of Sciences*, 100(10):6263–6268.

Zhang, P., Dundas, I. S., McIntosh, R. A., Xu, S. S., Park, R. F., Gill, B. S., and Friebe, B. (2015). Wheat-Aegilops Introgressions. In *Alien Introgression in Wheat: Cytogenetics, Molecular Biology, and Genomics*, pages 221–243. Springer.

Zhao, Y., Xie, P., Guan, P., Wang, Y., Li, Y., Yu, K., Xin, M., Hu, Z., Yao, Y., Ni, Z., Sun, Q., Xie, C., and Peng, H. (2019). Btr1-A Induces Grain Shattering and Affects Spike Morphology and Yield-Related Traits in Wheat. *Plant and Cell Physiology*, 60(6):1342–1353.

Chapter 2

Characterizing the oligogenic architecture of plant growth phenotypes informs genomic selection approaches in a common wheat population

Noah DeWitt[1, 2] Mohammed Guedira[1] Edwin Lauer[1] J. Paul Murphy[1] David Marshall[2] Mohamed Mergoum[3] Jerry Johnson[3] James B. Holland[1, 2] Gina Brown-Guedira[1, 2]

1 Department of Crop and Soil Sciences, North Carolina State University, Raleigh, NC 27695
2 USDA-ARS SEA, Plant Science Research, Raleigh, NC 27695
3 Department of Crop and Soil Sciences, University of Georgia, Athens, GA 30602

Published in *BMC Genomics*, May 2021, (DOI: doi.org/10.1186/s12864-021-07574-6)

2.1 Abstract

Genetic variation in growth over the course of the season is a major source of grain yield variation in wheat, and for this reason variants controlling heading date and plant height are among the best-characterized in wheat genetics. While the major variants for these traits have been cloned, the importance of these variants in contributing to genetic variation for plant growth over time is not fully understood. Here we develop a biparental population segregating for major variants for both plant height and flowering time to characterize the genetic architecture of the traits and identify additional novel QTL. We find that additive genetic variation for both traits is almost entirely associated with major and moderate-effect QTL, including four novel heading date QTL and four novel plant height QTL. *FT2* and *Vrn-A3* are proposed as candidate genes underlying QTL on chromosomes 3A and 7A, while *Rht8* is mapped to chromosome 2D. These mapped QTL also underlie genetic variation in a longitudinal analysis of plant growth over time. The oligogenic architecture of these traits is further demonstrated by the superior trait prediction accuracy of QTL-based prediction models compared to polygenic genomic selection models. In a population constructed from two modern wheat cultivars adapted to the southeast U.S., almost all additive genetic variation in plant growth traits is associated with known major variants or novel moderate-effect QTL. Major transgressive segregation was observed in this population despite the similar plant height and heading date characters of the parental lines. This segregation is being driven primarily by a small number of mapped QTL, instead of by many small-effect, undetected QTL. As most breeding populations in the southeast U.S. segregate for known QTL for these traits, genetic variation in plant height and heading date in these populations likely emerges from similar combinations of major and moderate effect QTL. We can make more accurate and cost-effective prediction models by targeted genotyping of key SNPs.

2.2 Introduction

Wheat is a major food crop, contributing nearly 20% of human calories and protein (FAO 2020). Wheat yield is highly polygenic, with variation in yield emerging from variation in other phenotypes each with different genetic bases. Plant growth traits such as heading date (when the spike emerges from the flag leaf) and adult plant height affect yield by both altering resource partitioning between tissues and changing how plants experience

environmental factors. A plant's height on a given date alters the physical position of the plant within its environment, influencing that plant's interactions with environmental factors like wind, weed competitors, and rain-splashed pathogens. Differences in heading date change the temporal position of plants at a given developmental stage, exposing them to different weather conditions and disease pressures. Wheat breeders typically select for optimal values of plant height and heading date for a given environment and production system in early generations based on unreplicated head rows. Beyond this selection, improvements in yield resulting from modern plant breeding programs have largely been generated without considering its underlying genetic architecture, including the dependence of final plant yield on variation in plant growth trajectories. Understanding the plant development factors that generate genetic variation in yield is critical to increasing the rate of genetic gain in wheat.

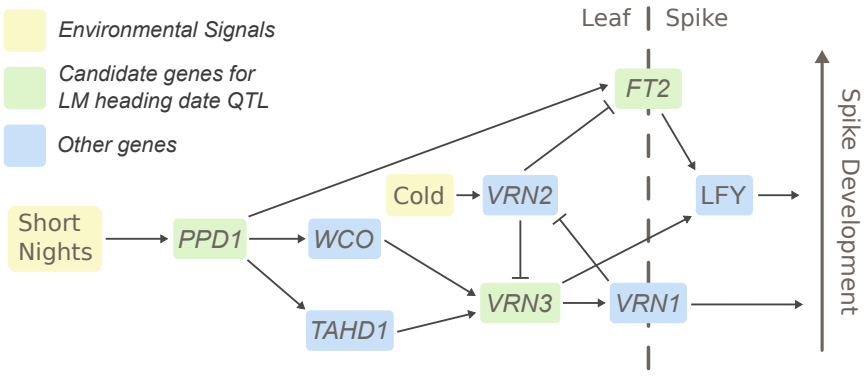

Figure 2.1: **Overview of the wheat flowering time pathway.** The gene network through which wheat plants receive and integrate signal about environmental conditions to determine heading date are outlined. Other, intermediate genes are not shown. Genes proposed as candidate for heading QTL in this population are highlighted in green. Other important genes in the flowering time pathway are highlighted in blue; *Wheat CONSTANS (WCO)*, *Triticum aestivum HD1 (TaHD1)*, *VERNALIZATION1 (VRN1)*, *VERNALIZATION2 (VRN2)*, and *LEAFY (LFY)*.

Allelic variation affecting core flowering time genes is strongly associated with the geographic distribution of wheat cultivars, and permits the cultivation of wheat in a wide range of environments. Winter wheat is sown in the fall, when it germinates but maintains the shoot apical meristem beneath the ground to prevent freeze damage. After the accumulation of signal through the vernalization (cold hours), photoperiod (night length), and

earliness-per-se (plant age) pathways, plants release from winter dormancy and transition to reproductive development. Allelic series in the *Vernalization1* (*Vrn1*) loci on the three chromosome 5 homeologues condition a spring or winter growth habit by controlling the sensitivity of plants to vernalization (Fig. 2.1) (Yan et al. 2004; Fu et al. 2005; Li et al. 2013). Additional alleles at these loci, some associated with copy number variation, may also modulate vernalization response in vernalization-sensitive winter lines (Fu et al. 2005; Díaz et al. 2012; Guedira et al. 2014, 2016; Kippes et al. 2018). *Photoperiod1* (*Ppd1*) is another core flowering time gene which integrates signals due to length of nights and allows plants to time flowering based on photoperiod. Variants affecting homeologous *Ppd1* loci on all three genomes lead to constitutive over-expression of *Ppd1* and a photoperiod-insensitive, earlier flowering habit (Beales et al. 2007; Nishida et al. 2013; Guedira et al. 2016). Breeding for an optimal heading date for a given environment allows plants enough time to add additional spikelets per spike prior to heading, which increases grain number, and to accumulate carbohydrates and fill grain, which increases grain size. Non-optimal flowering can expose plants to temperatures below freezing early in the season, or to excessive heat and drought late in the growing season. In southern U.S. field sites, *Vrn1* and *Ppd1* alleles have strong effects on final grain yield of winter wheat (Addison et al. 2016).

Introduction of "green revolution" dwarfing genes *Rht-B1* and *Rht-D1* – knock-out mutations in *DELLA* proteins – into US and CIMMYT germplasm dramatically improved yields by increasing wheat harvest index and preventing lodging due to applied inorganic nitrogen fertilizer (Hedden 2003). The effect of the *Rht1* genes is conditional on the environment and the quantity of assimilate produced by the variety, and has been associated with larger grain number but smaller grain size and weight (Borner et al. 1993). *Rht1* alleles disable plants' ability to respond to giberellic acid (GA-insensitivity), which may have negative effects on coleoptile length and early plant vigor that can decrease yield in some environments (Rebetzke and Richards 2000). Increase in seed number and grain yield seems to be related not to ear development but to the greater availability of assimilates during grain-fill with reduced biomass partitioned into stalks (Youssefian et al. 1992). Breeders generally select plants near some optimal height value, as too-short plants have a generally lower yield compared to semi-dwarfs characteristic of having only one *Rht* allele (Fischer and Quail 1990). An increasing number of dwarfing genes in wheat have been fine-mapped, and many, though not most, have been cloned.

Genomic selection is restructuring modern wheat breeding programs. The ability to

leverage data from past years to predict unobserved lines has tremendous potential to increase the rate of genetic gain. Beyond yield predictions, heading date and plant height predictions are valued by breeders, allowing them to exclude phenotypically extreme individuals without having to dedicate resources to planting and phenotyping in multiple environments. Standard GBLUP and rrBLUP models are optimized for highly polygenic traits like yield, but will underestimate QTL effect sizes and perform poorly with traits dominated by a smaller number of larger-effect QTL. Explicitly characterizing and taking into account large-effect QTL in traits where these QTL explain a substantial portion of additive genetic variation can increase prediction accuracy (Bernardo 2014; Sarinelli et al. 2019). This may have even more promise in biparental populations where the number of segregating causal variants is much smaller. If traits are mostly controlled by a few major variants, models using markers for just those variants can be predictive and more cost-effective.

Here we set out to understand the genetic basis of plant growth traits in a biparental common wheat population. Parents were chosen to generate major additive genetic variation for plant height and heading date and to characterize novel QTL for these traits. Parent SS-MPV57 carries the large-effect earliness allele *Ppd-D1a* as well as the smaller-effect allele *Ppd-A1a.1*, but no known dwarfing genes. Parent LA95135 caries the major dwarfing allele *Rht-D1b* but no known earliness genes except the smaller-effect *Ppd-A1a.1* allele. This parental selection contrasts with typical mapping studies, where the two parents generally differ for the trait of interest. Here, due to our understanding of already characterized major variants, we developed a population with the goal of generating transgressive segregation for our target phenotypes. A high-density sequence-based linkage map was supplemented with single SNP assays for putative causal variants in order to map novel QTL and study marker-trait associations for known variants. Phenotypic variation in each environment was partitioned into components associated with mapped QTL and the polygenic background in order to assess the relative importance of identified QTL. Different models were tested for prediction of both traits to determine if a simple QTL model would be sufficient in the context of a breeding program. Finally, a longitudinal analysis of multiple measures of plant height over time was used to determine QTL effects over the course of plant growth in a field season.

2.3 Materials and Methods

2.3.1 Population Development

Soft-red winter wheat lines developed by southeastern public-sector breeding programs were screened for alleles at known plant height and heading date variants using Kompetitive Allele-Specific PCR (KASP) markers. Louisiana State University forage cultivar LA95135 (CL-850643/PIONEER-2548//COKER-9877/3/FL-302/COKER-762) was chosen as a parent lacking major early-flowering alleles at the *Ppd-D1* or *Vrn-1* loci, but with a mid-season heading date when grown in North Carolina. Cultivar SS-MVP57 (FFR555W/3/VA89-22-52/TYLER //REDCOAT*2/GAINES) developed at Virginia Polytechnic Institute and State University displayed semi-dwarf stature but lacked dwarfing alleles at the *Rht1* loci. SS-MPV57 carries the *Ppd-D1a* allele conferring photoperiod insensitivity, and LA95135 has the *Rht-D1b* allele conferring semi-dwarfism. Parent lines were crossed, and F1 plants were selfed to generate an F_2 population (hereafter referred to as the LM population). The F_2 and later generations were inbred via the single-seed descent method until the F_5 generation, producing 358 F_5-derived recombinant inbred lines (RILs).

2.3.2 Phenotyping

During the winter of 2016-2017, an experiment was conducted in the greenhouse to evaluate heading date. Imbibed seeds from each RIL were placed in a cold chamber kept at 4°C for 8 weeks and were transplanted into plastic cones (volume 0.7L, 6.5 cm in diameter and 25 cm depth) containing soil mix. Plants were grown in a completely randomized design with four replications in a greenhouse set at 16 hr photoperiod and 20°C /15°C (day/night) temperature.

To evaluate the impact of vernalization on the genetic architecture of heading date and on effects of individual QTL, the greenhouse experiment was repeated with a low-vernalization treatment in the winter of 2017-2018. This experiment was performed as above, except that imbibed seeds were placed in the cold chamber for only four weeks prior to transplanting. In addition, the LM RIL population was evaluated in the field at Raleigh, NC and Kinston, NC during the 2017-2018 season, and in Raleigh, Kinston, and

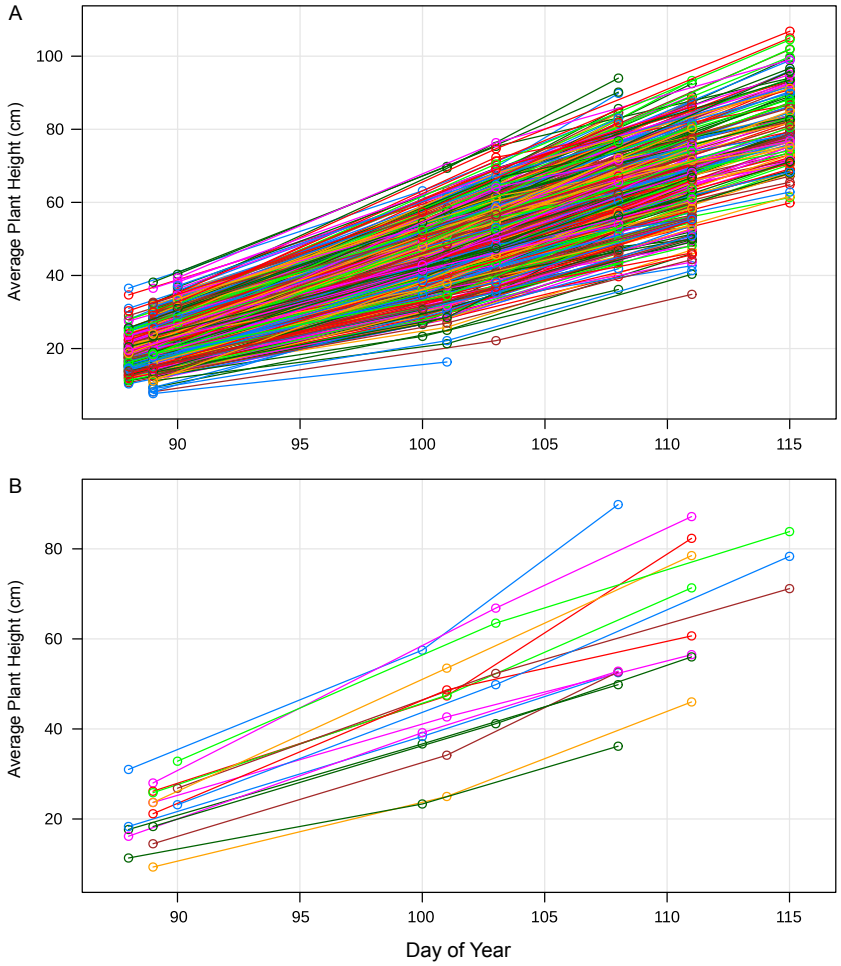

Figure 2.2: **Plant growth over time.** For each 1-m row plot (differently colored line), a total of three plant height values was collected in Raleigh in 2019. All plots are shown (A), as well as a random subset to better visualize plant growth (B). Mean plant growth follows a roughly linear pattern corresponding to the date collected, with different slopes and intercepts for each plot.

Plains, GA in the 2018-2019 season, sown in the fall at the locally recommended times for commercial winter wheat production. The 358 RILs were grown using an augmented set within replications design to facilitate planting of this large population. RIL experiments consisted of two fully replicated blocks of all 358 lines organized into five sets of 71 or 72 RILs. The order of the sets within each replication and the order of the RILs within each set were randomized at each location. Three parental checks were planted at the beginning of

each set of RILs, along with four or five additional parental checks randomized within each set.

Plots consisted of 1-m rows spaced 30 cm apart. Adult plant height was measured as the distance from the ground to the top of the spikes of a sample of tillers from the center of each row, excluding the awns. Heading date was measured as the day on which approximately half of the heads in each row had fully emerged from the flag leaf, typically a few days prior to anthesis. To study plant development over time, three measures of plant height were collected for each row plot in Raleigh in 2019, with two to four blocks measured roughly every ten days starting on March 29th and ending on April 25 (when most plants had fully headed). In this case, plant height at each time point was calculated as the mean of the height of three randomly chosen primary tillers from the ground to the base of the apical leaf sheath (Fig. 2.2). All measurements were collected on an android tablet with the Fieldbook app (Rife and Poland 2014).

2.3.3 Analysis of Phenotypes

For the greenhouse experiments where plants had been completely randomized within greenhouses, genotype values for RILs were calculated as the mean of the four replications of each line. For field experiments, best linear unbiased estimates (BLUEs) were calculated adjusting for these spatial effects. The software ASReml-R (Butler et al. 2017) was used to calculate BLUEs with an AR1xAR1 correlated residuals model:

$$Y_{ik} \sim \mu + G_i + u_{ik} + e_{ik}$$

Where Y_{ik} is the observed phenotype for an individual row plot, μ is the intercept, G_i is the fixed effect of genotype i, u_{ik} is the unit or "nugget" random residual effect for each observation k representing the component of the variance due to observation or measurement instead of spatial correlation, drawn from a distribution $u \sim iidN(0, \sigma_e^2)$, and e_{ik} is the spatially-correlated residual drawn from the distribution $e \sim N(0, \sigma_e^2 \Sigma_r(\rho_r) \otimes \Sigma_c(\rho_c))$, whose variance is the direct product of an $r \times r$ auto-correlation matrix $\Sigma_r(\rho_r)$ representing autoregressive correlations in the row direction and $c \times c$ correlation matrix $\Sigma_c(\rho_c)$ representing autoregressive correlations in the column direction. For all environments

and phenotypes, a full model with autocorrelated columns and rows was found to have a lower BIC and higher log likelihood than models with just the column autocorrelation or no spatial correction. BLUEs were calculated as the sum of the genotype effect and the intercept for each phenotype in each environment.

2.3.4 Genotyping and Linkage Map Construction

Tissue was collected from the F_5 greenhouse experiment, and seeds of the four $F_{5:6}$ plants from each line were bulked. Genotyping by sequencing (GBS; (Elshire et al. 2011)) was performed according to Poland *et al.* 2012 (Poland et al. 2012), with ninety-six individual samples barcoded, pooled into a single library, and sequenced on an Illumina HiSeq 2500. Tassel5GBSv2 pipeline version 5.2.35 (Glaubitz et al. 2014) was used to align raw reads to the International Wheat Genome Sequencing Consortium (IWGSC) RefSeqv1.0 assembly (https://wheat-urgi.versailles.inra.fr/Seq-Repository) using Burrows-Wheeler aligner (BWA) version 0.7.12 to call SNPs (Li et al. 2009). SNPs were filtered to retain samples with ≤ 20 percent missing data, ≥ 30 percent minor allele frequency and ≤ 10 percent of heterozygous calls per marker.

KASP markers taken from the literature or designed from exome capture data of the parents (triticeaetoolbox.org/wheat; Additional file 1: Table S1) were added to the GBS SNP data for chromosome regions with low marker density and for causal variants segregating in the population. Filtered SNPs were separated into chromosomes and ordered via alignment to the reference genome, and a custom script was run to filter out genotyping errors that would result in a false double recombination due to under or mis-calling of heterozygotes. The R package ASMap was used to construct the maps as an F_5 RIL population (Taylor and Butler 2017).

2.3.5 QTL Analysis

QTL mapping was performed in the R package r/QTL (Broman et al. 2003). Composite interval mapping was used for initial QTL identification, and intervals were narrowed using a multiple QTL model (MQM) as implemented in the refineqtl function. The addqtl function was used to identify additional QTL using identified QTL as covariates. Empirical signifi-

cance thresholds for a genome-wide $\alpha = 0.05$ were determined using 1000 permutations for each trait. QTL effects were estimated for significant QTL in each environment based on the estimated MQM positions using the fitqtl functions, which fits a multiple regression where for genotype values Y_i for each individual i, and n QTL Q, $Y_i \sim \sum_{h=1}^{n} Q_{ih} + e_i$.

Major variants *Ppd-D1* and *Rht-D1* alter functions of core genes in the flowering time and giberellic-acid response pathways, respectively, suggesting that their presence may alter the effects of other variants impacting those pathways. Taking advantage of the large number of lines in the LM population, two sub-populations of roughly 160 lines – each divided by genotype at the major-effect QTL – were created for mapping of each phenotype. Lines called as heterozygous for the major-effect QTL were excluded. The QTL mapping analysis was repeated for both of the sub-populations. Identified QTL interactions discovered this way were validated by modifying the above fitqtl model with a main effect for the identified QTL and an effect for its interaction with the major classifying QTL.

Variance analysis was performed in the R package lme4qtl, which allows for the fitting of random effects with supplied covariance matrices (Ziyatdinov et al. 2018). For known variants for which KASP marker genotypes of the causal polymorphisms were available, the genotypes were used directly, and for novel QTL genotype probabilities from the refineqtl object were used. For testing QTL, alleles were encoded in terms of the allele dosage of the LA95135 allele (0, 1, 2) without estimating a dominance effect.

While BLUEs estimated using the correlated errors model were used for QTL mapping, to estimate the relative importance of identified QTL in determining total phenotypic variation at the level of individual plots models were re-fit in each environment using the unadjusted phenotypes as the response. For each environment and phenotype, QTL effects and variance components for each the additive and non-additive effects of genotypes were specified with the mixed model:

$$Y_{ik} \sim \sum_{h=1}^{n} Q_{ih} + g_{Ai} + g_{Ii} + e_{ik}$$

Where for each phenotype Y of genotype i in row plot k, fixed effects for each QTL h were fit as regressions of allele dosage on phenotypes. g_{Ai} represents the random additive effect of genotype i with a variance specified by the realized relationship matrix ($g_A \sim \mathcal{N}(0, \mathbf{G}\sigma_g^2)$),

calculated using the A.mat function in the R package rrBLUP from the scaled GBS marker matrix ($\mathbf{G} = \frac{\mathbf{WW'}}{c}$, where $\mathbf{W}$ is the scaled marker matrix calculated as $W_{ik} = X_{ik} + 1 - 2p_k$ from the frequency of the 1 allele at marker k (p_k) and the marker matrix X_{ik}. c is a normalization value calculated as $c = 2\Sigma_k p_k(1-p_k)$) (Endelman 2011). g_{Ii} represents the non-additive random effect of genotype i with an independent variance ($g_I \sim \mathcal{N}(0, I\sigma_g^2)$).

A modified method from Nakagawa and Schielzeth 2013 (Nakagawa and Schielzeth 2013) was used to estimate variances associated with QTL and variance components from the specified model. Estimated coefficients for fixed effects are multiplied by the value of that effect (in this case, the allele dosage), and the variance of these values is taken as the variance associated with that fixed effect. This R^2-like estimator for mixed models is defined as:

$$R^2_{LMM} = \frac{\sigma^2_f}{\sigma^2_f + \sigma^2_r + \sigma^2_e}$$

Where σ^2_r is the variance of the random effects and σ^2_f is a variance of independent fixed effects calculated as $\sigma^2_f = \sum_{h=1}^{n} Var(\beta_h x_{hk})$ for coefficients and effects h and observations k. For both traits, all QTL were mapped to separate chromosomes, satisfying the assumption of independence. Using this approach, narrow-sense heritability in this population with n QTL in i individuals in k head rows is calculated as:

$$h^2 = \frac{\sum_{h=1}^{n} Var(\beta_h Q_{hk}) + \sigma^2_A}{\sum_{h=1}^{n} Var(\beta_h Q_{hk}) + \sigma^2_A + \sigma^2_I + \sigma^2_e}$$

Where we calculate per-observation QTL effects as the allele dosage of QTL h in plot k (Q_{hk}) times the estimated coefficient of each QTL (β_h), and the phenotypic variance associated with that QTL as the variance of these estimates. The total variance associated with all QTL is taken as the sum of these individual QTL variances, as all mapped QTL are located on separate chromosomes and are independent of one another. σ^2_A is the variance component associated with the random g_A genotype term fit with the relationship matrix, and σ^2_I as the variance component associated with the random independent g_I genotype

term, which represents some combination of epistatic effects, lack of linkage between observed markers and underlying causal variants, and deviation of the estimated genotype values from the true genotype values. Constructing the model in this way, we estimate the proportion of additive genetic variation associated with a QTL h (p_A) as:

$$p_A = \frac{Var(\beta Q_k)}{\sum_{h=1}^{n} Var(\beta_h Q_{hk}) + \sigma_A^2}$$

Where p_A is taken as the variance of the product of an estimated QTL effect by the allele dosage of that QTL in k rows, over the total additive genetic variation.

For investigating the effect of QTL on plant height variation over time, individual slopes of plant height over time measured multiple times for each row were calculated with a fixed intercept and a random intercept and time slope for each head row. The model was used to estimate plant height values for each row every day over the course of the month data was collected. A linear model fitting all relevant plant height and heading date QTL on plant height on every day was fit, and the partial R^2 values of each QTL calculated for each day were used to estimate the relative importance of each QTL at each time point.

2.3.6 Prediction of Phenotypes

Different prediction models were assessed to identify an optimal model for heading date and adult plant height. All models except for the simple QTL multiple regression model were fit in the R package BGLR (Pérez and De Los Campos 2014), which allows for flexible fitting of a variety of Bayesian and mixed effects models. A GBLUP model was fit solving the equation $y \sim \mu + u + e$ for u. y is a vector of BLUEs across environments for all RILs, with unobserved RILs assigned missing values, and u is a vector of random genotype effects with a variance $u \sim \mathcal{N}(0, \mathbf{G}\sigma_u^2)$, where $\mathbf{G}$ is the realized relationship matrix calculated previously from GBS markers.

A simple multiple-regression QTL model based on identified QTL was fit solving the equation $y \sim \mu + \sum_{h=1}^{n} \alpha_h \mathbf{Q}_h + e$ for n QTL, where $\mathbf{Q}_h$ encodes the LA95135 allele dosage for each QTL h in each individual, and α_h is the allele effect of QTL h.

A combined model was also fit specifying both a multiple-regression fixed-effects component for QTL effects, and random effects for each genotype constrained by the additive relationship matrix ($y \sim \mu + \sum_{h=1}^{n} \alpha_h \mathbf{Q}_h + \mathbf{I}u + e$, where $u \sim \mathcal{N}(0, \mathbf{G}\sigma_u^2)$).

BayesB and Bayesian LASSO models were both fit with the general model $y \sim \mu + \mathbf{X}u + e$, where $\mathbf{X}$ is a design matrix of markers coded by allele dosage of the LA95135 allele, and u is a vector of random marker effects. In the BayesB model, a certain proportion of markers given by the prior probability π $(u_i|\sigma_i^2, \pi)$ are assumed to have an effect size of 0, with the remainder having effects following a scaled-t distribution (Pérez and De Los Campos 2014). In the Bayesian LASSO, marker effects were estimated with a double exponential prior distribution that assumes a greater frequency of both larger marker effects and marker effects closer to zero than a normal distribution (Pérez and De Los Campos 2014). In both models, estimated genotype values are calculated as the sum of marker effects $\hat{Y}_i = \hat{\mu} + \sum_{j=1}^{n} x_{ij}\hat{u}_j$, where x_{ij} is the allele dosage of marker j in individual i, and $\hat{u}_j$ is the estimated marker effect.

A five-fold cross validation approach was used to compare the five models. RILs were randomly assigned to one of five folds, and genotype values from each environment from lines in four of the folds were used to predict the values of lines in the fifth fold, repeating for each fold in each environment for each model. Within each fold, QTL detection was re-performed as described in the QTL analysis section to identify the QTL used in the QTL regression and combined QTL and GBLUP model. This process was then repeated 40 times to get distributions of prediction abilities, calculated as the Pearson's correlation between predicted and observed genotype values across all five folds.

2.4 Results

2.4.1 Genetic Map Construction

After filtering, 5691 markers were assigned to 21 linkage groups representing 21 wheat chromosomes. Average chromosome map length was 208.3 cM, with a maximum individual chromosome length of 319.1 cM for chromosome 3B. Marker density on the D genome tended to be much lower than marker densities on the A and B genomes, as expected given

the much lower D genome diversity in hexaploid wheat (Akhunov et al. 2010).

2.4.2 Population Characterization

Generally, wheat cultivars' flowering habits are described by their genotypes at major heading date loci, but SS-MPV57 flowered later than LA95135 in all locations despite carrying the major earliness allele *Ppd-D1a* (Table 2.1). The difference in heading date was especially pronounced in the low-vernalization treatments both in the greenhouse (GH 2018) and in the field (Plains 2019), where SS-MPV57 flowered five and six days later, respectively, than LA95135. A similar pattern was observed for plant height: although LA95135 was the only parent genotyped for a major dwarfing allele (*Rht-D1b*), SS-MPV57 was substantially shorter in all locations (Table 2.1). For heading date in all locations, the mean genotype value of the RILs was approximately the mid-parent value. For plant height in Raleigh 2018 and Kinston 2018, the mean genotype value of the RILs was closer to the SS-MPV57 parent than the mid-parent value. The ranges of genotype values in Raleigh and Kinston were similar, but the range in heading date in Plains 2019 (26 days) was much larger. This is likely a result of the warmer winter temperatures at that site, delaying heading of lines with a greater vernalization requirement.

Table 2.1: Population characteristics. Means and ranges of estimated genotype values for all RILs, as well as parental values and plot-basis heritabilities (H), for site-year-phenotype combinations. Heading date for the GH experiments is recorded as days since transplanting (four weeks (HD_4W) or eight weeks (HD_8W) after vernalization), and for the field experiments time as day of year (DOY).

Loc.	**Year**	**Pheno**	$\mu_{\mathbf{LA}}$	$\mu_{\mathbf{MPV}}$	$\mu_{\mathbf{RILs}}$	**Range**	H
GH	2017	HD_{8W}	NA	NA	45.6	32-61	0.50
GH	2018	HD_{4W}	71.1	76.12	77.1	59-97	0.76
Ral	2018	HD	112.3	112.9	112.3	107-120	0.73
Kin	2019	HD	103.1	104.9	103.6	97-109	0.63
Pla	2019	HD	100.7	106.6	103.1	90-116	0.63
Ral	2018	PH	100.1	92.5	93.7	68-127	0.84
Kin	2018	PH	101.0	94.0	96.0	67-122	0.71
Kin	2019	PH	99.1	94.9	97.5	69-129	0.67

Table 2.2: Significant heading date QTL for four and eight week vernalization greenhouse experiments. The chromosome on which each QTL is found is indicated in the QTL name. For each QTL, the average difference in phenotype between two RILs homozygous for alternate alleles is given as twice the estimated allele effect of the LA95135 allele (2α), along with proportion of additive variation associated with each QTL ($\mathbf{p_A}$). The most significant markers for each QTL with a proposed candidate gene was a KASP marker associated with a previously identified causal polymorphism affecting that gene. Physical positions are given based on mapping of GBS markers to the IWGSC RefSeqv1.0 assembly.

Treatment	**QTL Name**	**Candidate Gene**	**Peak Marker**	**Position CI**	**LOD**	**2α** (days)	**$\mathbf{p_A}$**
4 Wk	*Qncb.HD-2D*	*Ppd-D1*	Ppd-D1	32-43 Mb	11.5	5.2	0.15
4 Wk	*Qncb.HD-3A*	*FT2*	FT2	118-478 Mb	21.2	6.5	0.25
4 Wk	*Qncb.HD-4D*	Rht-D1	Rht-D1	0-352 Mb	22.2	7.2	0.39
4 Wk	*Qncb.HD-5A*	NA	S5A_395681218	46-438 Mb	5.65	3.2	0.05
4 Wk	*Qncb.HD-5B*	NA	S5B_462554252	427-523 Mb	4.63	3.6	0.08
8 Wk	*Qncb.HD-2D*	*Ppd-D1*	Ppd-D1	33-44 Mb	15.9	3.9	0.60
8 Wk	*Qncb.HD-3A*	*FT2*	FT2	71-435 Mb	6.67	2.3	0.22
8 Wk	*Qncb.HD-5B*	NA	S5B_518684640	511-537 Mb	5.17	2.1	0.14

2.4.3 Known Variants and Novel QTL Impact Plant Growth

Genetic variation in quantitative traits like those measured in this study may result from the segregation of an unquantifiable number of small-effect QTL. Despite this, for both heading date and adult plant height the vast majority of additive genetic variation was associated with a small number of major QTL, some of which have been previously described and some of which are novel.

Heading Date

The RIL population was developed with the expectation that the major photoperiod-insensitive allele *Ppd-D1a* inherited from SS-MPV57 would segregate, and that potential novel early-flowering QTL from LA95135 could be mapped. Two preliminary greenhouse experiments were conducted to investigate the effect of vernalization treatments on heading date genetic architecture, with imbibed seeds given only four weeks of vernalization in the first experiment and a full eight weeks in the second. In addition, heading date notes were collected in three separate field experiments in Raleigh in 2018, and Kinston and Plains, GA in 2019. QTL were declared significant at $\alpha = .05$ based on 1000 permutations of

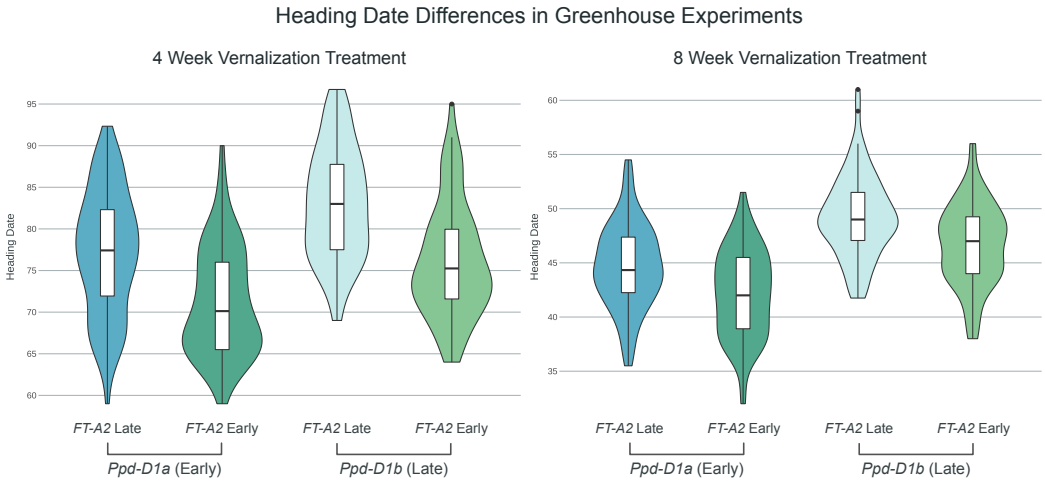

Figure 2.3: **Effect of *Qncb.HD-3A* and *Ppd-D1* QTL on heading date in two different vernalization treatments.** Density plots of BLUEs for heading date in two experiments, with RILs grouped by their genotype at *Ppd-D1* and a marker close to *FT-A2*. The allele effect of *Ppd-D1* is larger than that of *FT-A2* in the 8 week vernalization treatment (2.0 days versus 1.2), but the effect of the *FT-A2* marker is larger in the 4 week vernalization treatment (2.6 days vs 3.3 days).

the scanone function, but for all phenotypes significance values were near a LOD of 3.5. Together, *Ppd-D1*, *Rht-D1*, and four early-flowering alleles inherited from LA95135 were associated with differences in heading date in this experiment (Table 2.2, Table 2.3).

A heading date QTL on chromosome 2D co-localized with a known major-effect variant altering expression of the D-genome copy of pseudo-response regulator gene *Photoperiod-1* (*Ppd-D1a*) (Beales et al. 2007). This was the major QTL mapped in this experiment, associated with by far the highest LOD score in both the field environments (Table 2.3) and the eight week greenhouse treatment (Table 2.2). In the four week treatment the relative importance of *Ppd-D1* was diminished, primarily as a result of changes in the effects of other QTL.

A QTL in the centromeric region of chromosome 3A is mapped with low physical resolution (> 400 Mb), owing to the low recombination rates found in these regions in wheat (Table 2.2, Table 2.3). Contained in this interval is *FT-A2*, an ortholog of *FT* previously described by Shaw *et al.* (Shaw et al. 2019) as an important component of the wheat flowering time pathway (Fig. 2.1). A KASP assay designed from a polymorphism within *FT-A2* was

Table 2.3: Significant heading date QTL information from best environment. For each QTL, information from the experiment where that QTL had the largest estimated effect (Best Env) is given. The average difference in phenotype between two RILs homozygous for alternative alleles at each QTL is given as twice the estimated allele effect of the LA95135 allele (2α), along with proportion of additive variation associated with each QTL ($\mathbf{p_A}$). *Vrn-A3* only has a significant effect within the half of the population homozygous *Ppd-D1b*.

QTL Name	**Candidate Gene**	**Best Env**	**Peak Marker**	**Position CI**	**LOD**	**2α** (days)	**$\mathbf{p_A}$**
Qncb.HD-2D	*Ppd-D1*	Ral 19	Ppd-D1	59-64 Mb	48.6	3.4	0.67
Qncb.HD-3A	*FT2*	Pla 19	S3A_434822203	121-571 Mb	13.0	-2.6	0.12
Qncb.HD-4D	Rht-D1	Kin 19	Rht-D1	0-352 Mb	4.85	0.9	0.06
Qncb.HD-5A	NA	Pla 19	S5A_169302619	51.6-435 Mb	14.1	-2.7	0.15
Qncb.HD-5B	NA	Kin 19	S5B_511010094	436-476 Mb	7.7	-1.3	0.08
Qncb.HD-7A	*Vrn-A3*	Ral 18 (*Ppd-D1b*)	S7A_72104395	57.7-85.9 Mb	4.23	NA	NA

the peak marker for this QTL in both greenhouse experiments (Table 2.2), and had a much greater effect in the four week vernalization treatment than in the eight week treatment (Fig. 2.3). *Qncb.HD-3A* was also identified as significant in all field experiments, but with alternate peak markers in the long arm of chromosome 3A (Table 2.3).

In addition, two novel early-flowering alleles were identified on chromosomes 5A and 5B (Table 2.3). Both QTL are significant in all environments (Additional file 2: Table S3). *Qncb.HD-5A* is the third-most important QTL in most environments, but has an especially large effect in the Plains, GA field experiment. The QTL is also significant in the four week vernalization greenhouse treatment, but not the eight week treatment. The increased QTL effect in these two environments having shorter duration of cold temperature exposure suggests that *Qncb.HD-5A* may interact with genes involved in vernalization response. Due to its centromeric position, the confidence interval for the QTL contains 383 Mb of chromosome 5A (Table 2.3). Notably, despite the response of this QTL to vernalization treatment, this interval does not encompass the *Vrna-A1* locus.

Qncb.HD-5B is located in a more distal position on the long arm of the chromosome and was mapped to an interval of 61 Mb. This interval is proximal to *Vrn-B1*, excluding that locus as a candidate gene. Unlike *Qncb.HD-5A*, significance and effect sizes of *Qncb.HD-5B* are similar in both the four and eight week vernalization treatments (Table 2.2).

The major plant height QTL *Rht-D1* was also identified as having a pleiotropic effect on heading date in this population. In most environments the effect on heading date was

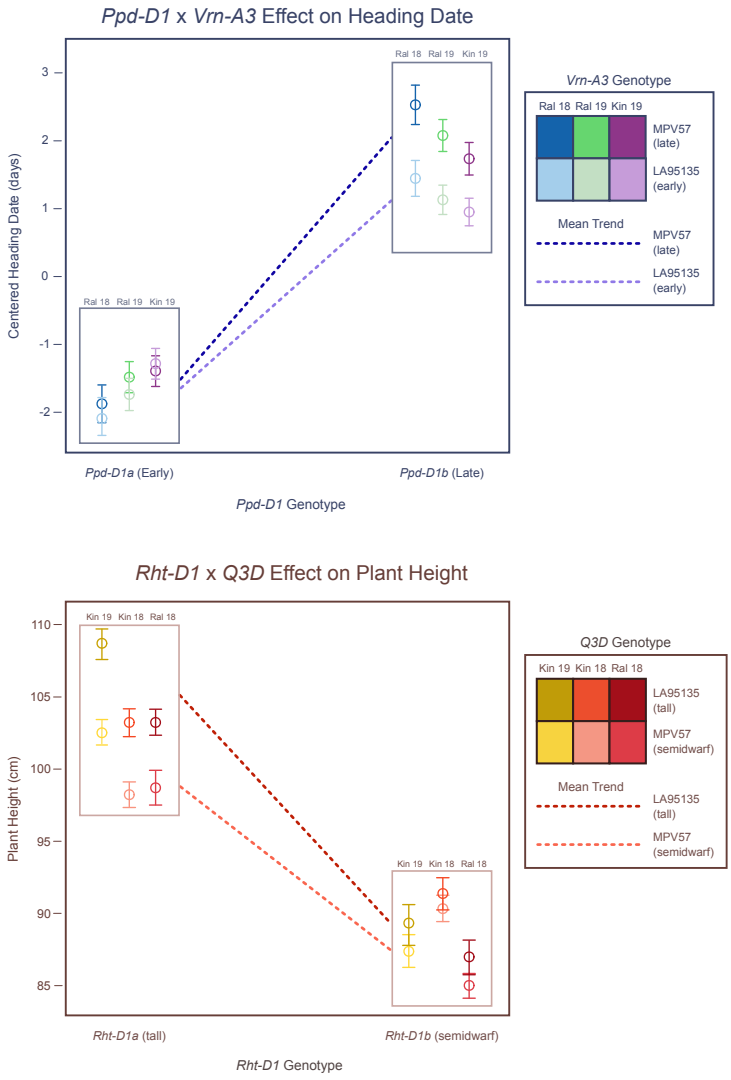

Figure 2.4: **Major variants diminish effects of other QTL.** *Vrn-A3* alters heading date in most environments, but only in a *Ppd-D1* sensitive background. The dwarfing effect of *Qncb.PH-3D* is greater in an *Rht-D1a* (tall) background.

minor, and not significant in the eight week greenhouse treatment or in Plains in 2019. However, in the four week greenhouse treatment *Rht-D1* was a highly significant QTL, with an average difference of over seven days between plants homozygous for wild type or semi-dwarf alleles (Table 2.2).

A benefit of large population sizes is the ability to subset the population by major-effect variant allele and perform QTL analyses on the sub-populations. In the case of the *Ppd-D1a*

insensitive allele, constitutive over-expression of *Ppd-D1* may obscure effects of variation elsewhere in the flowering pathway. After dividing the population by *Ppd-D1* allelic class and performing QTL analyses on the sub-populations, an additional early-flowering allele from LA95135 was identified on the short arm of chromosome 7A only in a *Ppd-D1b* photoperiod-sensitive background, and only in the field experiments (Table 2.3). The confidence interval for this QTL contains the *Vrn-A3* locus. *Vrn3* in wheat was identified as an *FT* ortholog (*TaFT1*), and serves as the primary integrator of flowering time signal, being translocated from the leaves to the shoot apical meristem to initiate the transition to reproductive growth (Fig. 2.1) (Yan et al. 2006). A variant in the D-genome copy of this gene, *Vrn-D3a*, was identified by (Chen et al. 2010) as a determinant of flowering time in winter wheat. A deletion of a GATA box in the promoter region of *Vrn-A3* has been recently associated with delayed flowering time in tetraploid durum wheat (Nishimura et al. 2018), and an additional polymorphism linked to differences in heading date and spikelets-per-spike has also been identified in common wheat (Chen et al. 2020). Screening the population with a KASP marker developed around the GATA box deletion (Additional file 1: Table S1) reveals that the population segregates for the deletion, with SS-MPV57 contributing the late-flowering deletion allele. While *Qncb.HD.7A* has a relatively small additive effect, it strongly interacts with *Ppd-D1a* (Fig. 2.4). In a background containing the insensitive over-expression *Ppd-D1* allele, there is no difference in heading date between lines with and without the *Vrn-A3* promoter deletion. In a *Ppd-D1b* background, however, the GATA box deletion is associated with significantly delayed heading date of approximately one day (Fig. 2.4). In wheat, *Ppd1* acts to trigger expression of *Vrn3* through signaling intermediates (Fig. 2.1), thus an interaction between the two fits with our understanding of their placement in a common pathway. This promoter deletion is a strong candidate for the variant underlying the chromosome 7A heading date QTL.

Adult Plant Height

Major QTL for plant height were initially mapped to chromosomes 4D, 2D, and 3D (Table 2.4). Using the MQM model, additional adult plant height QTL on chromosomes 1A, 2B, and 5B were also identified (Table 2.4). As expected, known variant *Rht-D1b* inherited from LA95135 was by far the largest-effect QTL across environments. Except for *Qncb.PH-5B*, all other reduced plant height alleles were inherited from SS-MPV57.

Table 2.4: Significant plant height QTL information from best environment. For each QTL, information from the experiment where that QTL had the largest estimated effect (Best Env) is given. The average difference in phenotype between two RILs homozygous for each QTL is given as twice the estimated allele effect of the LA95135 allele (2α), along with proportion of additive variation associated with each QTL ($\mathbf{p_A}$). The confidence interval for *Rht8* is consistent with prior studies placing the QTL distal to *Ppd-D1*.

QTL Name	**Candidate Gene**	**Best Env**	**Peak Marker**	**Position CI**	**LOD**	**2α** (cm)	**$\mathbf{p_A}$**
Qncb.PH-1A	NA	Kin 19	S1A_517409836	513-533 mb	4.01	3.6	0.03
Qncb.PH-2B	NA	Kin 18	S2B_662556874	530-691 mb	3.41	2.9	0.05
Qncb.PH-2D	*Rht8*	Kin 18	S2D_32151744	23.3-32.2 mb	27.9	9.4	0.32
Qncb.PH-3D	NA	Ral 18	S3D_476608044	477-527 mb	10.0	5.4	0.08
Qncb.PH-4D	*Rht-D1*	Kin 19	Rht-D1	0-352 mb	71.9	-19.6	0.68
Qncb.PH-5B	NA	Kin 18	S5B_511010094	463-524 mb	7.51	-5.6	0.05

The mapped position of the plant height QTL located on chromosome 2D is consistent with reported positions for *Rht8* (Gasperini et al. 2012). After the two major gibberellic-acid insensitive dwarfing genes *Rht-D1b* and *Rht-B1b*, the most commonly used gene is *Rht8*, which is tightly linked to *Ppd-D1* (Worland et al. 1998). In most environments, the marker most closely associated with *QPH.ncb-2D* is mapped closely distal to *Ppd-D1*. In an effort to tease apart the effects of photoperiod insensitivity and *Rht8* on plant height, we evaluated a terminal spike-compaction phenotype often associated with *Rht8* segregating in the LM population. This trait was rated in the field in Raleigh in 2019, and the major QTL was co-located with the plant height locus on the short arm of chromosome 2D (Additional file 1: Figure S1). We did not observe any significant interaction between *Rht8* and *Rht-D1*.

We identified *Qncb.PH-3D* as a novel plant height QTL, with a smaller effect than either *Rht-D1* or *Rht8* (Table 2.4). Despite the low marker density on chromosome 3D, *Qncb.PH-3D* was consistently localized to a 50-Mb interval on the long arm. *Rht-D1b* alters the function of an important component of the giberellic acid response pathway, so we may expect differential QTL effects in different *Rht-D1* backgrounds. We find that while *Qncb.PH-3D* was identified in all environments, the effect on plant height is much greater in a *Rht-D1a* (tall) background (Fig. 2.4). As SS-MPV57 is responsive to giberellic acid, the observed interaction between *Rht-D1* and *Qncb.PH-3D* will require further study, and may point to the identification of candidate genes for this QTL.

Three additional QTL (*Qncb.PH-1A*, *Qncb.PH-2B*, and *Qncb.PH-5B*) were also identified in one environment each, but when fit in the combined multiple QTL model all were

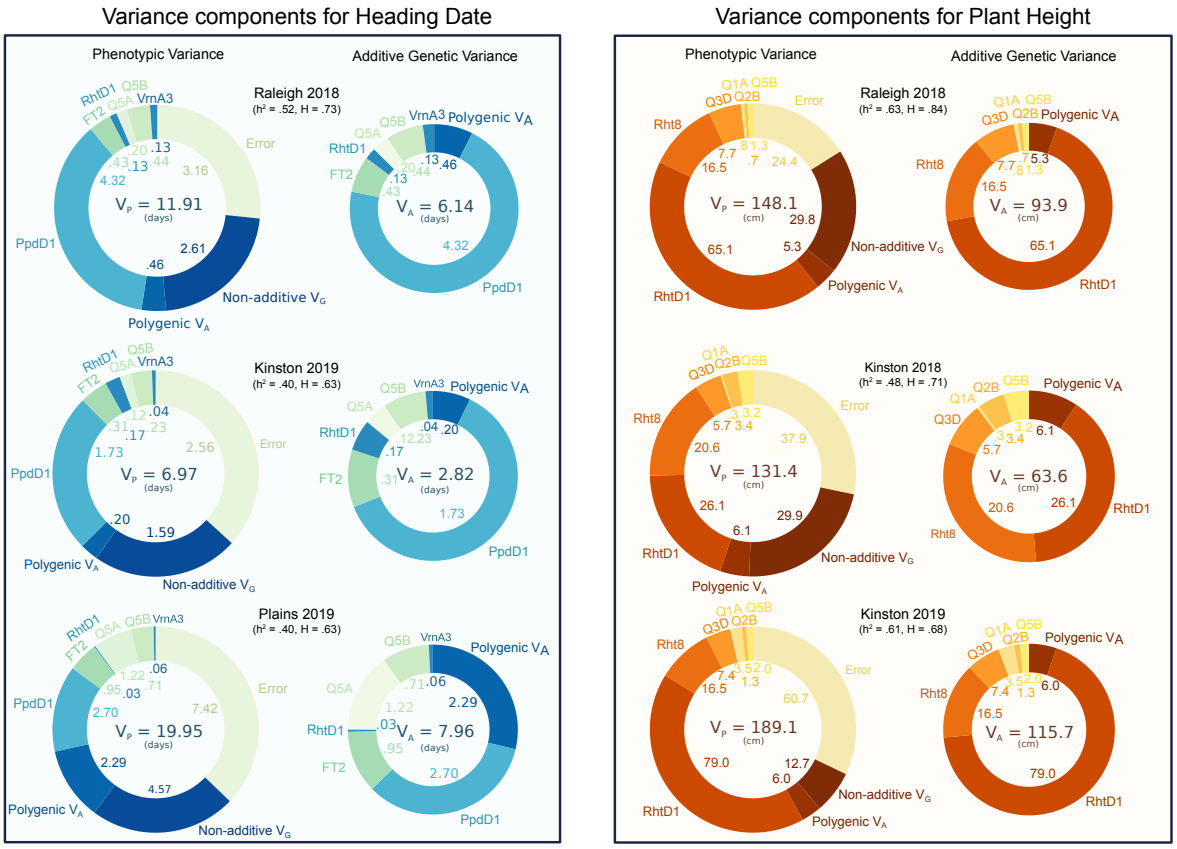

Figure 2.5: *Variance associated with QTL and variance components for heading date and plant height in multiple environments.* Non-additive genetic variation may be a result of epistatic interactions between QTL or mis-estimation of genotype values. *Ppd-D1* and *Rht-D1* dominate additive genetic variation for their respective phenotypes, but other mapped QTL explain a substantial portion of genetic variation. The scaling of total additive genetic variation is in large part due to the expression of *Ppd-D1* or *Rht-D1* effects.

significant with $p < .001$ in all environments (Additional file 2:Table S4).

2.4.4 QTL with Major and Moderate Effects Explain Most of Additive Genetic Variation and Generate Transgressive Segregation

Within-field phenotypic variance was partitioned in order to assess the genetic architecture of plant growth traits in this population and the relative importance of different mapped QTL in explaining observed differences (Fig. 2.5). For both heading date and adult plant height,

major effect QTL dominate additive genetic variation in most environments. Major-effect variant *Ppd-D1* was associated with a majority of additive genetic variation for heading date, except in the southern-most location of Plains, GA in 2019 (Fig. 2.5). In this environment, the polygenic additive genetic variation for heading date was similar to that associated with *Ppd-D1*. The modified architecture in a distinct environment suggests the presence of QTL with smaller effects conditional on photoperiod and vernalization signal. *FT2* and *Qncb.HD-5A* also increased in importance in the Plains 2019 environment, indicating that the effects of these moderate-effect QTL may also vary based on environmental conditions. Major-effect variant *Rht-D1* explained a majority of the additive genetic variation for plant height except in Kinston, NC in 2018 where *Rht8* explained a similarly sized proportion of variation (Fig. 2.5). The relative expression of these QTL in specific environments plays a large role in determining the observed variation both in genotype values and in phenotypes.

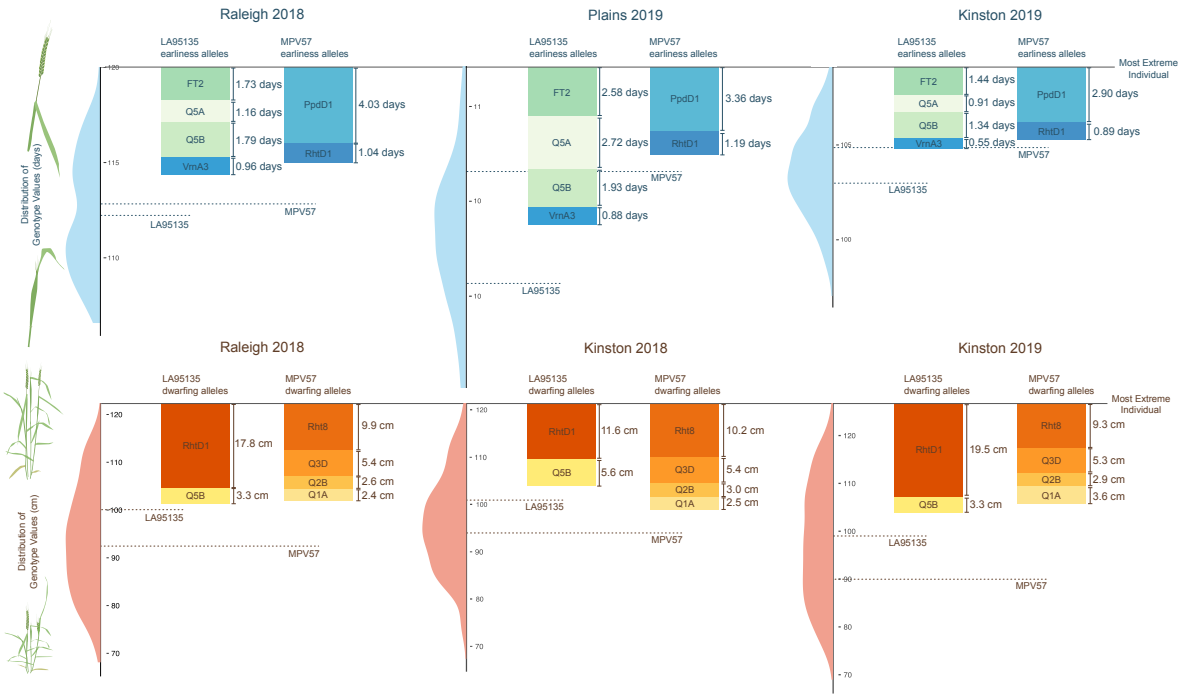

Figure 2.6: **Heading date and plant height characters of parental lines are mostly determined by major QTL.** For both heading date and plant height, the most phenotypically extreme individual was considered as the baseline for each environment and compared to both the distribution of genotype values and estimated QTL effects for the difference between two inbred lines (2α). Observed genotype values for the parental lines in each environment (dashed lines) are compared to the cumulative effects of their alleles.

A central question of this study is if variation in these plant growth traits is largely attributable to segregation of large-effect variants, or if identified variants are instead only contributing to largely polygenic differences in heading date and plant height. The plant height and heading date characters of the parental lines were found to be almost entirely determined by either major *Ppd-D1* and *Rht-D1* alleles, or cumulative effects of the stable, moderate-effect QTL identified in this study (Fig. 2.6). The transgressive segregation observed in this study, where both parents are phenotypically similar in terms of heading date and plant height, is being driven primarily by segregation of these major and moderate-effect QTL.

For heading date, the effects of *Ppd-D1* and *Rht-D1* were mostly sufficient to explain the observed phenotypes of SS-MPV57, and the phenotypes of LA95135 were mostly explained by the QTL effects of earliness alleles inherited from that parent (Fig. 2.6). In Raleigh in 2018, *Ppd-D1* has the largest effect, visible in the apparent bimodal distribution of genotype values. Plants in this environment experienced the coolest winter temperatures and had the latest mean heading dates (Table 2.1). The differences between the two parents is greatest in Plains in 2019, where the effect of *Ppd-D1* is relatively reduced and larger effects are observed for earliness alleles inherited from LA95135. Plants in this environment experienced the warmest winter temperatures and had the greatest range in heading dates.

For plant height, the effect of *Rht-D1b* largely determines the semi-dwarf character of LA95135, along with some contribution from novel plant height QTL on chromosome 5B. The semi-dwarf character of parent SS-MPV57 is largely generated by known dwarfing QTL *Rht8* and the novel QTL on chromosome 3D, with some contribution from novel QTL on chromosomes 1A and 2B.

2.4.5 QTL Models Out-Perform Genomic Selection for Oligogenic Traits

To assess the implications of the apparent oligogenic architecture of plant growth traits, a five-fold cross validation approach was used comparing a standard GBLUP model using genome-wide GBS markers to a simple multiple-regression QTL model based on previously estimated QTL effects (Table 2.5, Table 2.6).

Across all environments for both phenotypes, the simple QTL regression model is nearly as predictive as the top-performing model incorporating genome-wide marker information.

Table 2.5: Prediction accuracies for heading date. Mean prediction abilities and their standard deviations estimated from 40 replications of five-fold cross validations using QTL regression, GBLUP, QTL fixed effects plus GBLUP, Bayes B, and Bayesian Lasso models.

Model	**Ral18**		**Kin19**		**Pla19**	
	μ	**sd**	μ	**sd**	μ	**sd**
QTL Regres.	0.67	0.010	0.63	0.015	0.60	0.004
GBLUP	0.38	0.025	0.39	0.025	0.53	0.021
QTL/GBLUP	0.70	0.008	0.66	0.010	0.64	0.008
Bayes B	0.71	0.012	0.68	0.018	0.64	0.016
Bayes Lasso	0.63	0.018	0.58	0.024	0.63	0.022

Table 2.6: Prediction accuracies for plant height. Mean prediction abilities and their standard deviations estimated from 40 replications of five-fold cross validations using QTL regression, GBLUP, QTL fixed effects plus GBLUP, Bayes B, and Bayesian Lasso models.

Model	**Ral18**		**Kin18**		**Kin19**	
	μ	**sd**	μ	**sd**	μ	**sd**
QTL Regres.	0.77	0.004	0.69	0.007	0.79	0.005
GBLUP	0.24	0.032	0.35	0.024	0.26	0.032
QTL/GBLUP	0.80	0.005	0.74	0.007	0.81	0.006
Bayes B	0.79	0.007	0.71	0.009	0.79	0.007
Bayes Lasso	0.67	0.015	0.50	0.030	0.70	0.014

The GBLUP model commonly used in applied wheat breeding is comparatively ineffective in predicting heading date and especially plant height within the biparental population. Incorporation of genomic relationship information into the QTL regression model only offers slight performance increases compared to the base model, suggesting the genomic relationships do not add much additional information. The Bayes B model, designed to allow for marker effects of zero, performs the best for heading date (Table 2.5). For plant height, the GBLUP model with QTL fixed effects is superior (Table 2.6). In general, the Bayesian Lasso model is superior to the GBLUP model but inferior to the other models, except for heading date in Plains in 2019 where the relative proportion of additive genetic variation associated with the polygenic background was the highest.

2.4.6 Variation in plant growth is generated by major QTL

Plant height variation before maturity is caused in part by differences in development related to heading date variation, and thus may be controlled by QTL for both mature plant height and heading date. Multiple measures of plant height were collected from the RIL population planted in Raleigh during the 2019 field season, and a longitudinal model was used to estimate plant height over the measured time window. Identified heading date and adult plant height QTL were fit in a multiple regression model to estimate the proportion of phenotypic variation in plant height on a given day associated with each QTL. Variation in simulated genotype values were normalized by total QTL variation explained, and plotted over time to assess the relative importance of QTL in variation in plant height over time (Fig. 2.7).

As expected, the proportion of variation explained by the three adult plant height QTL (*Rht-D1*, *Rht8*, and *Qncb.PH-3D*) increases towards the end of the date range (March 29 to April 29, from near winter dormancy release to heading). Heading date QTL are more important than adult plant height QTL for early season plant height, when plants transition from vegetative to reproductive growth. The four heading date loci (*Ppd-D1*, *Qncb.HD-5A*, *FT-A2*, and *Vrn-A3*) continue to explain a large portion of variation in plant height as plants near heading, although their contribution diminishes as plants mature. Interestingly, the proportion of variation explained by QTL associated with *Vrn-A3* and *Rht8* were relatively consistent throughout development. *Rht-D1*, mapped as both a heading date and adult plant height QTL in this population, is associated with a large proportion of phenotypic variation throughout the date range.

2.5 Discussion

2.5.1 Unexplained Parental phenotypes result from novel QTL

Understanding the genetic basis of plant development is critical for understanding genetic variation for yield. In wheat, early flowering and plant height are understood to be largely determined by known large-effect variants; the mutations in the *DELLA* protein RHT1 (reduced height 1) on chromosomes 4B and 4D for plant height, and variation in

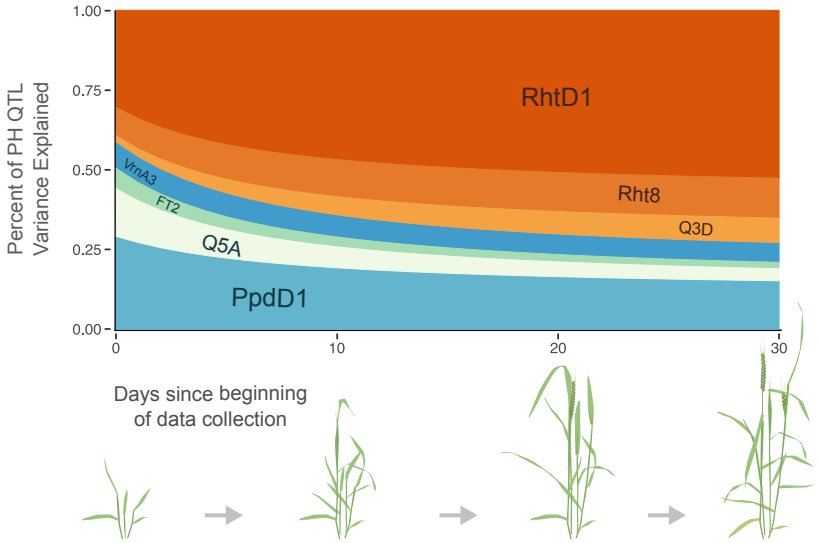

Figure 2.7: **Relative importance of QTL for plant height over time.** QTL associated with heading date (blue and green; *Ppd-D1*, *Qncb.HD-5A*, *FT2*, and *Vrn-A3*) explain over half of plant height variation associated with QTL at the beginning of data collection, but explain only approximately a quarter thirty days after data collecting began. The relative importance of plant height QTL (orange; *Rht-D1*, *Rht8*, and *Qncb.PH-3D*) increases over time.

vernalization-response *Vrn1* and photoperiod-response *Ppd1* genes for flowering time. Breeders generally select plants with plant heights and heading dates near some optimal values for their target environments, so that most cultivars have one of *Rht-B1b* or *Rht-D1b* but not both. In the southeastern US, most cultivars have some combination of early-flowering winter alleles of the *Vrn1* loci and one or more insensitive alleles of the *Ppd1* loci. Despite this, some cultivars with near-optimal values for heading date and plant height do not carry any known early flowering time or dwarfism alleles (for example, the two parents used in this study), and the relative importance of these major QTL versus other, smaller effect QTL in generating genetic variation for plant height and heading date is not known.

In other crop species such as maize, the majority of additive genetic variation in heading date and adult plant height is generated in a polygenic manner through the combination of many small-effect, unmapped QTL (Buckler et al. 2009; Peiffer et al. 2014). The importance of major-effect QTL in wheat (and other selfing species such as rice (Huang et al. 1996)) suggests that these traits may have a less polygenic basis in these species. Here we developed a biparental population by crossing cultivar LA95135, a cultivar with a normal flowering time but no early-flowering variants other than the weak photoperiod insensitive allele

Ppd-A1a.1, to SS-MPV57, a cultivar with a normal height but no known *Rht1* variants. Within this population, additive genetic variation for plant growth phenotypes emerges from known major-effect QTL and multiple novel moderate-effect QTL, instead of primarily from polygenic background effects of many small-effect QTL.

We find one plant height QTL on chromosome 2D, mapped distal to *Ppd-D1*, that likely represents *Rht8*. Cultivars having the *Rht8* dwarfing allele are responsive to gibberrellic-acid (Korzun et al. 1998), and the gene has a smaller effect on plant height than reported effects of *Rht-B1* and *Rht-D1*, in agreement with the allele effects estimated in this study (Rebetzke et al. 2012). Additionally, we propose newly characterized variants in genes *FT2* and *Vrn-A3* as candidates underlying QTL on chromosomes 3A and 7A, respectively. Additional novel plant height QTL were mapped to chromosomes 3D, 1A, 2B, and 5B, and additional heading date QTL to chromosomes 5A and 5B. When considered jointly, the effects of these QTL and *Ppd-D1* and *Rht-D1* are mostly sufficient to explain the phenotypes of the parental lines. Our ability to identify these novel QTL despite their comparatively small effect size may be attributable to the large population size, twice that of many winter wheat RIL populations.

Combining these moderate-effect QTL can produce plants with a short enough height and an early enough heading date. Like other non-*Rht1* dwarfing genes, *Rht8* and *Qncb.PH-3D* do not confer GA-insensitivity to SS-MPV57 (data not shown). A major limitation of the GA-insensitive *Rht1* genes is a reduced coleoptile length, which can lead to poor emergence and weak competition. While lines carrying *Rht8* alone are often too tall, semi-dwarf lines like SS-MPV57 produced by stacking *Rht8* with *Qncb.PH-3D* may perform better than *Rht1* semi-dwarfs in certain environments (Worland et al. 1998). The insignificant effect on plant height of *Qncb.PH-3D* in an *Rht-D1b* (semi-dwarf) background also reduces the potential of producing transgressive segregants that are too short from crosses between *Rht-D1b* and *Qncb.PH-3D*-dwarf cultivars. The position of *Qncb.PH-3D* distally on the long arm facilitates its fine-mapping, and identification of a predictive marker or the underlying causal polymorphisms will facilitate marker-assisted selection of this QTL in developing GA-sensitive semi-dwarf cultivars.

The use of major *Ppd1* and *Vrn1* variants in cultivar development also has associated drawbacks. In both cases, the early-heading character is a result of the plant losing its ability to receive signal from its environment – in wild-type photoperiod-sensitive genotypes, plants use information about changing night lengths to flower at an appropriate time, whereas photoperiod-insensitivity activates this pathway constitutively. Losing the ability

to respond to environmental cues may incur yield penalties in some situations. For example, autumn sown wheat lines with little or no vernalization requirement that are insensitive to photoperiod are susceptible to late spring freeze. However, requiring a long period of cold to potentiate flowering in environments with warm winters can result in delayed heading, even in lines having photoperiod insensitive alleles. This effect was observed in this study with the proportionally decreased effect of *Ppd-D1* in the Plains environment, which is farther south than the other locations and has shorter nights during the wheat growing season. A set of early flowering QTL with different environmental triggers or of more moderate effects, like those mapped here, provide breeders with additional tools to develop appropriate cultivars for various target environments. Fine-mapping and characterization of *Qncb.HD-3A*, *Qncb.HD-5A*, and *Qncb.HD-5B* will expand the flowering time toolbox for wheat breeders.

2.5.2 The oligogenic trait architecture of plant growth traits in wheat

In wheat, genetic variation in yield is dependent on yield components (e.g. kernel weight, kernel number per spike, spikelets per spike) that can be influenced by disease resistance, plant height, and heading date, among other traits. While yield variation itself is complex, this complexity may arise through a combination of variation in other traits which may not necessarily have a polygenic architecture. We observe only a small fraction of the total genetic variation for heading date and plant height in this population associated with lines' polygenic background. Even when not considering major-effect alleles *Rht-D1* and *Ppd-D1*, the remaining moderate-effect QTL explain more than twice the additive genetic variation as the polygenic background. While it is impossible to extend the results of this biparental study to wheat generally, the variation in heading date and plant height observed in this population is similar to the range of values observed in preliminary yield trials in breeding populations. It is likely the case that, while the particular variants differ from population to population, that the genetic architecture of plant height and heading date are similar across breeding populations in wheat.

2.5.3 Challenges and opportunities for genotype-based prediction of plant growth traits

The genetic architecture of plant growth has important implications for modern wheat breeding programs. Yield is the primary target of wheat breeders, and standard genomic selection models perform well for this trait in southeast U.S. wheat breeding programs (Sarinelli et al. 2019; Ward et al. 2019). Standard models shrink estimated effects of large-effect variants closer to zero, which will reduce accuracy of models for traits mostly conditioned by relatively few large-effect variants (Bernardo 2014). Given the effects of heading date and plant height variation on generating yield variation in wheat, if a handful of major QTL dominate these traits they may also have large effects on yield, complicating assumptions of these models. At the same time, heading date and plant height are themselves traits of interests to breeders, who screen biparental populations to remove transgressive segregants for these traits.

We show that the majority of additive genetic variation for heading date and plant height is controlled by large-effect QTL, such that a simple QTL model is sufficient for accurate prediction of phenotypes. In this case given marker information for major and moderate-effect QTL and a genotyped training population, a simple QTL model is likely to be effective for eliminating transgressive segregants for heading date and plant height. This model has the added benefit of being much cheaper than genomic selection if markers for polymorphisms linked to variants are available. Instead of genotyping a population of a set size with genome-wide markers, making predictions with genomic selection models, and then removing transgressive segregants for plant height and heading date, breeders can instead screen larger populations initially with simple makers for major QTL, and focus genotyping resources on lines predicted to be near optimal values for those secondary phenotypes. While QTL mapping was necessary to identify many important QTL for prediction in the QTL regression model in this population, this population was constructed specifically to segregate for novel heading date and plant height QTL. Our expanding knowledge of variants underlying these oligogenic traits results in the development of breeding populations where the major QTL will be known and predictions for heading date and plant height can be made. If we have genotypes for the causal polymorphisms underlying these QTL, we can make predictions in new populations regardless of their relationship to the training population lines. Fine-mapping and marker development for these and further novel QTL will then improve prediction models.

2.5.4 Plant growth QTL and variation for source traits

In the past few years, a number of variants impacting yield component traits that generate variation in sink tissues have been identified and cloned (Wang et al. 2018; Dixon et al. 2018; Kuzay et al. 2019; DeWitt et al. 2020). However, increasing the frequency of variants associated with larger grain size and number will only increase yields if plants produce sufficient carbohydrates ("source") to fill those grains. Similar characterization of important QTL underlying variation in physiological source traits will therefore also be critical to understand the components of yield variation. Variation in NDVI (normalized difference vegetation index) measurements or direct biomass samples, taken as proxies for source availability, is related to variation in the plant growth traits studied here. Both heading date and adult plant height can be viewed as components of the continuous phenotype of plant growth over time. While adult plant height is controlled by what are termed plant height QTL, juvenile plant height is often understood as winter dormancy release and is largely under the same genetic control as heading date (Guedira et al. 2014). To understand the genetic basis of plant growth in this population, we measured plant height over multiple days during development. We showed that variation in plant growth is influenced by a combination of heading date and plant height QTL. Studies of plant source traits may find it useful to consider phenotyping experiments for plant height and heading date as well to distinguish between QTL for heading date and plant height, and true source or biomass QTL. Understanding how plant height and heading date QTL interact to generate variation in plant growth over time will be critical to understanding how they impact source traits and in characterizing novel plant source QTL that can be deployed for higher yielding genotypes.

2.6 Conclusions

The polygenic nature of wheat yield results in part from major and moderate QTL for adaptation traits and other phenotypes that influence yield. It is therefore useful to consider and select for component phenotypes like disease resistance and plant growth traits that can influence yield separately, and to properly model these traits we need to first understand their genetic architectures. Already the simple genetic basis of many disease resistance genes has made MAS for disease resistance in wheat very useful to breeders, a success

story that could be replicated with plant growth traits given cost-effective predictions. Here, we show that component phenotypes of plant growth over time have an oligogenic basis dominated by QTL of major and moderate effect that allows for their prediction with simple QTL regression models. The movement towards genomic selection has called into question the utility of fine-mapping and positional cloning studies. We demonstrate the importance of major QTL and the poor performance of standard models in this study, illustrating the utility of understanding the important variants underlying these traits and others to crop improvement.

2.7 Competing interests

The authors declare that they have no competing interests.

2.8 Author's contributions

ND wrote the initial draft of the manuscript and performed all analyses. ND, MG, EL, JBH, and GBG edited the manuscript. MG, EL, and GBG designed the experiment and developed the population. ND, MG, JPM, DM, MM, and JJ planted the experiment and assisted with data collection and harvest. JBH assisted with data analysis and experimental design.

2.9 Acknowledgements

The authors thank staff of the USDA-ARS Plant Science Unit for assistance in genotyping and field evaluation of populations. Special thanks to Dr. Brian Ward, who gave valuable advice on many of the methods in this manuscript, Anna Rogers, who provided invaluable input on multiple aspects of the paper, and Kim Howell and Jared Smith, who extracted DNA and prepared GBS libraries. Support was provided by the Agriculture and Food Research Initiative Competitive Grant 67007-25939 (WheatCAP-IWYP) from the USDA NIFA.

References

Addison, C. K., Mason, R. E., Brown-Guedira, G., Guedira, M., Hao, Y., Miller, R. G., Subramanian, N., Lozada, D. N., Acuna, A., Arguello, M. N., Johnson, J. W., Ibrahim, A. M., Sutton, R., and Harrison, S. A. (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.

Akhunov, E. D., Akhunova, A. R., Anderson, O. D., Anderson, J. A., Blake, N., Clegg, M. T., Coleman-Derr, D., Conley, E. J., Crossman, C. C., Deal, K. R., Dubcovsky, J., Gill, B. S., Gu, Y. Q., Hadam, J., Heo, H., Huo, N., Lazo, G. R., Luo, M.-C., Ma, Y. Q., Matthews, D. E., McGuire, P. E., Morrell, P. L., Qualset, C. O., Renfro, J., Tabanao, D., Talbert, L. E., Tian, C., Toleno, D. M., Warburton, M. L., You, F. M., Zhang, W., and Dvorak, J. (2010). Nucleotide diversity maps reveal variation in diversity among wheat genomes and chromosomes. *BMC Genomics*, 11(1):702.

Beales, J., Turner, A., Griffiths, S., Snape, J. W., and 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.

Bernardo, R. (2014). Genomewide Selection when Major Genes Are Known. *Crop Science*, 54(1):68–75.

Borner, A., Worland, A. J., Plaschke, J., Schumann, E., and Law, C. N. (1993). Pleiotropic Effects of Genes for Reduced Height (Rht) and Day-Length Insensitivity (Ppd) on Yield and its Components for Wheat Grown in Middle Europe. *Plant Breeding*, 111(3):204–216.

Broman, K. W., Wu, H., Sen, S., and Churchill, G. A. (2003). R/qtl: QTL mapping in experimental crosses. *Bioinformatics*, 19(7):889–890.

Buckler, E. S., Holland, J. B., Bradbury, P. J., Acharya, C. B., Brown, P. J., Browne, C., Ersoz, E., Flint-Garcia, S., Garcia, A., Glaubitz, J. C., Goodman, M. M., Harjes, C., Guill, K., Kroon, D. E., Larsson, S., Lepak, N. K., Li, H., Mitchell, S. E., Pressoir, G., Peiffer, J. A., Rosas, M. O., Rocheford, T. R., Romay, M. C., Romero, S., Salvo, S., Villeda, H. S., Da Silva, H. S., Sun, Q., Tian, F., Upadyayula, N., Ware, D., Yates, H., Yu, J., Zhang, Z., Kresovich, S., and McMullen, M. D. (2009). The genetic architecture of maize flowering time. *Science*, 325(5941):714–718.

Butler, D. G., Cullis, B. R., Gilmour, A. R., Gogel, B. J., and Thompson, R. (2017). ASReml-R Reference Manual Version 4.

Chen, Y., Carver, B. F., Wang, S., Cao, S., and Yan, L. (2010). Genetic regulation of developmental phases in winter wheat. *Molecular Breeding*, 26:573–582.

Chen, Z., Cheng, X., Chai, L., Wang, Z., Du, D., Wang, Z., Bian, R., Zhao, A., Xin, M., Guo, W., Hu, Z., Peng, H., Yao, Y., Sun, Q., and Ni, Z. (2020). Pleiotropic QTL influencing spikelet

number and heading date in common wheat (Triticum aestivum L.). *Theoretical Applied Genetics*, page 3.

DeWitt, N., Guedira, M., Lauer, E., Sarinelli, M., Tyagi, P., Fu, D., Hao, Q., Murphy, J. P., Marshall, D., Akhunova, A., Jordan, K., Akhunov, E., and Brown-Guedira, G. (2020). Sequence-based mapping identifies a candidate transcription repressor underlying awn suppression at the B1 locus in wheat. *New Phytologist*, 225(1):326–339.

Díaz, A., Zikhali, M., Turner, A. S., Isaac, P., and 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).

Dixon, L. E., Greenwood, J. R., Bencivenga, S., Zhang, P., Cockram, J., Mellers, G., Ramm, K., Cavanagh, C., Swain, S. M., and Boden, S. A. (2018). TEOSINTE BRANCHED1 regulates inflorescence architecture and development in bread wheat (Triticum aestivum). *Plant Cell*, 30(3):563–581.

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., and Kawamoto, K. (2011). A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. *PLoS ONE*, 6(5):19379.

Endelman, J. B. (2011). Ridge Regression and Other Kernels for Genomic Selection with R Package rrBLUP. *The Plant Genome*, 4(3):250–255.

FAO (2020). *Crop Prospects and Food Situation #1, March 2020*. FAO.

Fischer, R. A. and Quail, K. J. (1990). The effect of major dwarfing genes on yield potential in spring wheats. *Euphytica*, 46(1):51–56.

Fu, D., Szűcs, P., Yan, L., Helguera, M., Skinner, J. S., Von Zitzewitz, J., Hayes, P. M., and Dubcovsky, J. (2005). Large deletions within the first intron in VRN-1 are associated with spring growth habit in barley and wheat. *Molecular Genetics and Genomics*, 273(1):54–65.

Gasperini, D., Greenland, A., Hedden, P., Dreos, R., Harwood, W. A., and Griffiths, S. (2012). Genetic and physiologcial analysis of Rht8 in bread wheat: an alternative source of semi-dwarfism with a reduced sensitivty to brassinosteroids. *Journal of Experimental Botany*, pages 4419–4436.

Glaubitz, J. C., Casstevens, T. M., Lu, F., Harriman, J., Elshire, R. J., Sun, Q., and Buckler, E. S. (2014). TASSEL-GBS: A High Capacity Genotyping by Sequencing Analysis Pipeline. *PLoS ONE*, 9(2):e90346.

Guedira, M., Maloney, P., Xiong, M., Petersen, S., Murphy, J. P., Marshall, D., Johnson, J., Harrison, S., and Brown-Guedira, G. (2014). Vernalization duration requirement in soft winter wheat is associated with variation at the VRN-B1 locus. *Crop Science*, 54(5):1960–1971.

Guedira, M., Xiong, M., Hao, Y. F., Johnson, J., Harrison, S., Marshall, D., and Brown-Guedira, G. (2016). Heading date QTL in winter wheat (Triticum aestivum L.) coincide with major developmental genes VERNALIZATION1 and PHOTOPERIOD1. *PLoS ONE*, 11(5).

Hedden, P. (2003). The genes of the Green Revolution.

Huang, N., Courtois, B., Khush, G. S., Lin, H., Wang, G., Wu, P., and Zheng, K. (1996). Association of quantitative trait loci for plant height with major dwarfing genes in rice. *Heredity*, 77(2):130–137.

Kippes, N., Guedira, M., Lin, L., Alvarez, M. A., Brown-Guedira, G. L., and Dubcovsky, J. (2018). Single nucleotide polymorphisms in a regulatory site of VRN-A1 first intron are associated with differences in vernalization requirement in winter wheat. *Molecular Genetics and Genomics*, 293(5):1231–1243.

Korzun, V., Röder, M. S., Ganal, M. W., Worland, A. J., and Law, C. N. (1998). Genetic analysis of the dwarfing gene (Rht8) in wheat. Part I. Molecular mapping of Rht8 on the short arm of chromosome 2D of bread wheat (Triticum aestivum L.). *Theoretical and Applied Genetics*, 96(8):1104–1109.

Kuzay, S., Xu, Y., Zhang, J., Katz, A., Pearce, S., Su, Z., Fraser, M., Anderson, J. A., Brown-Guedira, G., DeWitt, N., Peters Haugrud, A., Faris, J. D., Akhunov, E., Bai, G., and Dubcovsky, J. (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.

Li, G., Yu, M., Fang, T., Cao, S., Carver, B. F., and Yan, L. (2013). Vernalization requirement duration in winter wheat is controlled by TaVRN-A1 at the protein level. *Plant Journal*, 76(5):742–753.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16):2078–2079.

Nakagawa, S. and Schielzeth, H. (2013). A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, 4(2):133–142.

Nishida, H., Yoshida, T., Kawakami, K., Fujita, M., Long, B., Akashi, Y., Laurie, D. A., and Kato, K. (2013). Structural variation in the 5 upstream region of photoperiod-insensitive alleles Ppd-A1a and Ppd-B1a identified in hexaploid wheat (Triticum aestivum L.), and their effect on heading time. *Molecular Breeding*, 31(1):27–37.

Nishimura, K., Moriyama, R., Katsura, K., Saito, H., Takisawa, R., Kitajima, A., and Nakazaki, T. (2018). The early flowering trait of an emmer wheat accession (Triticum turgidum L. ssp. dicoccum) is associated with the cis-element of the Vrn-A3 locus. *Theoretical and Applied Genetics*, 131(10):2037–2053.

Peiffer, J. A., Romay, M. C., Gore, M. A., Flint-Garcia, S. A., Zhang, Z., Millard, M. J., Gardner, C. A., McMullen, M. D., Holland, J. B., Bradbury, P. J., and Buckler, E. S. (2014). The genetic architecture of maize height. *Genetics*, 196(4):1337–1356.

Pérez, P. and De Los Campos, G. (2014). Genome-Wide Regression and Prediction with the BGLR Statistical Package. *Genetics*, 198:483.

Poland, J., Endelman, J., Dawson, J., Rutkoski, J., Wu, S., Manes, Y., Dreisigacker, S., Crossa, J., Sánchez-Villeda, H., Sorrells, M., and Jannink, J. L. (2012). Genomic selection in wheat breeding using genotyping-by-sequencing. *Plant Genome*, 5(3):103–113.

Rebetzke, G. J., Ellis, M. H., Bonnett, D. G., Mickelson, B., Condon, A. G., and Richards, R. A. (2012). Height reduction and agronomic performance for selected gibberellin-responsive dwarfing genes in bread wheat (Triticum aestivum L.). *Field Crops Research*, 126:87–96.

Rebetzke, G. J. and Richards, R. A. (2000). Gibberellic acid-sensitive dwarfing genes reduce plant height to increase kernel number and grain yield of wheat. *Australian Journal of Agricultural Research*, 51(2):235–245.

Rife, T. W. and Poland, J. A. (2014). Field book: An open-source application for field data collection on android.

Sarinelli, J. M., Murphy, J. P., Tyagi, P., Holland, J. B., Johnson, J. W., Mergoum, M., Mason, R. E., Babar, A., Harrison, S., Sutton, R., Griffey, C. A., and Brown-Guedira, G. (2019). Training population selection and use of fixed effects to optimize genomic predictions in a historical USA winter wheat panel. *Theoretical and Applied Genetics*, 132(4):1247–1261.

Shaw, L. M., Lyu, B., Turner, R., Li, C., Chen, F., Han, X., Fu, D., and Dubcovsky, J. (2019). FLOWERING LOCUS T2 regulates spike development and fertility in temperate cereals. *Journal of Experimental Botany*, 70(1):193–204.

Taylor, J. and Butler, D. (2017). R package ASMap: Efficient genetic linkage map construction and diagnosis. *Journal of Statistical Software*, 79.

Wang, W., Simmonds, J., Pan, Q., Davidson, D., He, F., Battal, A., Akhunova, A., Trick, H. N., Uauy, C., and Akhunov, E. (2018). Gene editing and mutagenesis reveal inter-cultivar differences and additivity in the contribution of TaGW2 homoeologues to grain size and weight in wheat. *Theoretical and Applied Genetics*, 131(11):2463–2475.

Ward, B. P., Brown-Guedira, G., Tyagi, P., Kolb, F. L., Van Sanford, D. A., Sneller, C. H., and Griffey, C. A. (2019). Multienvironment and Multitrait Genomic Selection Models in Unbalanced Early-Generation Wheat Yield Trials. *Crop Science*, 59(2):491–507.

Worland, A. J., Korzun, V., Röder, M. S., Ganal, M. W., and Law, C. N. (1998). Genetic analysis of the dwarfing gene Rht8 in wheat. Part II. The distribution and adaptive significance of allelic variants at the Rht8 locus of wheat as revealed by microsatellite screening. *Theoretical and Applied Genetics*, 96(8):1110–1120.

Yan, L., Fu, D., Li, C., Blechl, A., Tranquilli, G., Bonafede, M., Sanchez, A., Valarik, M., Yasuda, S., and Dubcovsky, J. (2006). The wheat and barley vernalization gene VRN3 is an orthologue of FT. *Proceedings of the National Academy of Sciences of the United States of America*, 103(51):19581–19586.

Yan, L., Helguera, M., Kato, K., Fukuyama, S., Sherman, J., and Dubcovsky, J. (2004). Allelic variation at the VRN-1 promoter region in polyploid wheat. *Theoretical and Applied Genetics*, 109(8):1677–1686.

Youssefian, S., Kirby, E. J., and Gale, M. D. (1992). Pleiotropic effects of the GA-insensitive Rht dwarfing genes in wheat. 2. Effects on leaf, stem, ear and floret growth. *Field Crops Research*, 28(3):191–210.

Ziyatdinov, A., Vázquez-Santiago, M., Brunel, H., Martinez-Perez, A., Aschard, H., and Soria, J. M. (2018). lme4qtl: linear mixed models with flexible covariance structure for genetic studies of related individuals. *BMC Bioinformatics*.

Chapter 3

A network modeling approach provides insights into the environment-specific yield architecture of wheat

Noah DeWitt[1, 2] Mohammed Guedira[1] J. Paul Murphy[1] David Marshall[2] Mohamed Mergoum[3] Christian Maltecca[4] Gina Brown-Guedira[1, 2]

1 Department of Crop and Soil Sciences, North Carolina State University, Raleigh, NC 27695
2 USDA-ARS SEA, Plant Science Research, Raleigh, NC 27695
3 Department of Crop and Soil Sciences, University of Georgia, Athens, 30602, GA, USA
4 Department of Animal Science, North Carolina State University, Raleigh, NC, USA 27695

Minor Revisions in *Genetics*, February 2022

3.1 Abstract

Wheat (*Triticum aestivum*) yield is impacted by a diversity of developmental processes which interact with the environment during plant growth. This complex genetic architecture complicates identifying quantitative trait loci (QTL) that can be used to improve yield. Trait data collected on individual processes or components of yield have simpler genetic bases and can be used to model how QTL generate yield variation. The objectives of this experiment were to identify QTL affecting spike yield, evaluate how their effects on spike yield proceed from effects on component phenotypes, and to understand how the genetic basis of spike yield variation changes between environments. A 358 $F_{5:6}$ RIL population developed from the cross of LA-95135 and SSMPV-57 was evaluated in two replications at five locations over the 2018 and 2019 seasons. The parents were two soft red winter wheat cultivars differing in flowering, plant height, and yield component characters. Data on yield components and plant growth were used to assemble a structural equation model (SEM) to characterize the relationships between QTL, yield components and overall spike yield. The effects of major QTL on spike yield varied by environment, and their effects on total spike yield were proportionally smaller than their effects on component traits. This typically resulted from contrasting effects on component traits, where an increase in traits associated with kernel number was generally associated with a decrease in traits related to kernel size. In all, the complete set of identified QTL was sufficient to explain most of the spike yield variation observed within each environment. Still, the relative importance of individual QTL varied dramatically. Path analysis based on coefficients estimated through SEM demonstrated that these variations in effects resulted from both different effects of QTL on phenotypes and environment-by-environment differences in the effects of phenotypes on one another, providing a conceptual model for yield genotype-by-environment interactions in wheat.

3.2 Introduction

Wheat breeders have increased grain yield and stability by selecting for greater yield potential, despite little understanding of the genetic basis of that potential. Breeders and geneticists have successfully characterized the genetic basis of disease resistance and used marker-assisted selection (MAS) to deploy resistance in new cultivars, but the character-

ization and use of beneficial variants for yield potential has been less successful. One approach to disentangle the complexity of grain yield is to break it down into component traits (Grafius 1964; Brinton and Uauy 2019). In this view, per-acre wheat yield is a product of the number of spikes per square meter and spike yield (SY), which can be thought of as the product of kernel number per spike (KN) and thousand kernel weight (TKW) (Figure 3.1). Thousand kernel weight can be further decomposed into kernel area (KA), kernel width (KW), kernel length (KL), and kernel density). Kernel number per spike depends on the number of spikelets per spike (SpS), and the mean number of kernels per spikelet (KpS, related to floret fertility). Increased kernel number, rather than kernel weight, has driven increased wheat yield (Fischer 2008). Researchers understand that plants typically compensate for an increase in kernel number through a decrease in kernel size and vice-versa, but the extent to which this is driven by pleiotropy or physiology is unclear (Knott and Talukdar 1971). Loci with positive impacts on grain size with no kernel number tradeoff exist, suggesting it is possible to improve yield by identifying yield component QTL (Sadras 2007; Griffiths et al. 2015). Evidence that most modern wheat cultivars are sink-limited (Borrás et al. 2004), suggests that negative correlations between grain size and number may result from the mechanisms of underlying genes rather than resource constraints.

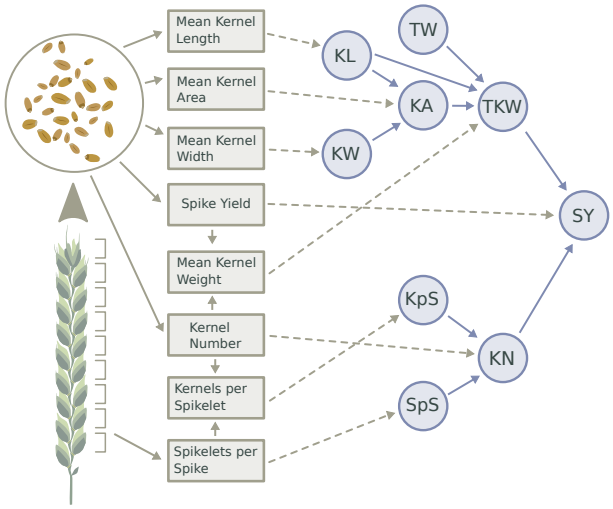

Figure 3.1: **Phenotyping of yield component traits in the LM population** The full set of phenotypes collected as in this figure is representative of the Raleigh NC, Kinston NC, and Plains GA locations in 2019. Collected phenotypes are related to the phenotypes within the conceptual yield component model, where phenotypes are taken to be the product of parent phenotypes, such that $SY = TKW * KN$, $TKW = TW * KW * KA$, and so on.

Selecting for increases in sink tissue is therefore a promising strategy for increasing yield. Mapping and cloning of the major variants influencing floral and grain morphology may facilitate prediction through MAS and improved GS models, but importantly should provide insight into the biology underlying phenotypic variation. The development of genomic resources in the wheat community has led to the mapping and cloning of well-studied wheat alleles with large effects on floral and grain morphology, including super-domestication gene *Q* controlling the spelt character (Simons et al. 2006), *S* controlling the semi-dwarf and spherical grain characters of *sphaerococcum* wheat (Cheng et al. 2020), and *B1* controlling the awnless character in most germplasm (DeWitt et al. 2020). More recently, new variants affecting yield components have been identified and cloned; *TaGW-A2* has been characterized as a functional variant in a negative regulator of grain weight (Su et al. 2011; Wang et al. 2018), while a major variant in floral developmental gene *WAPO-A1* increases the number of spikelets per spike (Kuzay et al. 2019).

Variation in plant growth also impacts yield and its components. Major genes have been identified that control adult plant height and heading date. The physiological effect of the *Rht* genes, mutations in DELLA proteins that greatly reduce plant height, is conditional on the environment and the quantity of assimilate produced by the cultivar. *RhtD1* is associated with larger kernel number but smaller kernel size and weight (Borner et al. 1993), a trade-off related not to spike development but to the greater availability of assimilates during grain-fill due to partitioning less biomass into stems (Youssefian et al. 1992). Semi-dwarf wheat cultivars having either the *Rht-D1a* or *Rht-B1a* allele are widely cultivated, as breeders generally select plants near some optimal height, because too-short plants have a generally lower yield (Fischer and Quail 1990) and tall plants tend to lodge. Variation for heading date generated by major-effect variants underlies the broad geographic distribution of cultivated wheat. Heading date variation alters the duration plants spend in certain developmental stages, which can impact developmental traits like kernel size and spikelet number per spike (Vargas et al. 2007). Plants with a mean heading date outside of the optimal range for a specific environment may expose themselves to early-season cold stress, late-season heat stress, or drought stress at sensitive developmental stages. Previous studies have demonstrated major effects of *Vrn1* and *Ppd1* alleles on final grain yield, often in an environment-specific manner related to climatic conditions (Addison et al. 2016, unpublished results). A variant may have a large effect on a single plant growth phenotype that will translate into a smaller effect on yield component traits and, in turn, to yield variation. For example, the insensitive allele at the major heading date gene *Ppd-D1*, alters

expression of intermediates that up-regulate other genes in the flowering time pathway, hastening the development of the shoot apical meristem which in turn reduces the number of spikelets per head (Boden et al. 2015). Decreases in spikelet number may decrease spike yield, depending on the relationships between yield components in a given environment. Variation for these developmental traits then alters yield in an environment-specific manner. Using QTL for the improvement of yield requires an understanding of how those QTL alter yield. A "yield QTL" useful for MAS should alter yield components in a way that consistently produces an increase in yield in the targeted environment without altering plant growth over time.

The structural equation modeling (SEM) framework fits multivariate data to a network structure that links variables via effects. Path analysis and SEM have been used in a variety of fields (e.g. Wright 1920; Crespi and Bookstein 1989; Kaplan and Phillips 2006). In the context of crop genetics, they can model the translation of sequence variation into yield variation through intermediate phenotypes. SEM has been recently used to increase the power of GWAS by incorporating information on multiple correlated traits (Momen et al. 2018, 2019). In wheat, SEM has been used within chromosome substitution lines to identify QTL for yield via effects on grain number and weight (Dhungana et al. 2007; Mi et al. 2010). At the same time, SEM has been used to show the importance of variation in plant growth traits in generating variation in seed number, size, and yield in small numbers of genotypes (Vargas et al. 2007). We chose to apply SEM in this study because, unlike standard multi-trait models that assume direct effects of QTL on all traits, SEM allows linking QTL to spike yield via a graph of intermediate phenotypes with a known structure to understand the basis of changes in QTL effects on spike yield between environments.

In DeWitt et al. (2021), we characterized the genetic basis of heading date and plant height in the LA-95135 x SS-MPV-57 biparental RIL population. These traits follow an oligogenic architecture, with over 90% of additive genetic variation for heading date and plant height associated with six QTL in most environments. Here we build off this prior work to understand the relationship between these QTL and yield, characterize new QTL that are associated with yield component traits but not heading date or plant height, and understand the basis of their contribution to spike yield variation in an environment-specific manner.

3.3 Materials and methods

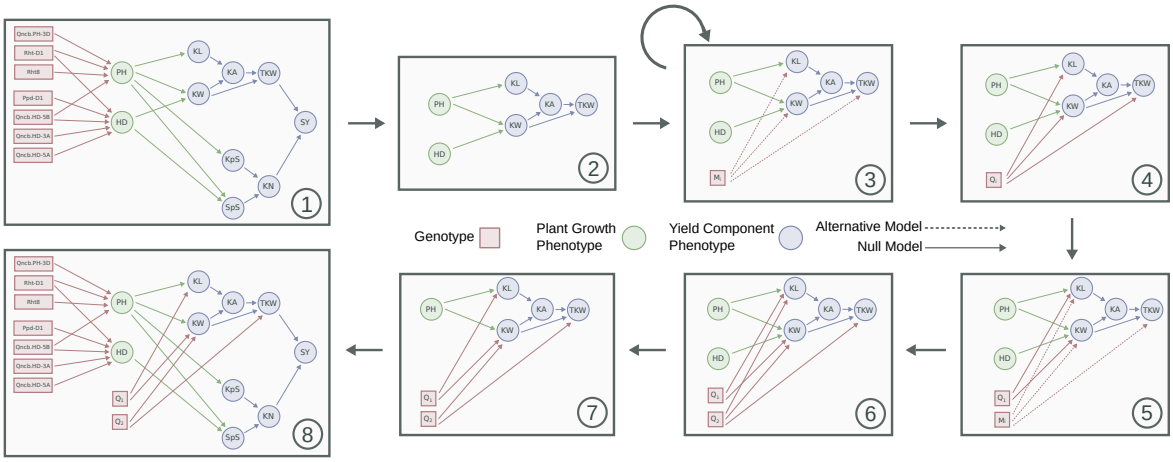

Figure 3.2: **SEM QTL Mapping Model** Steps in the QTL mapping model. The iteration for QTL having effects on TKW, KL, and KW is shown.

3.3.1 Population Development and Phenotyping

A recombinant inbred lines (RIL) population developed from a cross between cultivars SS-MPV57 (FFR555W/3/VA89-22-52/TYLER//REDCOAT*2/GAINES) and LA95135 (CL-850643/PIONEER-2548//COKER-9877/3/FLORIDA -302/COKER-762) was used in this study. The parents of the population were chosen to generate major additive genetic variation for plant height and flowering time and to characterize novel QTL for these traits. Parent SS-MPV57 carries major earliness gene *Ppd-D1a*, but no major dwarfing genes, and parent LA95135 caries major dwarfing gene *Rht-D1a*, but no major earliness genes. At the same time, recently cloned major variants *WAPO-A1*, *TaGW-A2*, and *B-A1* (Supplementary Table 1) segregate in this population, allowing for the investigation of their effects on different yield components in a common background (Kuzay et al. 2019; Su et al. 2011; Wang et al. 2018; DeWitt et al. 2020). After crossing, F_1 plants were selfed to generate a population of F_2 lines. Plants were advanced via the single-seed descent method until the F_5 generation, producing 358 RILs.

The LM RIL population was evaluated in the field at Raleigh, NC (RAL18) and Kinston,

NC (KIN18) during the 2018 field season, and in 2019 at Raleigh (RAL19), Kinston (KIN19), and Plains, GA (GA19). The 358 RILs were grown in an augmented block experiment with three replications of the parents per block, and two replications of each RIL at each location. Each of the five blocks consisted of 72 RIL entries, and plots consisted of 1-m rows spaced 30 cm apart. Plant height was collected as the height of the top of the spikes, excluding the awns, of a sample of tillers from the center of each row,. Heading date of each row was noted as the day of year when half of the heads had fully emerged from the flag leaf, typically a few days before anthesis.

Grain yield in this study is treated as the product of kernel number and size (Figure 3.1). Kernel size was collected as thousand kernel weight (TKW) and kernel size components (kernel length, KL; kernel width, KW), while spikelet number per spike (SpS) and kernel number per spikelet (KpS) were collected as components of kernel number. Six representative spikes from each row were harvested and number of spikelets recorded for all field experiments. Seeds were weighed and counted to determine kernel weight. Rows at the RAL18 location were hand harvested and threshed using a Vogel thresher. A 15 ml seed sample was cleaned and grain morphometric parameters (kernel length, width, area and weight) were obtained using a MARVIN grain analyzer (GAT Sensorik GMBH, Neubrandenburg, Germany). Kernel morphometric parameters and total spike yield for the KIN18 experiment and all 2019 field experiments were obtained by sampling seed threshed from the six-spike sample of each row using a scale and a Vibe QM3 grain analyzer (Vibe Image Analytics Ltd., Capitola, CA). In all locations, SpS was calculated as the mean number of fertile spikelets on the sampled wheat spikes.

3.3.2 Analysis of Phenotypes

For QTL mapping, values for each genotype were calculated for each combination of phenotype and environment as best linear unbiased estimates (BLUEs). To correct for effects related to plants' physical positions within a field, the software Echidna (Gilmour 2018) was used to calculate BLUEs with an AR1xAR1 spatial adjustment model:

$$y_i \sim \mu + G_i + u_i + e_i$$

Where y_i is the observed phenotype for an individual row, μ is the intercept, G_i is the fixed effect of genotype, and u_i is the unit or "nugget" random effect for each observation representing the component of the variance associated with sampling or other uncorrelated error drawn from a distribution $u \sim iidN(0, \sigma_e^2)$, and e_i is the spatially-correlated residual drawn from the distribution $e \sim N(0, \sigma_e^2 \Sigma_r(\rho_r) \otimes \Sigma_c(\rho_c))$, whose variance is the direct product of an rxr auto-correlation matrix $\Sigma_r(\rho_r)$ representing rows and a cxc correlation matrix $\Sigma_c(\rho_c)$ representing columns. For structural equation modelling, phenotypes were centered and scaled to a mean of zero and standard deviation of one.

3.3.3 Network Modeling

The structure of the conceptual yield component model (Figure 3.1) was extended to estimate allele effects of QTL on spike yield as the product of effects on component phenotypes by fitting a Structural Equation Model using the R package lavaan (Figure 3.2). In this model, a QTL effect on spike yield is the sum of the products of effects on the base yield components (KL, KW, KpS, SpS) and the mediated effects of the QTL on downstream phenotypes via the base phenotypes. Correlation coefficients between linked variables are taken as causal effects of variables on other variables using prior knowledge of the relationships between phenotypes. This model was extended with information on phenotypes that may affect yield components – plant height, heading date, and leaf rust severity – to generate a graph relating QTL affecting these traits to spike yield through a set of intermediate phenotypes (Figure 3.2).

The effects of QTL on traits and the effect of traits on each other were estimated by fitting the graph in each environment in lavaan. Yield component traits were allowed to have effects on the other yield components that were set later in plant development (for example, altered values of spikelets per spike (SPS) can affect kernel length (KL), but not vice-versa). The total allele effect of a QTL on spike yield was calculated using path analysis, multiplying the effects of the QTL on associated phenotypes by the effects of those phenotypes on subsequent phenotypes, and summing all paths.

3.3.4 QTL Mapping

A linkage map was constructed from GBS and KASP markers as described in DeWitt et al. (2021). QTL mapping of plant height and heading date genes was performed as described previously using the R package r/QTL (Broman et al. 2003). Composite interval mapping was used for initial QTL identification, and intervals were narrowed using a multiple QTL model (MQM) as implemented in the *refineqtl()* function. The *addqtl()* function was used to identify additional QTL using identified QTL as covariates. Significance thresholds for an $\alpha = 0.05$ were determined using 1000 permutations of a given test.

Given the degree of correlation between our phenotypes, most identified yield QTL are expected to have pleiotropic effects on more than one yield component trait. Thus, a multivariate model that included components was expected to have more power to detect a QTL for spike yield than a traditional QTL model. In addition, to identify yield component QTL that are not heading date or plant height QTL with mediated effects on yield components, those phenotypes should be considered as covariates in the analysis. We implemented a network mapping approach to account for the observation that each specific QTL does not have pleiotropic effects on all phenotypes, but affect different subsets of traits. To model the existence of "per-se" yield component QTL the effect of markers in a network model was tested directly using the following steps:

1. Specify a prior graph featuring effects of known genes on phenotypes, and relationships between phenotypes (Figure 2, panel 1).
2. Test each combination of base yield component phenotypes, starting with the maximal case of all five, and moving to each combination of four, then each combination of three, and so on. For each set of phenotypes:
3. Only consider the variables in the graph that affect the current set of phenotypes under consideration (Figure 2, panel 2).
4. Generate 1000 random permutations of the genotype data (Figure 2, panel 3). For each permutation:
    (a) For each permuted marker, compute the likelihood of the null model of no effect of the tested marker on the base yield components using the *sem()* function in lavaan. Compute the likelihood of the alternative model of an effect of the tested

marker on the base yield components, and take the difference as a log of odds (LOD) score.

(b) Return the LOD value of the permuted marker with the largest LOD score as representing the iteration.

5. Order the vector of maximum LOD values and take the 95th percentile value as a conservative empirical significance threshold at $\alpha = .05$.

6. For each non-permuted marker, compute and compare the likelihoods of alternative and null models using the sem() function in lavaan. Take the model with the highest LOD score as an updated model for that set of phenotypes (Figure 4 panel 4).

7. Refit the model with the added marker and remove any non-significant paths between the new marker and phenotypes (Figure 2, panel 5)

8. Remove any other paths that have become non-significant because of addition of new variables.

9. Repeat steps a-f until no further markers pass the LOD threshold (Figure 2, panels 6-7).

10. Replace the sections of the full graph considered in the current iteration with the updated subgraph containing any new QTL and their relationships to phenotypes (Figure 2, panel 8).

Genotype permutations were performed with the *sim.geno()* function in r/QTL. Number of QTL used in significance threshold testing was limited to 8 to prevent over-fitting of the model and maintain a conservative significance threshold, and mapped QTL within 2 cM of an existing QTL were treated as the same QTL affecting different phenotypes. The procedure was repeated separately for each environment. In the Plains 2019 environment where plant height was not collected, plant height was modeled as a latent variable measured by the plant height variables collected in the other four environments. The final model for each environment was fit using the *sem()* function in lavaan to estimate coefficients for path analysis of the effects of individual QTL on spike yield and its components.

3.3.5 Variance and Path Analysis

To assess the relative importance of mapped QTL in generating additive genetic variation in collected phenotypes, variance analysis was performed in the R package lme4qtl, which allows for the fitting of random effects with supplied covariance matrices (Ziyatdinov et al. 2018). For known variants for which KASP marker genotypes of the causal polymorphisms were available, the genotypes were used directly, and for novel QTL genotype probabilities from the refineqtl object were used. For testing QTL, alleles were encoded in terms of the allele dosage of the LA95135 allele (0, 1, 2) without estimating a dominance effect.

For each environment and phenotype, QTL effects and variance components for both the additive and non-additive effects of genotypes were specified with the mixed model:

$$Y_{ik} \sim \sum_{h=1}^{n} Q_{ih} + g_{Ai} + g_{Ei} + e_{ik}$$

Where for each phenotype Y of genotype i in row k, fixed effects for each QTL h were fit as regressions of allele dosage on phenotypes. The term g_{Ai} represents the the random additive effect of genotype i with a variance specified by the realized relationship matrix ($g_A \sim \mathcal{N}(0, \mathbf{G}\sigma^2_g)$), calculated from the scaled GBS marker matrix ($\mathbf{G} = \frac{\mathbf{W}'\mathbf{W}}{c}$, where $\mathbf{W}$ is the scaled marker matrix and c is a normalization value calculated from marker frequencies). The term g_{Ii} represents the non-additive random effect of genotype i with an independent variance ($g_I \sim \mathcal{N}(0, I\sigma^2_g)$).

A modified method from Nakagawa and Schielzeth (2013) used in DeWitt et al. (2021) was used to estimate variances associated with QTL and variance components from the specified model. Variances associated with fixed effects in a mixed linear model were estimated by multiplying the vector of values for an effect by its estimated effects and taking the variance of the vector ($\sigma^2_{fixed} = Var(\sum_{h=1}^{n} \beta_h x_{hk})$ for coefficients and effects h and observations k). Linked QTL *Ppd-D1* and *Rht8* were treated as a single locus to satisfy assumptions of independence between variables, using the mean genotype probabilities of the two QTL.

The above model was modified to estimate variance components of spike yield by collectively estimating two additional random genotype effects associated with yield com-

ponents and plant growth (heading date and plant height) QTL. Relationship matrices were calculated as above except that genotypes for plant height and heading date QTL were used to calculate a relationship matrix to estimate the variance associated with plant growth QTL, and genotypes for yield component traits identified using the SEM model were used to calculate a relationship matrix to estimate variance associated with yield component QTL. To verify that this model was appropriately partitioning variation, 100 iterations of a null model were run in which the same number of randomly sampled genome-wide SNP were used to construct the plant growth and yield component relationship matrices.

3.4 Results and discussion

3.4.1 Initial QTL Identification

To investigate the genetic architecture of SPS and TKW in this population, single-trait QTL mapping was performed in the RAL18 and KIN18 environments. Twelve QTL were significantly associated (LOD > 3.54) with SpS, and 8 QTL were significantly associated (LOD > 3.46) with TKW. Based on peak markers and confidence intervals, multiple QTL (Table 3.1) co-located with loci that were previously determined to affect plant height and heading date in the population (DeWitt et al. 2021). The *Rht-D1b* dwarfing allele was the major QTL for TKW and in KIN18 had a positive effect on SpS. Photoperiod insensitivity gene *Ppd-D1a* decreased SpS in both environments. Previously identified earliness variants on chromosomes 3A and 5A *(Qncb.HD-3A, Qncb.HD-5A*) were co-located with QTL for fewer SPS, and a variant in Vrn-A3 affecting heading date is co-located with a QTL for TKW on the short arm of chromosome 7A.

Recently characterized variants in wheat genes *Tb-B1*, *B-A1*, *WAPO-A1*, and *TaGW-A2* were predicted to segregate based on KASP marker genotypes of the parents (Supplementary Table 1). Yield component QTL were mapped near each of these genes. For example, predictive marker for the *B-A1* locus determining presence or absence of awns on the spike was the peak marker for TKW and SpS on the long arm of chromosome 5A (Table 3.1). In addition, a predictive marker for *Lr9* conferring resistance to leaf rust (caused by *Puccinia triticina*) was significant also for both TKW and SpS at KIN18, where plants were affected by a naturally occurring leaf rust epidemic. Beyond these known variants, QTL affecting

Table 3.1: **Significant Yield Component QTL Information from Best Environment.** Data on thousand kernel weight (TKW) and spikelets-per-spike (SPS) were collected in Raleigh and Kinston in 2018. For QTL significant in two environments, the effects, LOD score, and confidence interval from the most significant environment are displayed. QTL for other phenotypes mapped near the confidence interval for a trait are given. Many QTL are mapped in the same regions as QTL for plant height (PH) and heading date (HD). QTLs mapped in the same region for both SPS and TKW increase one while decreasing the other.

SpS: Chr.	**Gene**	**Peak Marker**	**Position CI**	**LOD**
2A		S2A_223722835	79-661 Mb	5.89
2B		S2B_612673903	147-654 Mb	3.88
2D	*Ppd-D1*	*Ppd-D1*	32-44 Mb	15.07
3A		S3A_609909640	485-621 Mb	4.64
3D		S3D_476608044	434-549 Mb	3.58
4B	*Tb-B1*	S4B_416305737	52-620 Mb	3.71
4D	*Rht-D1*	*Rht-D1*	0-35 Mb	12.1
5A.1		S5A_480804629	480-494 Mb	8.62
5A.2	*B-A1*	*B1*	698-703 Mb	12.9
6B	*LR9*	*LR9*	701-721 Mb	6.45
6D		S6D_250410640	55-323 Mb	4.22
7A	*WAPO-A1*	S7A_673096342	672-675 Mb	23.46
TKW: Chr.	**Gene**	**Peak Marker**	**Position CI**	**LOD**
3D		S3D_496955493	477-530 Mb	3.58
4D	*Rht-D1*	*Rht-D1*	0-35 Mb	37.76
5A.1		S5A_439551041	331-455 Mb	6.77
5A.2	*B-A1*	*B1*	698-702 Mb	5.59
6A	*TaGW-A2*	S6A_355433083	97-509 Mb	5.81
6B	*LR9*	*LR9*	701-721 Mb	11.42
7A.1	*VRN-A3*	S7A_66220298	58-77 Mb	4.75
7A.2		S7A_149426090	128-516 Mb	5.51

SPS were identified on chromosomes 2A, 2B, 3D, and 6D, and a QTL on chromosome 7A was identified for TKW (Table 3.1).

Of the 20 QTL, seven were significant in both locations (Table 3.1); however, the relative proportion of additive genetic variation associated with QTL varied between locations (Figure 3.3). No individual QTL increased both SPS and TKW, but in some the magnitude of

the negative effect on one phenotype was much smaller than the magnitude of the positive effect on the other.

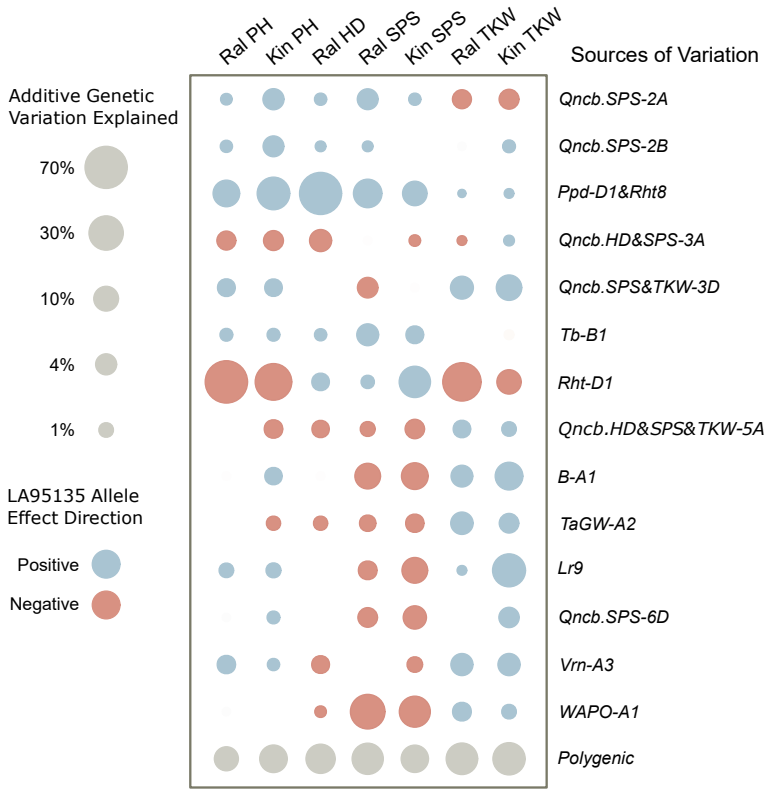

Figure 3.3: **Additive genetic variation associated with yield component QTL.** Additive genetic variation for phenotypes collected in Raleigh and Kinstson in the 2018 field season (PH; plant height, HD; heading date, SPS; spikelets-per-spike, TKW; thousand kernel weight)is partitioned into components associated with mapped SPS and TKW QTL and all other QTL (polygenic effect). QTL that have a positive effect on SPS generally have a negative effect on TKW, and vice versa.

3.4.2 Network Mapping of Yield Component QTL Using SEM

During 2019, data was collected for correlated morphometric traits KpS, KL, and KW in addition to SpS and TKW (Supplementary figure 1). Considering the analysis of the yield components in 2018, a SEM was fit that utilized plant height and heading date in each environment to control for effect of plant growth genes on spike yield and allowed QTL to

alter multiple yield component traits pleiotropically. In KIN19 and GA19, where leaf rust was observed and individual plots scored for disease severity, leaf rust score was added in as an additional variable. The network mapping model identified 53 QTL in KIN19 (Supplemental Figure 2, Figure 3.4), 62 QTL in RAL19, and 57 QTL in GA19 (Supplementary Table 3). In this model, QTL for TKW and SpS co-located with plant height or heading date loci were generally not detected. The exceptions were *Rht-D1* and *Ppd-D1*, suggesting they may have both direct and indirect effects on yield components. All other QTL identified for TKW and SpS in 2018 were identified by the model at one or more locations in 2019.

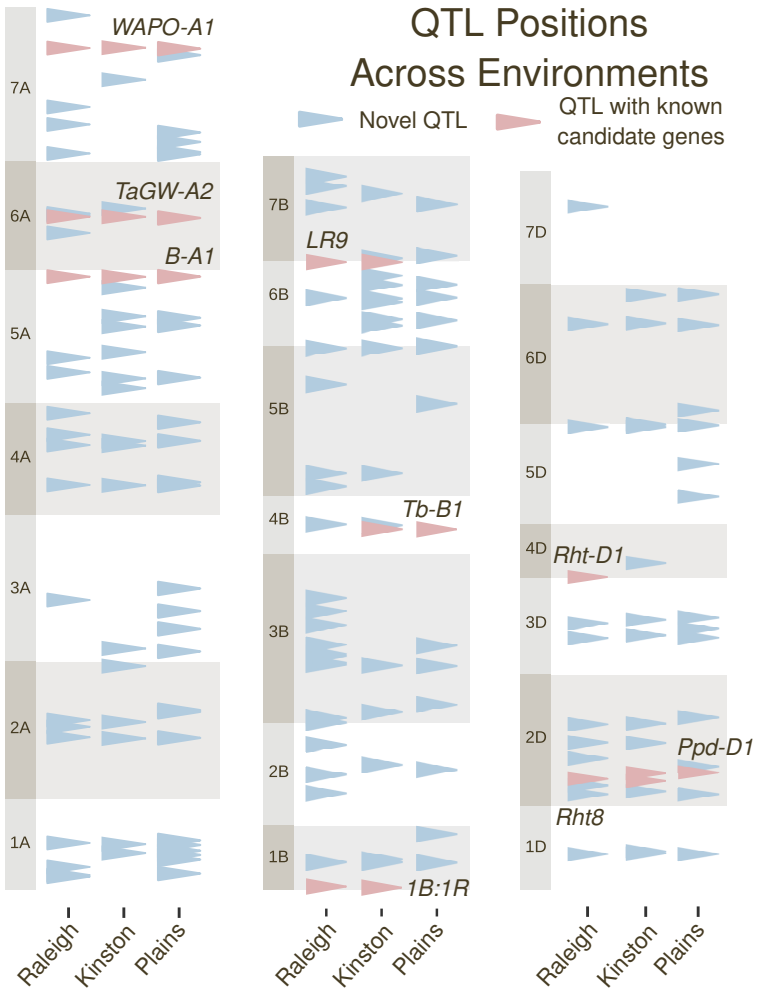

Figure 3.4: **Mapped QTL across locations** QTL added to the model for any combination of traits, plotted by genetic position across all three subgenomes. QTL associated with a known gene are highlighted in red, whereas novel QTL are in blue.

3.4.3 Path Analysis of Phenotypes

Correlations between phenotypes may arise due to underlying genes with pleiotropic effects, linkage between loci, or causal relationships between traits. The correlation between heading date and plant height was insignificant ($p > .05$) after including the effects of QTL affecting both traits, implying that the correlation of unadjusted phenotypes resulted from pleiotropic effects or close linkage between underlying genes (ie. *Ppd-D1* and *Rht8*). Compensation is often observed in wheat, resulting in negative correlations between base yield component phenotypes KL, KW, KpS, and SpS. While these negative correlations can arise due to pleiotropic effects of the underlying genes, they may also result from source limitations. For example, in an environment or genetic background where carbohydrate availability limits grain yield, increasing the number of spikelets per spike may decrease the numbers of kernels per spikelets, and increasing kernel number may reduce mean kernel size.

Structural equation models allow for the estimating the mediating effects and allow analysis of these relationships for individual loci. Loci identified by network mapping of QTL were fit in a full SEM for two locations where all developmental and yield component traits were evaluated (RAL19 and KIN19) to estimate phenotypic relationships after correcting for direct QTL effects (Figure 3.5). As yield component traits are plastic at different developmental stages, only effects of base yield component traits on subsequently developing traits were estimated. For example, increasing SPS may decrease kernel size but not vice versa, as spikelet number is set prior to grain fill. As kernel density was not directly measured in this experiment, plant growth phenotypes and mapped QTL were allowed to affect TKW directly.

Generally positive relationships of similar effect were observed between kernel measurements (KL, KW, TKW) and SY in both environments, with larger kernels contributing positively to SY, particularly at KIN19 (Figure 3.5). As expected, there were also strong positive relationships between yield components KN and TKW and their sub-components. Comparison of the estimated graphs indicated differences in the relationships between developmental traits and base yield components at each environment. At RAL19, there was little evidence for source limitation, although a slight negative effect of increasing spikelets per spike on kernel width was observed. In comparison, the KIN19 environment appeared more source-limited, with negative relationships between SpS and KpS, and

strong negative effects of increasing kernel number on kernel size components KL and KW (Figure 3.5). Consistent with the understood physiology of spike development, increasing heading date increased the number of spikelets per spike in both locations. Later heading was also associated with increased KW at RAL19. However, a slight decrease in KW was observed at KIN19 where plants experienced higher temperatures during grain fill. In addition, decreasing plant height was associated with increasing KpS at KIN19. These effects were consistent with the appearance of source constraints in Kinston as suggested by the negative relationships between base yield components. Increased leaf rust severity (RS) decreased KW and KL, but had a slightly positive effect on KpS. Taken together, this analysis indicates that variability in relationships among phenotypes across environments provided a mechanism for QTL to have environment specific effects on spike yield, even while having consistent direct effects on the primary phenotype.

3.4.4 Genetic architecture of spike yield

Additive genetic variation in spike yield and its components was largely associated with identified QTL, but the relative importance of those QTL varied from across environments (Figure 3.6). When the yield component or plant height relationship matrices were populated with random markers instead of genotypes for identified QTL they did not capture variance for the phenotypes. The QTL with the overall largest contribution to additive genetic variation was *Rht-D1*. At RAL19 and GA19, *Rht-D1* explained with XX% of additive genetic variation in SY, was primarily associated with TKW and explained little additive genetic variation in KN. In contrast, at KIN19 *Rht-D1* was associated with 54.6% of additive genetic variation for TKW and 34.0% for KN, and 67.7% of additive genetic variation in spike yield. This is consistent with the large effect of plant height on both KW and KpS at KIN19 observed in the path analysis (Figure 3.5).

Resistance gene *Lr9* showed a pattern of environment-specific expression consistent with the presence of disease. Leaf rust pressure was highest at GA19 where *Lr9* was associated with 34.8% of additive genetic variation in SY, moderate at KIN19 where *Lr9* was associated with 5.6% of variation, and absent at RAL19 where the locus explained 1.6% of variation.

None of the per-se yield components had effects comparable to *Rht-D1* on spike yield in

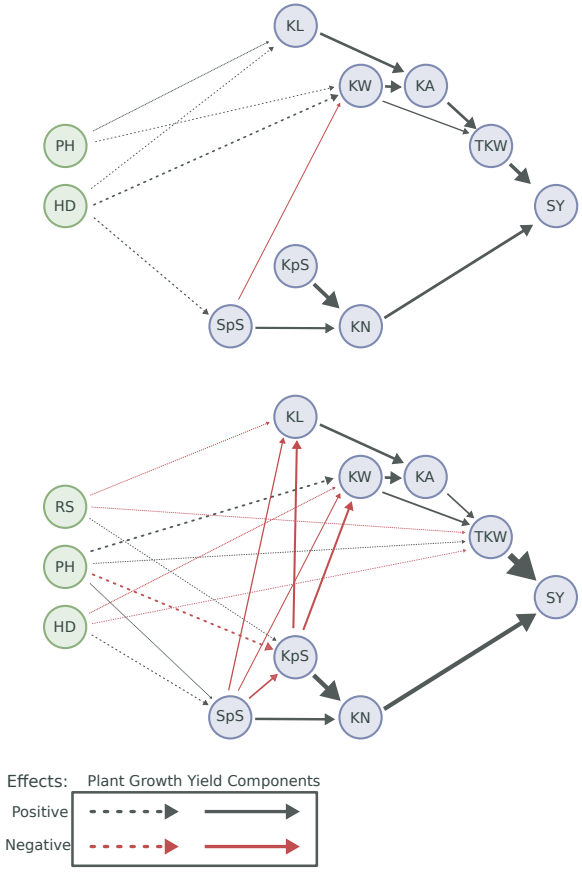

Figure 3.5: **Phenotype graph for Raleigh and Kinston 2019.** Inter-phenotype coefficients for full SEM model (including QTL effects which are not shown) fit in Raleigh and Kinston 2019. Thickness of paths demonstrate standardized effect size, while color indicates direction of effect. Leaf rust score (RS) is shown in Kinston but not Raleigh, where no pathogen affected plants. Negative correlations between yield components in Kinston after correcting for genetic effects suggests environment-specific source constraints.

any location and relative importance of other loci varied. All yield QTL together explained more than a third of additive genetic variation in SY at RAL19. Although *B1* explained a substantial portion of additive genetic variation in negatively correlated traits TKW and KN in this environment, it was associated with almost no additive genetic variation in SY. The *B1* awn suppressor explained a substantial proportion of additive genetic variation for SY at GA19 at KIN19, but very little for KN (Fig. 9). Interestingly, *WAPO-A1* was associated with 8.0% of additive genetic variation for SY at KIN19 and almost no additive genetic variation for SY at RAL19 and GA19. The differential importance of QTL in different environments suggests that the magnitudes of QTL effects on spike yield change from environment to

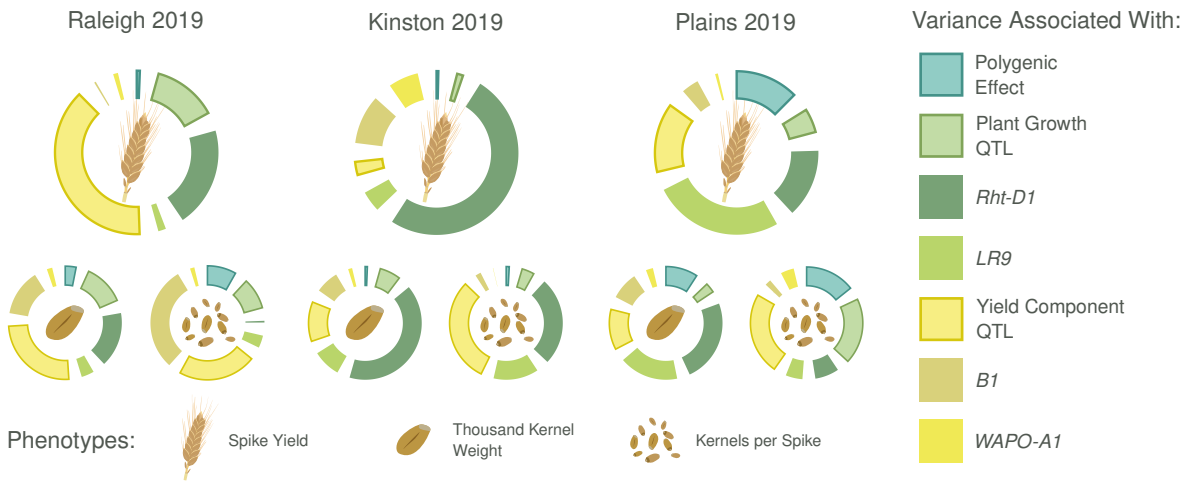

Figure 3.6: **Spike yield additive genetic variance components** Partitioning of additive genetic variance associated with different subsets of marker genotypes. The genetic architecture of spike yield changes dramatically from environment to environment, with a relationship to the genetic architecture of component traits TKW (thousand kernel weight) and KN (kernel number).

environment through both differential effects on the yield components themselves and altered relationships between yield components.

The small proportion of additive genetic variation associated with the polygenic background across environments validates the ability of the mapping model to identify the loci generating yield component variation within each environment. The polygenic effect was largest at GA19, explaining over 15% of variation, but less than 5% at RAL19 and KIN19. The combined contribution of all other identified plant growth and yield component QTL was small for GA19 and KIN19. In contrast, at RAL19 2.7% of the additive genetic variance was associated with these yield component QTL, and 15.7% with plant growth QTL.

3.4.5 Major plant height and heading date allele effects

Differences in the relative importance of individual loci and categories of loci were examined in-depth using path analysis in the full network model. *Rht-D1* explained a substantial portion of additive genetic variation in all environments but dominated total additive genetic variation in KIN19. To understand the mechanism by which the *Rht-D1* effect

differed between RAL19 and KIN19, predicted allele effects on spike yield were estimated via path analysis starting with the estimated QTL effects of *Rht-D1* in both environments (Figure 3.7). In both environments, the *Rht-D1a* allele has a slight positive effect on heading date, but strong negative effects on plant height and kernel width. *Rht-D1a* reduces growth by preventing the degradation of the DELLA protein it alters, rendering plants insensitive to gibberellic acid (Peng et al. 1999). This affects the growth of cells in developing kernels and the elongating stem tissue (Keyes et al. 1990). *Rht-D1a* was associated with an increase in heading date in both environments, contributing to slight increases in SpS and KN. Reduced plant height due to *Rht-D1a* at KIN19 also contributed to increased KpS and further increased KN. In RAL19, a small negative effect of plant height on KW resulted from the direct negative effect of *Rht-D1a* on this trait. In KIN19, a stronger association between plant height and KW was observed along with a larger effect of *Rht-D1* on plant height; decreased KW in this environment result from both the direct effect of *Rht-D1a* and the indirect effect mediated by plant height and KpS. While *Rht-D1a* was associated with decreased TKW in both environments, the effect in KIN19 was larger in part due to a stronger effect of plant height on KW. In addition, the decrease in thousand kernel weight had a larger negative effect on spike yield in KIN19 than in RAL19, so that *Rht-D1* explained a larger proportion of additive genetic variation in yield in the KIN19 environment. Thus, the specific *Rht-D1* by environment interactions were due to both differential effects of *Rht-D1* on plant height and KW in the two environments, as well as different relationships between phenotypes in the two separate environments.

3.4.6 Major per-se yield component allele effects

Allele coefficients and relationships between yield component phenotypes were used to investigate allele effects across the whole set of yield components in each environment (Figure 3.8). While the direction of effects on base yield component phenotypes remains constant, effect sizes often change between environments. Additional changes in the effect from environment to the environment is produced by altered relationship between phenotypes. *B1* has a much larger estimated effect in KIN19 than in RAL19, tracking with its much greater importance for determining spike yield in that environment. At the same time, *B1* generates major variation in both TKW (thousand kernel weight) and KN (kernel number) in RAL19. The presence of awns is associated with increases in kernel weight and decreases in KN in both environments due to the positive effects on kernel density and length. At

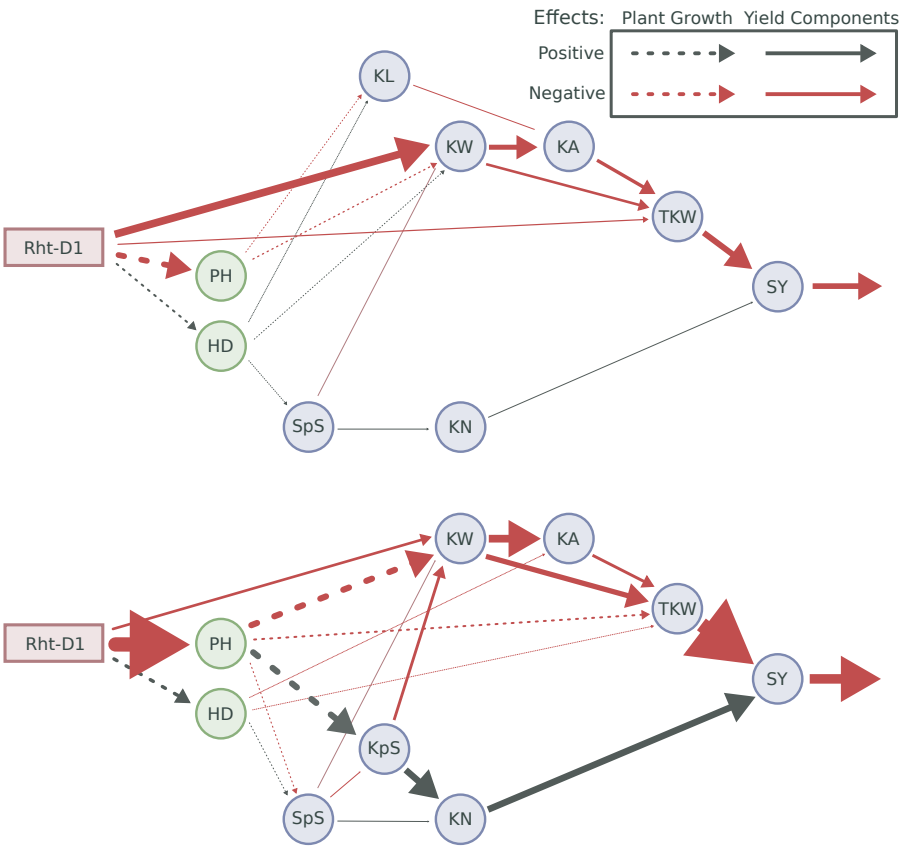

Figure 3.7: **Phenotype graph for *Rht-D1* allele effects.** Inter-phenotype coefficients for full SEM model (including QTL effects which are not shown) fit in Raleigh and Kinston 2019. Thickness of paths demonstrate standardized effect size, while color indicates direction of effect. Leaf rust score (RS) is shown in Kinston but not Raleigh, where no pathogen affected plants. Negative correlations between yield components in Kinston after correcting for genetic effects suggests environment-specific source constraints.

KIN19 the increase in kernel weight was larger than the decrease in KN, leading to a major positive effect on spike yield. However, at RAL19, the contrasting effects on component traits are similar and produce almost no spike yield effect. The presence of awns was also associated with a positive SY effect in GA19, with a smaller positive effect on TKW than in KIN19, but no negative effect on KN.

WAPO-A1, a recently identified gene involved in determining the number of spikelets per spike, was the largest per-se QTL for SpS in the LM population. If *WAPO-A1* can consistently increase yield by increasing KN without altering kernel size, MAS may be used to increase

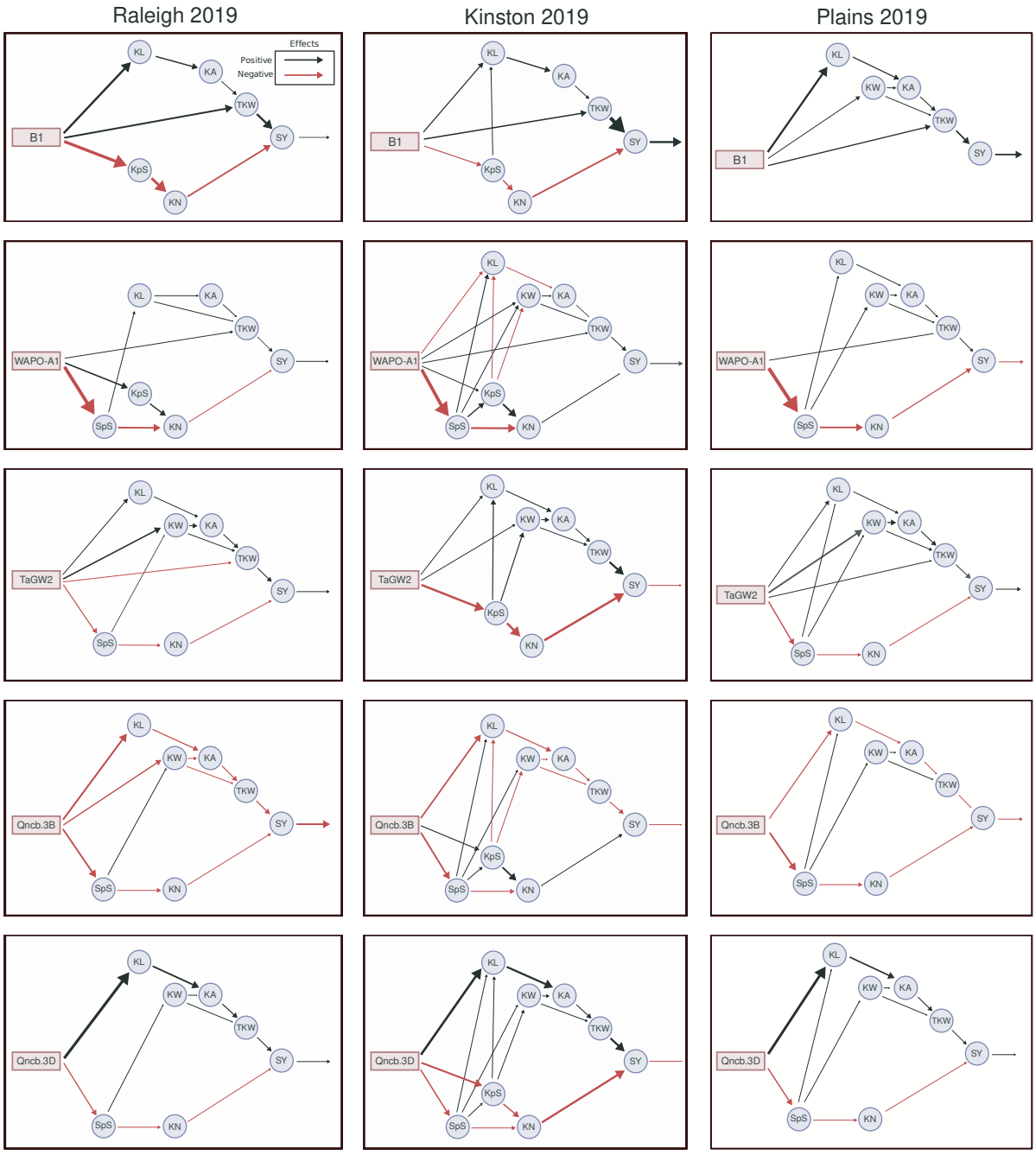

Figure 3.8: **Allele effects of major yield component QTL** QTL models fit in each environment (columns) were used to estimate significant ($p < 0.10$) allele effects for previously mapped QTL on multiple base scaled yield component phenotypes, as well as effects on TKW (thousand kernel weight) presumed to be mediated by their effects on test weight (kernel density). Effects of those allele effects on base phenotypes on additional phenotypes are estimated by path analysis using the estimated model, by the multiplication of the sum of effects on a phenotype by that phenotypes standardized effect on other phenotypes. Many QTL generate substantial effects on yield components and little effect on spike yield, due to contrasting effects on component phenotypes. QTL effects vary significantly by environment, but only in magnitude and not in effect direction.

yield by increasing the frequency of *WAPO-A1a* in breeding populations. However, the only environment where the increased SPS allele is associated with an increase in spike yield is GA19 (Figure 3.8). The *WAPO-A1* allele from LA95135 has a significant negative effect on SpS and is also associated with an increase in KpS and kernel density in RAL19 and KIN19, and an increase in KW in KIN19. An increase in kernel size associated with the decreased SpS allele also emerges because of the negative effect of SpS on KW in both environments. The effect at KIN19 on spike yield is greater, where decreasing SpS increases KpS, leading to a minor increase in both KN and a larger increase in kernel weight for the lower SpS allele. Increasing yield by increasing spikelet number through MAS for *WAPO-A1* will necessitate understanding the environmental conditions that lead to these altered relationships. The *TaGW-A2* gene is a negative grain size regulator. The LA95135 allele of a SNP polymorphism in the promoter region segregating in the population was associated with an increase in spike yield at RAL19 and GA19 but a slight decrease at KIN19 (Figure 3.8). In all environments the LA95135 allele is associated with increased TKW due to positive effects on KL and KW, and also with a decrease in KN. At RAL19 this decrease in KN resulted from a decrease in SPS, and at KIN19 with a decrease in KPS. In both cases, the decrease in KN is associated with an increase in TKW. The total negative effect of the increased KW allele on SY in KIN19 results from a relatively larger decreases in KN compared to the other two environments, where it is associated with an overall increase in SY.

The LA95135 allele of a QTL mapped to the centromeric region of chromosome 3B is associated with decreases in both kernel size and SpS at RAL19 and GA19. The LA95135 allele of *Qncb-3B* decreases KL and SpS in all environments, but is associated with increases in KpS at KIN19. This allele has a major negative effect on spike yield in RAL19 and GA19, as it decreased both TKW and KN. At KIN19, an increase in KpS contributed to a comparatively smaller negative effect on spike yield despite similar effects on KL and SpS. The QTL having effects in the same direction on both kernel size and number is further evidence that the pleiotropy observed for per-se yield component QTL in this study is not arising solely from source constraints.

Qncb-3D is the major per-se KL QTL in this population, increasing KL with smaller negative effects on KpS and SpS. In all environments, effects on spike yield are small and the QTL is associated with almost no additive genetic variation. Small positive or negative estimated effects of the allele result from contrasting positive effects on TKW and negative effects on KN.

3.5 Conclusions

Modeling the relationships between genotypes and phenotypes using SEM can help us understand the mechanisms behind observed marker-trait associations for yield, and their relative utility in marker-assisted selection. In all cases, the relative importance of a QTL in generation variation in spike yield relates to its predicted effect on spike yield as calculated by allele effects on base components and relationships between yield component phenotypes. Dissection of environment-specific spike yield effects demonstrates that variation in these effects emerges in part from altered allele effects on base yield components (most importantly, the presence or absence of an effect on a specific trait) as well as altered relationships between phenotypes in different environments. The different pleiotropic effects of major yield component QTL on base yield components suggests that these different QTL are affecting base yield components through different genetic pathways or physiological relationships. RNA-sequencing and careful tracking of plant physiology in NILs contrasting for these major QTL may yield insights into how differences in sequence variation alter base yield components in environmentally specific ways.

We found no evidence in this study for the presence of QTL that consistently alter spike yield without altering plant phenology. Identified QTL are collectively associated with nearly all of the additive genetic variation in spike yield within each environment. However, the relative effects of QTL in each environment on spike yield, and therefore their relative contribution to yield variation, change in magnitude from environment to environment. Genomic selection models trained in these populations are then not averaging together QTL effects that vary from one location to the next due to error variation in effect estimates, but a combination of error variation and real differential expression of QTL. As the direction of QTL effects largely remained the same within these environments, these models still produce useful results – but suggests that there is room for improvement through an understanding and modeling of effect variation. There is no simple relationship between any identified QTL and spike yield. To truly make use of major genes in understanding and predicting yield, the underlying environmental factors altering differential QTL expression must be identified and explicitly modeled. Structural equation modeling and network analyses similar to those utilized in this study may be one framework in which to do so.

3.6 Data availability

Genotype and phenotype data for the biparental population is available in Supplemental files on figshare. Code used to perform QTL mapping using SEM can be found on Github (https://github.com/noahddewitt/SEM_QTL).

3.7 Acknowledgments

The authors thank staff of the USDA-ARS Plant Science Unit for assistance in genotyping and field evaluation of populations. Special thanks to Kim Howell, Kou Vang, and Jared Smith, who extracted DNA, prepared GBS libraries, and assisted with field work. Additional thanks to James B. Holland, Amanda Peters-Haugrud, Katherine Running, Meng Su, and Dylan Larkin, who all provided valuable feedback on early drafts of the manuscript.

References

Addison, C. K., Mason, R. E., Brown-Guedira, G., Guedira, M., Hao, Y., Miller, R. G., Subramanian, N., Lozada, D. N., Acuna, A., Arguello, M. N., Johnson, J. W., Ibrahim, A. M., Sutton, R., and Harrison, S. A. (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.

Boden, S. A., Cavanagh, C., Cullis, B. R., Ramm, K., Greenwood, J., Jean Finnegan, E., Trevaskis, B., and Swain, S. M. (2015). Ppd-1 is a key regulator of inflorescence architecture and paired spikelet development in wheat. *Nature Plants*, 1.

Borner, A., Worland, A. J., Plaschke, J., Schumann, E., and Law, C. N. (1993). Pleiotropic Effects of Genes for Reduced Height (Rht) and Day-Length Insensitivity (Ppd) on Yield and its Components for Wheat Grown in Middle Europe. *Plant Breeding*, 111(3):204–216.

Borrás, L., Slafer, G. A., and Otegui, M. E. (2004). Seed dry weight response to source-sink manipulations in wheat, maize and soybean: A quantitative reappraisal. *Field Crops Research*, 86(2-3):131–146.

Brinton, J. and Uauy, C. (2019). A reductionist approach to dissecting grain weight and yield in wheat. *Journal of Integrative Plant Biology*, 61(3):337–358.

Broman, K. W., Wu, H., Sen, S., and Churchill, G. A. (2003). R/qtl: QTL mapping in experimental crosses. *Bioinformatics*, 19(7):889–890.

Cheng, X., Xin, M., Xu, R., Chen, Z., Cai, W., Chai, L., Xu, H., Jia, L., Feng, Z., Wang, Z., Peng, H., Yao, Y., Hu, Z., Guo, W., Ni, Z., and Sun, Q. (2020). A Single Amino Acid Substitution in STKc_GSK3 Kinase Conferring Semispherical Grains and Its Implications for the Origin of Triticumsphaerococcum. *The Plant cell*, 32(4):923–934.

Crespi, B. J. and Bookstein, F. L. (1989). A path-analytic model for the measurement of selection on morphology. *Evolution*, 43(1):18–28.

DeWitt, N., Guedira, M., Lauer, E., Murphy, J. P., Marshall, D., Mergoum, M., Johnson, J., Holland, J. B., and Brown-Guedira, G. (2021). Characterizing the oligogenic architecture of plant growth phenotypes informs genomic selection approaches in a common wheat population. *BMC Genomics*, 22(1).

DeWitt, N., Guedira, M., Lauer, E., Sarinelli, M., Tyagi, P., Fu, D., Hao, Q., Murphy, J. P., Marshall, D., Akhunova, A., Jordan, K., Akhunov, E., and Brown-Guedira, G. (2020). Sequence-based mapping identifies a candidate transcription repressor underlying awn suppression at the B1 locus in wheat. *New Phytologist*, 225(1):326–339.

Dhungana, P., Eskridge, K. M., Baenziger, P. S., Campbell, B. T., Gill, K. S., and Dweikat, I. (2007). Analysis of genotype-by-environment interaction in wheat using a structural equation model and chromosome substitution lines. *Crop Science*, 47(2):477–484.

Fischer, R. A. (2008). The importance of grain or kernel number in wheat: A reply to Sinclair and Jamieson. *Field Crops Research*, 105(1-2):15–21.

Fischer, R. A. and Quail, K. J. (1990). The effect of major dwarfing genes on yield potential in spring wheats. *Euphytica*, 46(1):51–56.

Gilmour, A. R. (2018). Echidna Mixed Model Software. In *Proceedings of the World Congress on Genetics Applied to Livestock Production*.

Grafius, J. E. (1964). A Geometry for Plant Breeding. *Crop Science*, 4(3):241–246.

Griffiths, S., Wingen, L., Pietragalla, J., Garcia, G., Hasan, A., Miralles, D., Calderini, D. F., Ankleshwaria, J. B., Waite, M. L., Simmonds, J., Snape, J., and Reynolds, M. (2015). Genetic Dissection of Grain Size and Grain Number Trade-Offs in CIMMYT Wheat Germplasm. *PLOS ONE*, 10(3):e0118847.

Kaplan, R. H. and Phillips, P. C. (2006). Ecological and developmental context of natural selection: maternal effects and thermally induced plasticity in the frog bombina orientalis. Technical Report 1.

Keyes, G., Sorrells, M. E., and Setter, T. L. (1990). Gibberellic Acid Regulates Cell Wall Extensibility in Wheat (Triticum aestivum L.). *Plant Physiology*, 92(1):242–245.

Knott, D. R. and Talukdar, B. (1971). Increasing Seed Weight in Wheat and Its Effect on Yield, Yield Components, and Quality. *Crop Science*, 11(2):280–283.

Kuzay, S., Xu, Y., Zhang, J., Katz, A., Pearce, S., Su, Z., Fraser, M., Anderson, J. A., Brown-Guedira, G., DeWitt, N., Peters Haugrud, A., Faris, J. D., Akhunov, E., Bai, G., and Dubcovsky, J. (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.

Mi, X., Eskridge, K., Wang, D., Baeziger, S. P., Campbell, T. B., Gill, K. S., and Dweikat, I. (2010). Bayesian mixture structural equation modelling in multiple-trait QTL mapping. *Genetics Research*, 92:239–250.

Momen, M., Ayatollahi Mehrgardi, A., Amiri Roudbar, M., Kranis, A., Mercuri Pinto, R., Valente, B. D., Morota, G., Rosa, G. J. M., and Gianola, D. (2018). Including Phenotypic Causal Networks in Genome-Wide Association Studies Using Mixed Effects Structural Equation Models. *Frontiers in Genetics*, 9(OCT):455.

Momen, M., Campbell, M. T., Walia, H., and Morota, G. (2019). Utilizing trait networks and structural equation models as tools to interpret multi-trait genome-wide association studies. *Plant Methods*, 15(1).

Nakagawa, S. and Schielzeth, H. (2013). A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, 4(2):133–142.

Peng, J., Richards, D. E., Hartley, N. M., Murphy, G. P., Devos, K. M., Flintham, J. E., Beales, J., Fish, L. J., Worland, A. J., Pelica, F., Sudhakar, D., Christou, P., Snape, J. W., Gale, M. D., and Harberd, N. P. (1999). 'Green revolution' genes encode mutant gibberellin response modulators. *Nature*, 400(6741):256–261.

Sadras, V. O. (2007). Evolutionary aspects of the trade-off between seed size and number in crops.

Simons, K. J., Fellers, J. P., Trick, H. N., Zhang, Z., Tai, Y. S., Gill, B. S., and Faris, J. D. (2006). Molecular characterization of the major wheat domestication gene Q. *Genetics*, 172(1):547–555.

Su, Z., Hao, C., Wang, L., Dong, Y., and Zhang, X. (2011). Identification and development of a functional marker of TaGW2 associated with grain weight in bread wheat (Triticum aestivum L.). *Theoretical and Applied Genetics*, 122(1):211–223.

Vargas, M., Reynolds, M. P., Dhungana, P., and Eskridge, K. M. (2007). Structural equation modeling for studying genotype environment interactions of physiological traits affecting yield in wheat. *Journal of Agricultural Science*, 145:151–161.

Wang, W., Simmonds, J., Pan, Q., Davidson, D., He, F., Battal, A., Akhunova, A., Trick, H. N., Uauy, C., and Akhunov, E. (2018). Gene editing and mutagenesis reveal inter-cultivar differences and additivity in the contribution of TaGW2 homoeologues to grain size and weight in wheat. *Theoretical and Applied Genetics*, 131(11):2463–2475.

Wright, S. (1920). The relative importance of heredity and environment in determining the piebald pattern of guinea-pigs. *Proceedings of the National Academy of Sciences*, 6(6):320–332.

Youssefian, S., Kirby, E. J., and Gale, M. D. (1992). Pleiotropic effects of the GA-insensitive Rht dwarfing genes in wheat. 2. Effects on leaf, stem, ear and floret growth. *Field Crops Research*, 28(3):191–210.

Ziyatdinov, A., Vázquez-Santiago, M., Brunel, H., Martinez-Perez, A., Aschard, H., and Soria, J. M. (2018). lme4qtl: linear mixed models with flexible covariance structure for genetic studies of related individuals. *BMC Bioinformatics*.

Chapter 4

Sequence based mapping identifies a candidate transcription repressor underlying awn suppression at the *B1* locus in wheat

Noah DeWitt[1, 3] Mohammed Guedira[1] Edwin Lauer[1] Martin Sarinelli[1] Priyanka Tyagi[1] Daolin Fu[2] QunQun Hao[2] J. Paul Murphy[1] David Marshall[2] Alina Akhunova[4] Katherine Jordan[4] Eduard Akhunov[4] Gina Brown-Guedira[1, 3]

1 Department of Crop and Soil Sciences, North Carolina State University, Raleigh, NC 27695
2 Department of Plant Sciences, University of Idaho, Moscow, ID, USA 83844
3 USDA-ARS SEA, Plant Science Research, Raleigh, NC 27695
4 Department of Plant Pathology, Kansas State University, Manhttan, KS, USA 66506

Published in *New Phytologist*, August 2019 (DOI: doi.org/10.1111/nph.16152)

4.1 Abstract

Awns are stiff, hair-like structures which grow from the lemmas of wheat (*Triticum aestivum* L.) and other grasses that contribute to photosynthesis and play a role in seed dispersal. Variation in awn length in domesticated wheat is primarily controlled by three major genes, most commonly the dominant awn suppressor *Tipped1* (*B1*). This study identifies a transcription repressor responsible for awn inhibition at the *B1* locus. Association mapping was combined with analysis in bi-parental populations to delimit *B1* to a distal region of 5AL co-localized with QTL for number of spikelets per spike, kernel weight, kernel length, and test weight. Fine-mapping located *B1* to a region containing only two predicted genes, including C2H2 zinc finger transcriptional repressor *TraesCS5A02G542800* up-regulated in developing spikes of awnless individuals. Deletions encompassing this candidate gene were present in awned mutants of an awnless wheat. Sequence polymorphisms in the *B1* coding region were not observed in diverse wheat germplasm while a nearby polymorphism was highly predictive of awn suppression. Up-regulation of the *B1* repressor is the major determinant of awn suppression in global wheat germplasm. It is associated with increased number of spikelets per spike and decreased kernel size.

4.2 Introduction

Awns are stiff, hair-like structures common in grass inflorescences. In Poaceae, awns emerge from the lemma of young spikelets at an early developmental stage. Awns are important for seed dispersal in wild relatives of wheat with a brittle rachis; spikelets attach to passing animals via barbs lining the awns. In addition, awns balance spikelets as they fall, and their expansion and contraction in response to changes in humidity drives them into the soil (Elbaum et al. 2007). Awns may also deter herbivores from ingesting heads (Grundbacher 1963), making awn suppression important for developing forage cultivars (Cash et al. 2009). The absence of awns in rice is considered a key domestication trait facilitating grain harvest and storage (Toriba and Hirano 2014), but the history of the evolution and spread of awn suppression in domesticated wheat is not well understood.

Major variation for awn length in domesticated wheats emerge from different combinations of three dominant genes: *B1* (*Tipped 1*), *B2* (*Tipped 2*) and *Hd* (*Hooded*) (McIntosh et

al., 2013). On its own, the *B1* awn suppressor produces an apically awnletted phenotype, characterized by short awns at the end of the spike and absent elsewhere (Watkins and Ellerton 1940). The *B2* allele reduces awn length most dramatically towards the top and bottom of the wheat spike, while the *Hd* allele reduces awn length consistently and can produce curved, "hooked" awns (Watkins and Ellerton 1940). Combination of these genes produces a nearly awnless or completely awnless phenotype (Yoshioka et al. 2017). None of the genes controlling awn length in wheat have been cloned, but mapping studies place *Hd* and *B2* on the short arm of chromosome 4A and the long arm of chromosome 6B, respectively (Sourdille et al. 2002; Yoshioka et al. 2017). The *B1* locus has a long history as a physical marker, and is located distal to the major genes controlling vernalization requirement (*VRN-A1*) and the spelt head type (*Q*) on 5AL (Kato et al. 1998). Recent fine mapping has narrowed the *B1* region to a 7.5 cM interval on the distal end of 5AL closely linked to marker *BW8226_227* (Mackay et al. 2014). In *Aegilops tauschii*, the donor of the D-genome in hexaploid wheat, an additional dominant awn suppressor *Anathera* (*Antr*) was located distally on 5DS (Nishijima et al. 2018). Deletion of the short arm of chromosome 3B in Chinese Spring also produces an awned phenotype, suggesting that further uncharacterized genes are involved in awn development (Ma et al. 2012).

Awn suppression in barley is controlled by two major genes: the homeobox gene Knox3 (Müller et al., 1995) underlying the *Hooded* (*K*) locus that replaces awns with sterile flowers, and a *SHORT INTERNODES* transcription factor (Yuo et al., 2012) underlying the short awn 2 (lks2) locus (Roig et al., 2004, Takahashi 1955). In rice awn suppression is an important domestication trait, and is derived from mutations in genes involved in awn development: a helix-loop-helix protein I (*An-1*), an auxin response factor I, a YABBY transcription factor *DROOPING LEAF (DL)*, *LONG AND BARBED AWN (LABA1)* involved in cytokinin biosynthesis, and an EPIDERMAL *PATTERNING FACTOR-LIKE 1 (EPFL1)* gene *REGULATOR OF AWN ELONGATION 2 (RAE2)* or *GRAIN NUMBER, GRAIN LENGTH AND AWN DEVELOPMENT1 (GAD1)* (Bessho-Uehara et al., 2016, Hua et al., 2015, Jin et al., 2016, Luo et al., 2013, Toriba & Horina 2013). However, awn suppression in wheat does not appear to be related to these genes (Yoshioka et al., 2017).

Variation in awn length across modern wheat cultivars and landraces suggests awns are variably adaptive in different environments. Previous studies support a role for awns in supplying photosynthate to developing grain of wheat and barley (Grundbacher 1963; Kjack and Witters 1974; Motzo and Giunta 2002; Li et al. 2006; Tambussi et al. 2007; Amjad Ali

et al. 2010; Maydup et al. 2014), and the location of awns on the wheat head facilitates movement of carbohydrates into kernels and positions them favorably for photosynthesis (Amjad Ali et al. 2010; Evans et al. 1972; Li et al. 2006). Awns may continue contributing to photosynthesis if leaves senesce early or are damaged by disease or drought (Tambussi et al. 2007), and the absence of awns can halve the rate of net ear photosynthesis (Evans et al. 1972). Awn tissue, potentially due to its silica coating, tolerates water deficit better than other important photosynthetic tissues such as the flag leaf (Tambussi et al. 2005; Peleg et al. 2010). Besides increased area for photosynthesis, in warmer climates awns can play a role in cooling the wheat spike during grain fill (Motzo and Giunta 2002). Awned wheats have been demonstrated to perform better in hotter climates or under water stress. The presence of awns is associated with smaller numbers of larger kernels (Rebetzke et al. 2016). Similarly in rice, wild type awned plants have larger kernels with a reduced number per panicle when compared to awnless *GAD1* mutants (Jin et al. 2016).

Awnless or awnletted wheats are widely cultivated and are the dominant morphology in many parts of the world. Potential explanations for the prevalence of awn inhibited types is their association with a reduced incidence of pre-harvest sprouting (Cao et al. 2016; King and Richards 1984), use of wheat as forage, and historical ease of harvest. In warm growing regions like the southeastern U.S. where awnless varieties are historically dominant, the proportion of awned varieties has increased over the past two decades. Given the importance of awn status in selection of cultivars for local adaptation and end use, and the influence of awns on spike and kernel morphology, a better understanding of the genetic basis of awn suppression should also improve our understanding of these processes. In this study, we investigated the relationship of awn suppression with kernel quality and spike morphology in association and bi-parental mapping populations. Fine-mapping was combined with analysis of mutant lines and gene expression to identify a candidate gene responsible for awn suppression at the *B1* locus. Our companion paper Huang et al. (2020) presents parallel identification of *B1*, and through transcriptome analyses of awnletted *B1* over-expressing plants they propose possible pathways through which *B1* may act for awn inhibition.

4.3 Materials and Methods

4.3.1 Genome Wide Association Analysis

An association mapping panel of 640 elite soft winter wheat breeding lines was grown in two replications as 1 m rows spaced 30 cm apart at Raleigh, North Carolina during the 2016-2017 growing season. At heading, presence or absence of a fully awned phenotype was noted. These lines were entries in collaborative yield testing nurseries in the southeast soft wheat growing region of the United States, the Gulf Atlantic Wheat Nursery (GAWN) and the SunWheat Nursery, over a period of nine growing seasons. Association analyses were performed for grain yield and test weight using historical data available for the 640 entries. The GAWN and SunWheat yield trials were evaluated at one location in up to seven states from 2008 to 2016: Arkansas (Stuttgart or Marianna), Florida (Citra or Quincy), Georgia (Plains), Louisiana (Winnsboro), North Carolina (Kinston), Texas (Farmersville) and Virginia (Warsaw). Experimental designs at each environment were randomized complete block designs with one to three replications. Plot size was typical of yield trial plots for wheat in the region at a minimum of 1.3 m wide and 3.1 m long. The data set was balanced for individual years, where the same set of genotypes was evaluated across different locations, and unbalanced between different years. Mixed model analyses of grain yield and test weight data were performed as described in Sarinelli et al. (2019), except that the dataset in the present study was expanded to include the SunWheat nursery. Best linear unbiased estimates (BLUEs) of each genotype were calculated as the estimated genotypic effect plus overall mean and used as the response variable for association mapping.

Genotyping by sequencing (GBS; Elshire et al. (2011)) was performed according to Poland et al. (2012), with ninety-six individual samples barcoded, pooled into a single library, and sequenced on an Illumina HiSeq 2500. Tassel5GBSv2 pipeline version 5.2.35 (Glaubitz et al. 2014) was used to align raw reads to the International Wheat Genome Sequencing Consortium (IWGSC) RefSeqv1.0 assembly (https://wheat-urgi.versailles.inra.fr/Seq-Repository/Assemblies) using Burrows-Wheeler aligner (BWA) version 0.7.12 and call SNPs Li et al. (2009). SNPs were filtered to retain samples with ≤ 20 percent missing data, ≥ 5 percent minor allele frequency and ≤ 10 percent of heterozygous calls per marker. Missing SNPs were imputed using Beagle (Browning & Browning, 2016).

GWAS for awn status, grain yield and test weight were conducted in R version 3.3.1

(R Core Team 2016) using the GAPIT package (Lipka et al. 2012). Population structure and relatedness between individuals were accounted for using the first three principal components of the genomic relationship matrix, determined using the GBS markers and the prcomp function in R version 3.3.1 (R Core Team 2016). Markers were declared significant based on the Bonferroni corrected p-value at alpha = 0.01.

4.3.2 Bi-Parental Mapping Populations

A population of 341 F5-derived RILs was developed from the cross of awned cultivar LA95135 with awnless SS-MVP57 (LM population). LA95135 possesses the *Rht-D1b* (dwarfing) allele, and the *B1* allele for awns. SS-MPV57 has the *Rht-D1a* allele and the *Ppd-D1a* allele conferring photoperiod insensitivity, as well as the *B1* allele for awn suppression. Awned (NIL37-3) and awnless (NIL37-14) sister lines were derived from a F5 plant of RIL37 heterozygous at *B1* and homozygous for *Rht-D1b* and *Ppd-D1b*.

During the winter of 2016-2017, the LM population was grown in the greenhouse to evaluate spike morphology and kernel weight. Imbibed kernels from each RIL were placed in a cold chamber kept at 4°C for 8 weeks and were transplanted into plastic cones (Volume 0.7L, 6.5 cm in diameter and 25 cm depth) containing soil mix. Plants were grown in a completely randomized design with four replications in a greenhouse set to a 16-hour photoperiod and a 20°C /15°C (day/night) temperature. The primary tillers of each plant were used for evaluation of spike length, spikelet number per spike, kernel weight and presence or absence of awns.

The LM RIL population was evaluated in the field at Raleigh, NC and Kinston, NC during the 2017-2018 season. The 341 RILs were grown in an incomplete block design with two replications at each location. The population was divided into five blocks, each consisting of 72 entries and the two parents of the population. Sister lines NIL37-3 and NIL37-14 were grown at Raleigh, NC in an experiment with two blocks of 25 replications each.

Plots consisted of 1 m rows spaced 30 cm apart. Six representative spikes from each row were harvested and number of spikelets per spike recorded. Kernels were weighed and counted to determine kernel weight. Rows were hand harvested and threshed using a Vogel thresher. A 15 mL sample of grain was cleaned and morphometric parameters (grain length, width, area, and weight) were obtained using a MARVIN grain analyzer (GAT

Sensorik GMBH, Neubrandenburg, Germany). An estimate of test weight was obtained by measuring the weight of grain in the 15 mL sample. BLUEs were calculated for individual RILs using the R package lme4, treating genotype as a fixed effect and all other terms as random (Bates et al., 2015).

The RIL population was genotyped and SNP identified using the GBS protocol as described above, except that missing data were not imputed. A linkage map was constructed with these data, KASP markers for major effect loci *Rht-D1*and Ppd-D1, and the presence or absence of awns as a physical marker using the R packages R/qtl and ASMap (Broman et al. 2003; Taylor and Butler 2017). QTL analysis was performed using composite interval mapping, with significance thresholds for an alpha = 0.05 determined using 1000 permutations.

Three F2 populations (referred to as GxM, GxS, and LxS) were developed from crosses between awned breeding lines from the association mapping panel, GA06493-13LE6 (G) and LA09264C-P2 (L), with the awnletted cultivars SS-MPV57 (M) and SS8641 (S). After heading, a total of 950 plants were evaluated for awn suppression. Fully awned plants were placed in one category while awnletted or awnless plants were placed in a separate category.

Genomic DNA of the RIL and F2 populations was evaluated with KASP markers developed from sequences flanking GBS SNPs in the *B1* region (Tables S1 and S2) in the LM mapping population. In addition, polymorphisms were identified from exome capture data obtained using a previously described assay (Krasileva et al. 2017) for parental lines of mapping populations used in the USDA/IWYP-WheatCAP project that included LA95135 and SS-MPV57 (https://www.triticeaecap.org/qtl-cloning- projects). Linkage maps of the distal region of the long arm of chromosome 5A containing *B1* and selected KASP markers for each F2 and the RIL population were developed using R/qtl (Broman et al. 2003). Scaffold assemblies of awnless winter wheats with the *B1* suppressor (Cadenza, Paragon, Robigus, and Claire) and the awned tetraploid wheat Kronos were downloaded from the Earlham Institute website (https://opendata.earlham.ac.uk/opendata/data/). Scaffolds containing portions of the *B1* region in the awnless assemblies and Kronos were identified using BLAST and aligned to the IWSGC reference assembly RefSeqv1.0 of Chinese Spring using LAST (Kielbasa et al., 2011). SNPs between the awned and awnless wheats for KASP marker design were identified with Samtools (Li et al. 2009).

4.3.3 Analysis of Mutant Lines

Kernels of the awnless cultivar Brundage were treated using fast neutron irradiation with a center total dose of 7 Gy air at the McClellan Nuclear Radiation Center (McClellan, CA, USA). M1 kernels were planted at the Parker Farm in Moscow, Idaho. The main tiller was harvested from each M1 plant and kernels were planted as 1 m rows in the 2017-2018 crop season. Rows were noted as being either awnless or segregating for presence of awns. Markers in the *B1* region were evaluated on DNA isolated from one awnless plant and up to four awned plants from segregating rows. DNA samples of Chinese Spring and the Chinese Spring deletion line 5AL-6 (TA4535-6) missing the terminal 32% of 5AL were included as controls to evaluate genome specificity of markers. Progenies from the families with the smallest deletion around *B1* were grown in the greenhouse in Raleigh in 2018 to confirm the awned phenotype.

4.3.4 Identification of Haplotypes in Diverse Wheat Germplasm

Genomic DNA from a panel of 455 winter and 1984 spring wheat accessions from the core collection of the USDA-ARS National Small Grains Collection (NSGC) representing global diversity was evaluated using KASP markers developed from a 127 kb region flanking the *B1* locus. Data on the presence or absence of awns for winter wheat accessions was gathered from the U.S. National Plant Germplasm System descriptor data (https://npgsweb.ars-grin.gov). The spring wheat accessions were grown as single 1 m rows at Raleigh, NC. At heading time, single tillers were selected for each accession, the presence or absence of awns was noted and genomic DNA was isolated from the flag leaf.

4.3.5 Sequencing the *B1* region

The two parents of the RIL population and eight individuals representing diverse haplotypes for markers flanking *B1* were selected for Sanger sequencing (Table S3). Amplification of the *TraesCS5A02G542800* coding sequence and 7 kb of surrounding non-repetitive sequence was performed with a nested PCR design to obtain specificity using NEB Longamp polymerase (New England Biolabs, Ipswich, MA). The region was divided into 2.5 kb and 4.5 kb regions, and forward and reverse primers were developed for each region. Forward

sequencing primers were designed every 500-800 bp for Sanger sequencing of the region. CodonCode Aligner software (CodonCode Corporation, www.codoncode.com) was used to check base calls and assemble sequence reads into contigs.

4.3.6 Gene Expression

To evaluate expression of candidate genes, tissue from immature inflorescences was collected from primary and secondary tillers of plants of LA95135 and SS-MPV57, as well as the awned cultivar AGS 2000 (*b1*) and awnless cultivar Massey (*B1*). Individual samples from spikelets at similar stages were grouped for analysis to assess developmental variation in gene expression. RNA was isolated from plant tissue using the Zymopure RNA Extraction kit (Zymo Research, Irvine, CA), and reverse transcribed with the ThermoFischer Reverse Transcription kit. The predicted exons of candidate genes *TraesCS5A02G542700* and *TraesCS5A02G542800* were aligned to similar sequences on chromosomes 4B and 4D to design genome-specific primers for qPCR with an amplicon size of 100-150 bp and a Tm of 60-62^{o}C (Table S4). Quantitative PCR reactions were performed using a CFX384 real time PCR machine (Bio-Rad Laboratories, Hercules, CA) with Sybr green qPCR Master Mix (Applied Biosystems, Foster City, CA). Reactions included primers for the candidate gene along with a β *-ACTIN* control at an annealing temperature of 61o C (Table S4). There was a minimum of six awned and awnless samples at each developmental stage and three technical replications were performed per sample. Cq values were calculated for each replication using the Biorad CFX Maestro software and normalized to expression relative to the endogenous control β *-ACTIN* with $2^{(ACTINCT-TARGETCT)}$.

As an additional evaluation of gene expression over the course of apical development, the sequences of genes in the *B1* region were submitted to the WheatExp wheat expression database (https://wheat.pw.usda.gov/WheatExp/). β *-ACTIN* primer sequences were also submitted to verify that its expression is consistent during apical development.

4.4 Results

4.4.1 *B1* awn suppression is associated with test weight, spikelets per spike and kernel weight

Of the 640 lines evaluated in the eastern soft winter wheat panel, 58% had awns. Association mapping utilizing 14,567 GBS markers identified 30 markers significantly associated (adjusted P-value < 0.01) with the presence or absence of awns (Table S5). The significant markers were located distally on the long arm of chromosome 5A, consistent with the location of the *B1* awn suppressor (Fig. 1). Alignment of markers to IWGSC RefSeqv1.0 of Chinese Spring wheat placed the *B1* locus in a 25 Mb region between 681,455,268 bp and 706,705,101 bp. A SNP located at 698,528,417 bp on chromosome 5A (5A28417) was highly significant (P-value = 7. x 10-57) and co-segregated with awn status in this set of lines. Association analysis of historical data for test weight and grain yield of the GAWN and SunWheat regional testing nurseries identified a significant association of test weight with SNP *5A28417* (P-value = 4.1 x 10-7; Fig. 1a). Test weight is a measure of the weight of a standard volume of grain and is a general indicator of grain quality parameters such as kernel size and density. No significant markers for grain yield were identified.

Analysis of the LM RIL population developed from the cross of fully awned cultivar LA95135 with awnless SS-MPV57 indicated that awn inhibition was controlled by a single locus on chromosome 5A, co-segregating with SNP 5A28417. Observed awn phenotypes in RILs having residual heterozygosity at *B1* illustrated that the awn suppression was mostly dominant with short awns present on lemmas of apical spikelets of heterozygous plants (Fig 2a).

Multiple QTL for spikelets per spike and kernel morphometric traits were identified in the LM RIL population, including highly significant QTL associated with awn suppression at the *B1* locus and with markers predictive of major effect photoperiod locus *Ppd-D1* and plant height gene *Rht-D1*(Fig. 3, Table S6). The *Rht-D1b* semi-dwarfing allele was associated with the largest effects, contributing to decreases in kernel weight, kernel width and kernel area (Table S6). The *Ppd-D1a* allele for photoperiod insensitivity was also associated with decreased kernel weight in the greenhouse and a 2% and 6.1% decrease in number of spikelets per spike in the field and greenhouse, respectively. The presence or absence of

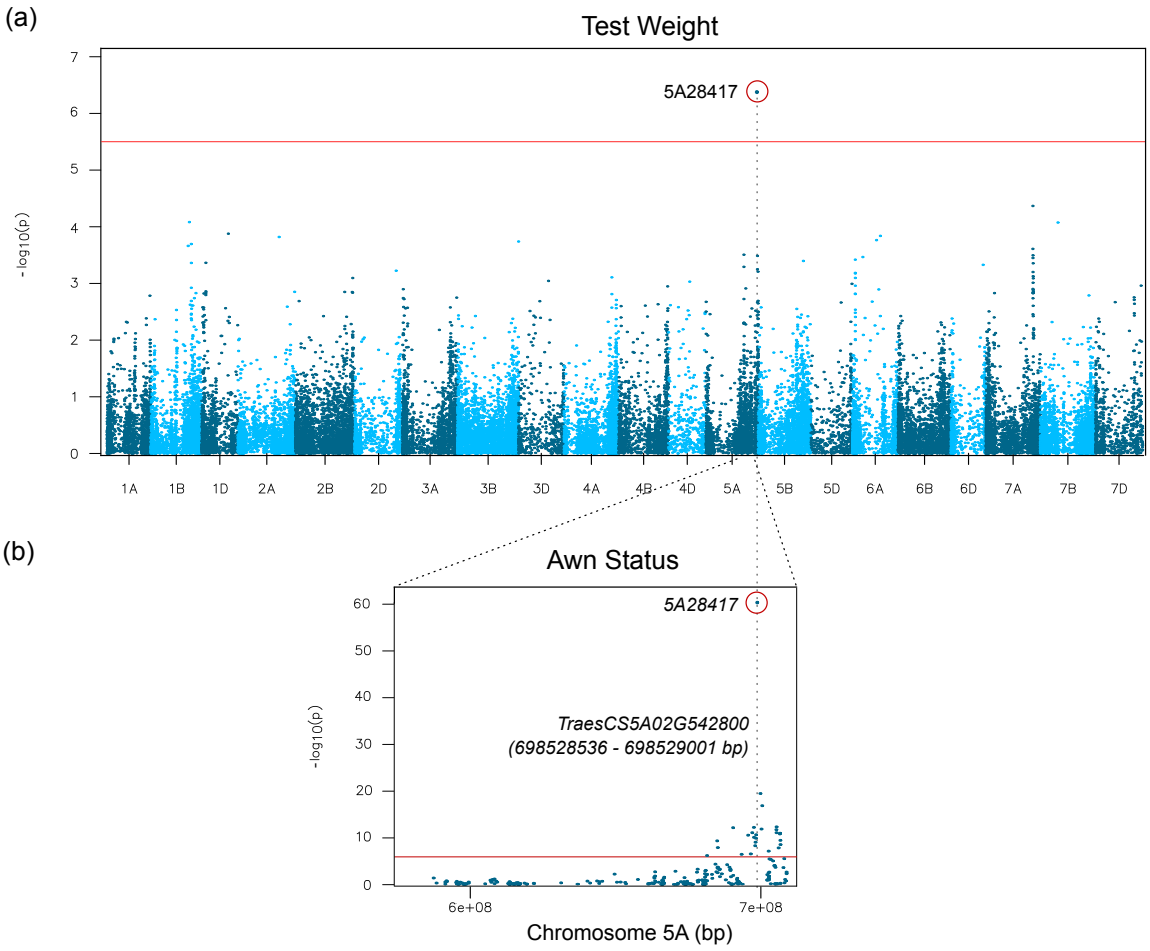

Figure 4.1: **Association analysis identifies the candidate awn inhibition gene on wheat chromosome 5A.** Genome-wide association analysis for test weight(a) and awn status (b). The most significant marker awn status (*5A28417*) also is the most predictive marker for test weight and is located 219 bp upstream of candidate zinc finger *TraesCS5A02G542800*.

awns, included in the genetic map as a physical marker, was significantly associated with QTL for number of spikelets per spike (LOD = 8.4), kernel weight (LOD = 6.5), kernel length (LOD = 9.5), and test weight (LOD = 5.6) (Fig. 3). In the greenhouse experiment, the presence of awns increased thousand kernel weight by 1.47 g (5.1%), and decreased spikelets per spike by 0.48 spikelets. In the field experiment, presence of awns increased kernel weight by 0.88 mg (3.2%), and decreased spikelets per spike by 0.37 spikelets (Fig. 2b). In addition, QTL at *B1* were significantly associated with estimated test weight and kernel length in 2018 field data (Fig. 2b). Highly significant ($P < .001$) differences in mean spikelets per spike, kernel weight, estimated test weight, and kernel length were observed when awned

and awnless F5-derived sister lines from heterozygous inbred line RIL37 were evaluated in the 2018 field experiment (Fig. 2c, 2d). Most notably, awned line NIL37-3 had 0.85 fewer spikelets per spike, a 6.3% increase in kernel weight, and an 11.6% greater estimated test weight compared with awnless sister line NIL37-14.

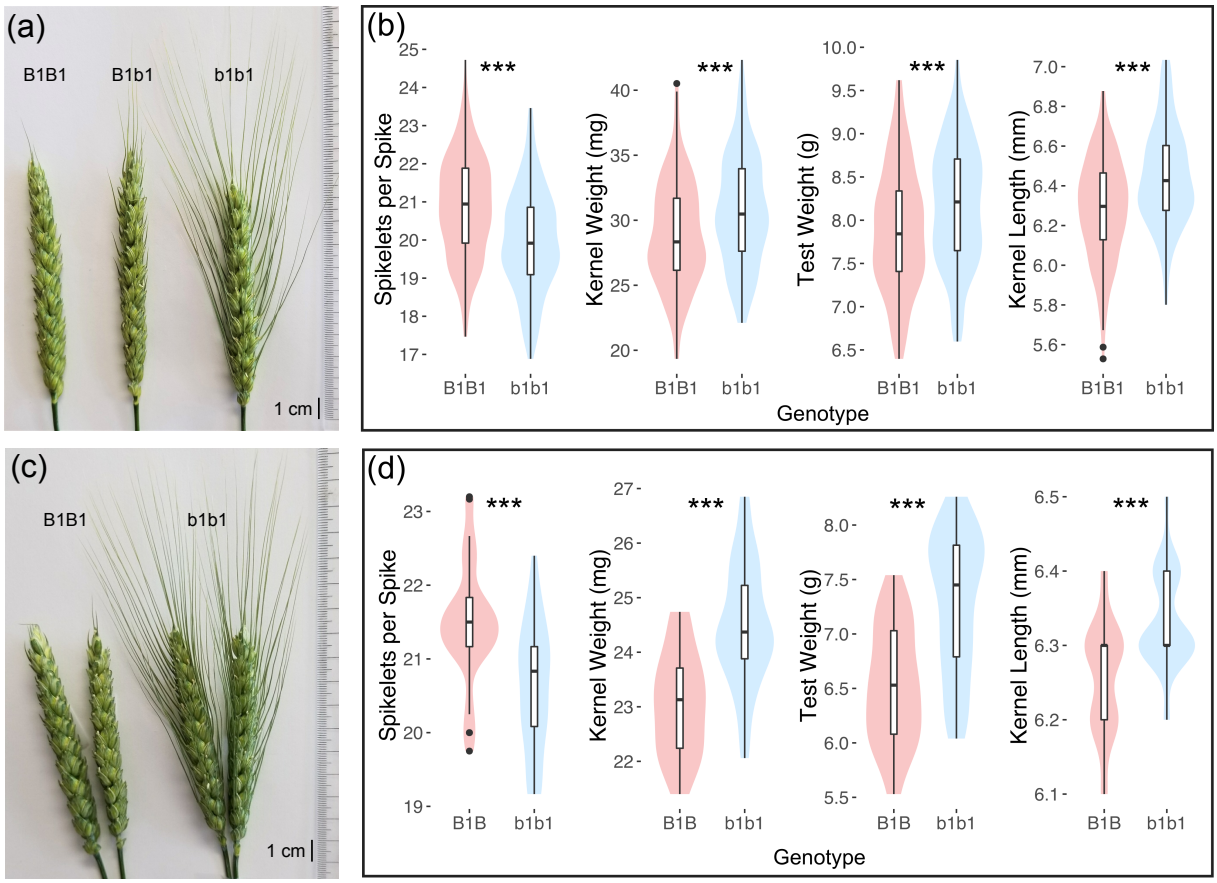

Figure 4.2: **Differences observed between awned and awnless recombinant inbred lines (RIL) and near isogenic lines (NIL) of wheat.** (a) Spikes from a segregating field row from the LA951359SS-MPV57 (LxM) RIL population show the effect of the mostly dominant *Tipped1 (B1)* allele. (b) Comparison of BLUEs calculated for spikelets per spike, kernel weight, test weight and kernel length indicate differences between awned vs awnless lines in the LxM RIL population. Significant difference are indicated (***, $P < 0.001$). (c) Spikes for HIF-derived NIL families 37-14 (left) and 37-3 (right), differing at the *B1* locus. (d) The *B1* NILs also differ significantly (***, $P < 0.001$) for all four traits in a highly replicated field experiment in Raleigh, NC in 2018.

4.4.2 Fine Mapping Identifies *B1* Candidate Genes

Significant SNP *5A28417* identified in the *B1* QTL region was located 219 bp upstream of predicted gene model *TraesCS5A02G542800* located on chromosome 5A from 698,528,636 bp to 698,529,001 bp in Chinese Spring RefSeqv1.0 (IWGSC 2018). The LM RIL population and three F2 populations developed from crosses between selected awned and awnless individuals from the association panel were genotyped for 8 KASP markers targeting an 8.52 Mb region flanking *TraesCS5A02G542800* (Fig. 4 Table S2). In each population, an awnless phenotype was conferred by a single dominant allele co-segregating with KASP marker 5A28417. Exome capture of the LA95135 and SS-MPV57 parents of the LM RIL population did not reveal polymorphisms in the *TraesCS5A02G542800* coding sequence. Polymorphisms identified in predicted genes proximal (*TraesCS5A02G542600* and *TraesCS5A02G5426700*) and distal (*TraesCS5A02G542900*) were targeted for marker development. KASP marker *5A15019* targeted an A/G variant in intron five of predicted gene *TraesCS5A02G542700* and BW8226_227 targeted a T/G polymorphism in exon two of *TraesCS5A02G542600*. Marker 5A30334 targeted a C/T variant in exon one of predicted gene *TraesCS5A02G542900*.

Recombination within the bi-parental populations narrowed the genomic region to a 127 kb region containing two predicted genes (Fig. 4). Recombination events observed between the awn phenotype and marker 5A30334 in the LM RIL population located the SNP in *TraesCS5A02G542900* 0.2 cM distal to *B1* (Fig. 4a). However, this did not exclude potential causative variation in the *TraesCS5A02G542900* promoter region. No recombination was observed between *B1* and 5A28417, *5A15019* and *BW8226_227* in the LM population. Of the 950 individuals evaluated from the three F2 populations, individual GM#101 from the cross between GA06493-13LE6 and SS-MPV57 was determined to be homozygous for BW8226_227 and heterozygous for *5A28417* and 5A15019. The awn phenotype segregated in a progeny test of 16 F3 plants derived from GM#101, placing the SNP in *TraesCS5A02G542600* 0.2 cM proximal to *B1*. Mackay et al. (2014) also identified *BW8226_22* as associated with but proximal to *B1*, both in a MAGIC population and a diversity panel. These results narrowed candidate genes underlying *B1* awn suppression to predicted genes *TraesCS5A02G542700*, *TraesCS5A02G542800*, and the promoter region of *TraesCS5A02G542900*.

Analysis of awned M2 plants of the awnless variety Brundage identified deletions on the distal part of 5AL encompassing the candidate genes (Table 1). Genomic DNA of awned and awnless M2 plants along with Chinese Spring and deletion line 5AL-6 (TA4535-6) missing

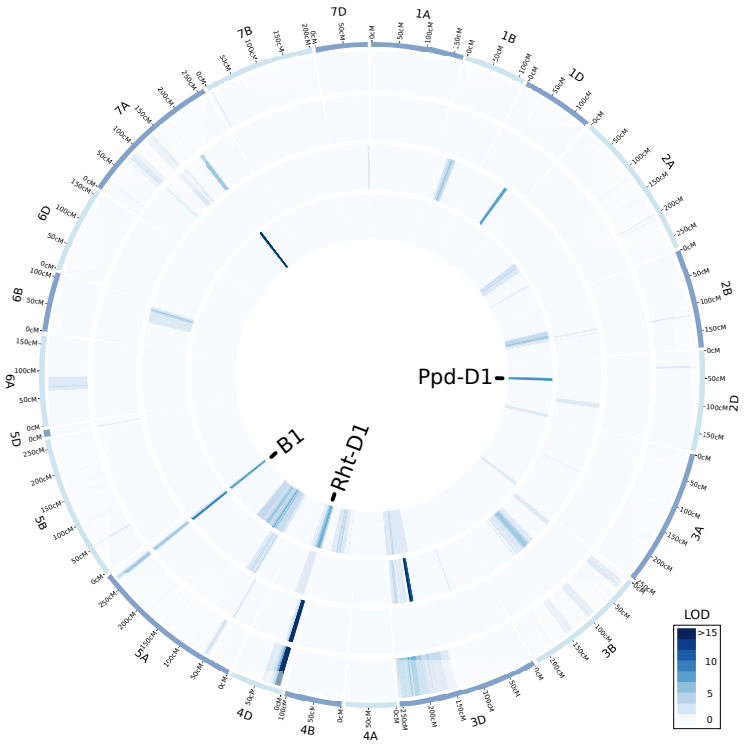

Figure 4.3: **QTL mapping in a wheat population segregating for *Tipped1 (B1)* awn suppression.** Heatmap of LOD scores for composite interval mapping in the LxM population of traits significantly impacted by presence or absence of awns. Traits from the outer to inner circle are thousand kernel weight, test weight, kernel length and spikelets per spike. Location of the *Rht-D1*, *Ppd-D1*, and *B1* loci are noted in the center. Data on thousand kernel weight, test weight and kernel length were collected from the field in Raleigh, NC in 2018. Data on spikelets per spike were collected from the field in Raleigh and Kinston, NC in 2018.

the terminal 32% of 5AL was used to amplify 11 KASP markers targeting 5AL from 696 Mb to 706 Mb. For all markers, amplification was observed for wild-type Brundage, awnless plants from each M2 family, and from Chinese Spring. No amplification was obtained for deletion line 5AL-6, suggesting genome specificity of the primers. Deletions in the region were observed for all 53 awned plants selected from 17 segregating M2 families (Table 1). As expected with gamma irradiation, large deletions were observed with the majority of the M2 lines having lost more than 7.5 Mb of the surrounding region. The smallest deletion was observed for mutant line Br-187 (Fig. 5) and was estimated to be between 19.6 kb and 392.8 kb in size and included candidate genes *TraesCS5A02G542700* and *TraesCS5A02G542800*. Amplification of markers *5A10592* and *5A30334* indicated that *TraesCS5A02G542900* was

not deleted from Br-187, eliminating it as a candidate gene (Table 1; Table S1). An awnletted phenotype was observed for the *B1* hemizygous F_1 from the cross of Br-187 with awnletted cultivar NC-Neuse (Fig. 5).

Evaluation of awned and awnless wheats in the exome capture data set identified additional SNPs proximal to and within predicted gene *TraesCS5A02G542700* annotated as universal stress protein family with protein kinase domain. Haplotypes were compared to sequence data of the orthologous gene TRIUR3_34498 from *Triticum Urartu* (https://www.ebi.ac.uk/ena/data/view/GCA_000347455.1), the awned progenitor of the A-genome in wheat. A SNP unique to awned wheats and *T. urartu* predicted to create a missense mutation in exon 9 (*5A16541*) was targeted for development of a KASP assay. The winter wheat diversity panel was screened with the 5A16541 marker and previously developed marker for intronic SNP 5A15019. Of the 455 lines, 55 having the SS-MPV57 (*B1*) *5A15019* allele and 99 individuals with the SS-MPV57 *5A16541* allele possess awns, suggesting that neither of these polymorphisms in *TraesCS5A02G542700* underlie *B1* awn suppression.

4.4.3 Characterization of *B1* Candidate Gene *TraesCS5A02G542800*

Our fine mapping narrowed the *B1* locus to a 127 kb region of mostly repetitive sequences in the Chinese Spring reference genome containing predicted genes *TraesCS5A02G542700* and *TraesCS5A02G542800*. Marker analysis of wheat germplasm suggested that polymorphisms in predicted gene *TraesCS5A02G542700* were not predictive of awn suppression. Thus, *TraesCS5A02G542800*, annotated as a 366 bp single exon predicted C2H2 zinc finger transcription factor seemed a strong candidate for the dominant *B1* awn suppressor. A predicted protein sequence was used to identify related genes in the Uniprot database. While similar C2H2 zinc fingers in related grass species were identified, there are no publications characterizing these genes. The most similar characterized gene to *TraesCS5A02G542700* is KNUCKLES (KNU) in Arabadopsis thaliana, a C2H2 zinc finger involved in floral development sharing conserved zinc finger and N-terminal ethylene-associated response motifs (EAR-like) (Fig. 6).

Variation within the *TraesCS5A02G542700* coding sequence was not was found between the Chinese Spring RefSeq v1.0, scaffold assemblies of winter wheats with the *B1* sup-

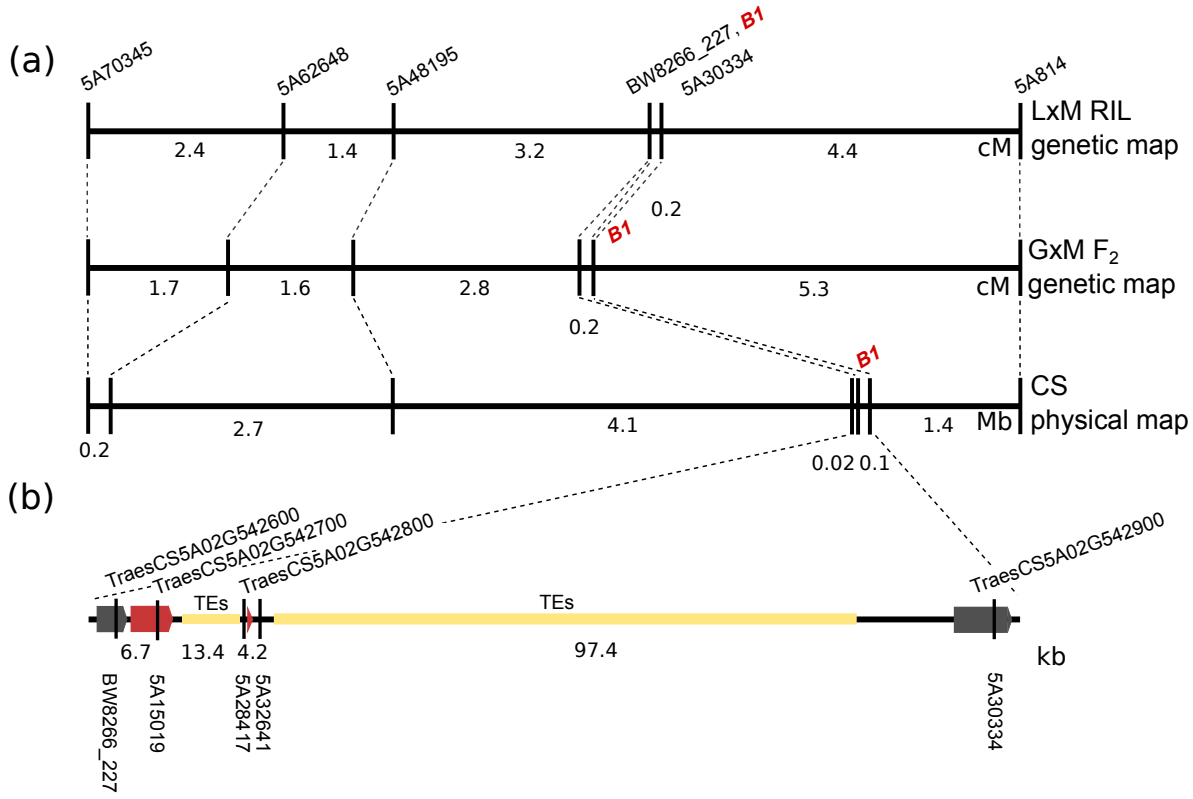

Figure 4.4: **Fine mapping of *B1* in wheat.** (a) Genetic distances in theB1region calculated from the LA95135xSS-MPV57 (LxM)recombinant inbred line (RIL) population and F_2 population GA06493-13LE6xSS-MPV57 (GxM), compared to physical distances obtained using the International Wheat Genome Sequencing Consortium RefSeqv1.0 Chinese Spring reference genome. (b) The fine-mapped *B1* region in Chinese Spring containing candidate gene *TraesCS5A02G542800* is shown relative to the genetic and physical maps. The *B1* locus co-segregates with single nucleotide polymorphism (SNP) markers *5A15019*, *5A28417* and *5A32641*. Genes co-segregating with awn status are highlighted in red. Approximate positions of SNP markers are shown, including the most significant genotyping by sequencing marker in both the RIL population and association mapping results(5A28417), and the marker most predictive of awn status in global germplasm (*5A32641*).

pressor (Cadenza, Paragon, Robigus, and Claire) and the awned tetraploid wheat Kronos (https://opendata.earlham.ac.uk/opendata/data/). In Chinese Spring, the gene is surrounded by more than 100kb of repetitive elements, hindering assembly and comparison of the region in the scaffold assemblies. In the available *B1* cultivar scaffolds, however, at least 10kb of additional repetitive elements were observed proximal to our candidate gene suggesting insertion or removal of transposable elements since the divergence of the

available *B1* assemblies and Chinese Spring.

Figure 4.5: **Deletion of *B1* in awned wheat mutants.** Wild-type Brundage wheat (Br-WT), Brundage mutant line 187 (Br-187) having the smallestdeletion in the terminal part of chromosome 5AL surrounding *TraesCS5A02G542800*. The awnletted phenotype is observed in the F_1 hybrid from the cross Br-187xNC-Neuse (*B1*).

B1 is located in a region of 5AL originating from a translocation with the 4A chromosome that occurred in the A-genome diploid progenitor of common wheat (Dvorak et al., 2018). Nucleotide BLAST was used to identify two regions on chromosomes 4B and 4D syntenic to the fine-mapped region. Sequences similar to the candidate C2H2 zinc finger were found in both regions, annotated as *TraesCS4D02G476700LC* on the long arm of chromosome 4D (78% nucleotide identity), and unannotated on chromosome 4B (80% nucleotide identity). The 4BL sequence contains a frameshift mutation and is not annotated in either the high confidence or low confidence gene set of the IWGSC v1.1 annotation (https://wheat-urgi.versailles.inra.fr/Seq-Repository/Annotations). *TraesCS4D02G476700LC* differs somewhat from *TraesCS5A02G542700* in the region between the zinc finger and EAR-like motifs, including a 10 amino acid insertion in the A-genome from positions 90-99 (Fig. 6). Phylogenetic analysis by Huang et al. (2019) grouped *B1* and its homoeologs with orthologous proteins of the progenitor species *T. urartu* and *Aegilops tauschii*.

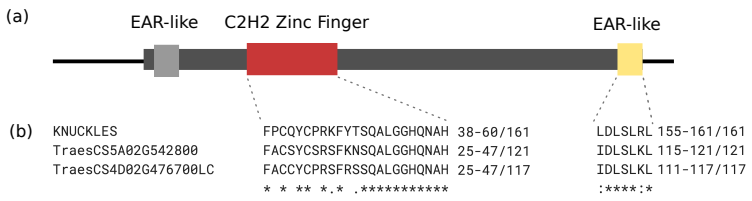

Figure 4.6: **Wheat *B1* is a C2H2 zinc finger protein with ethylene-associated response (EAR) motif.** (a) Gene model of *TraesCS5A02G542800* with potential conserved domains. (b) Alignment of functional motifs from characterized *Arabidopsis* C2H2 zinc finger KNUCKLES with candidate B1 gene *TraesCS5A02G542800* and its D-genome homoeolog *TraesCS4D02G476700LC*. Cystedeine and histidine residues responsible for zinc finger binding and core plant zinc finger sequence QALGGH are conserved between the wheat genes and KNUCKLES (Takatsuji, 1999 from Payneet al., 2004.) A C-terminal EAR-like motif implicated in recruitment of histones in KNUCKLES is conserved (Hiratsuet al., 2002 from Payneet al., 2004). A potential N-terminal EAR motif is not conserved.

4.4.4 Up-regulation of *TraesCS5A02G542800* is Associated with *B1* Awn Suppression.

In data from WheatEXP, *TraesCS5A02G542700* expression increases as the early inflorescence develops, and decreases after the emergence of the head (Fig. S1), while *TraesCS4D02G476700LC* increases in expression later in the development of the spike. The WheatFP browser (http://bar.utoronto.ca/efp_wheat/cgi-bin/efpWeb.cgi) (Ramírez-González et al. 2018; Winter et al. 2007) was used to localize gene expression in an awned spring variety Azhurnaya. *TraesCS5A02G5426800* is most highly expressed in the developing inflorescence, with some expression in the developing awns, ovaries, and grain tissues, while *TraesCS4D02G476700LC* is mostly expressed later in spike development (Fig. S2, S3). Similar analyses of predicted gene *TraesCS5A02G5426700* indicate that this gene is expressed in most tissues, with higher expression in spikes after the development of awn tissue (Fig. S4).

Gene expression data from qPCR determined that *TraesCS5A02G5426800* showed a 7-fold average increase in gene expression ($P < 1e\text{-}7$) in awnless plants compared to awned plants (Fig. 7) whereas *TraesCS5A02G5426700* was expressed similarly in both awned and awnless individuals (Fig. S5). Developing inflorescences were grouped into four stages based on development (Fig. 7b). Within each group, expression of *TraesCS5A02G5426800* was higher in awnless individuals, suggesting a mechanism of awn suppression associated

with this gene.

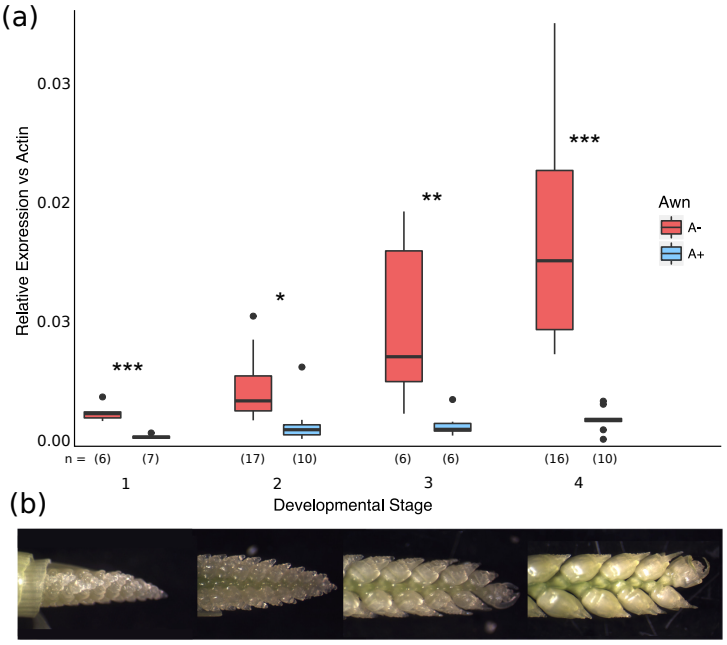

Figure 4.7: **Expression of *TraesCS5A02G542800* increases in spikes of awnless wheat.** (a)Expression of *TraesCS5A02G542800* in apical meristems of awned (blue, A+) and awnless (red, A) wheat plants at different developmental stages where 1 = younger meristems; and 4 = older meristems. Expression is relative to the reference gene $\beta - Actin$. Significant difference between awned and awnless individuals are indicated(***, $P < 0.001$; **, $P < 0.01$; *, $P < 0.05$).Number of biological replicates (n) is given per group, with three technical replications per biological replicate. (b) Representative spikes at each developmental stage.

Given fine-mapping of awn status to this interval, differences in expression imply the existence of cis-regulatory variation either directly upstream in the promoter of *TraesCS5A02G5426800* or in nearby enhancers or repressors. A 7 kb region of non-repetitive sequences surrounding the *B1* candidate gene (3 kb upstream and 4 kb downstream) was sequenced in a set of 10 individuals, including LA95135, SS-MVP57, and eight accessions of T. aestivum selected from the NSGC wheat core collection based on diversity in geographic origin and haplotypes of KASP markers in the region from 698.51 Mb to 698.62 Mb of 5AL (Table S4). No polymorphisms in the *TraesCS5A02G5426800* coding sequence were observed, and sequence variation proximal to the gene in the promoter region (including marker 5A28417) was not predictive of the awn suppression phenotype. A 30 bp deletion

4,005 bp downstream of the *TraesCS5A02G5426800* start codon was most predictive of awn suppression in this set. KASP marker 5A32641 designed around this deletion co-segregated with *B1* in the bi-parental mapping populations (Fig. 4).

4.4.5 Haploid Diversity in Global Wheat Germplasm

Marker 5A32641 was highly predictive of awn suppression in 2439 winter and spring wheat accessions in the USDA NSGC core set. Of the 455 winter wheat accessions evaluated 57% were awnless, compared to 45% of 1984 spring wheat accessions (Table S7). The 30 bp deletion distal to *TraesCS5A02G5426800* was present in 98% of awnless winter wheat accessions and all but 59 of 696 awnless spring lines (92%). Awnless lines without the 30 bp deletion downstream of *TraesCS5A02G5426800* may possess either the *Hd* or *B2* awn suppressors. Of the 1558 awned accessions, only 18 were homozygous for the deletion. Of these, 13 accessions are landraces from Sudan, Egypt, and Oman, suggesting they may share a rare genetic variant in either the candidate gene region or at the *B1* target. Overall, these results suggest that the dominant *B1* inhibitor is the primary determinant of awn suppression in wheat globally.

Eight haplotypes were identified when six SNP markers flanking *B1* were examined in conjunction with marker 5A32641 (Table 2). Marker haplotype 1 (Hap1) present in the awned cultivar LA95135 and Hap8 present in awnless cultivar SS-MPV57 were the most common, detected in 55% and 35% of accessions respectively (Fig. 8). The majority of awnless wheats were categorized as having the *B1*-associated Hap8. A small number of awnless spring wheat accessions from various geographic regions possessed Hap7 characterized by the 30 bp deletion distal to the candidate *B1* gene and differing from Hap8 at marker *5A28417* proximal to *B1*. In contrast, seven of the eight haplotypes were observed in awned accessions, with Hap1 being by far the most common. Higher haplotype diversity associated with the ancestral *B1* allele is expected assuming *B1* originated and spread during or after the domestication of cultivated wheat.

Hap1 was also observed in awnless accessions that may possess the *B1* allele and some combination of the *Hd* and *B2* alleles. Of these lines, 67% originated from central and south Asia, primarily Nepal and India, suggesting regional variation in control of awn suppression (Fig. 8). The greatest diversity in haplotypes was in accessions from Central Asia where

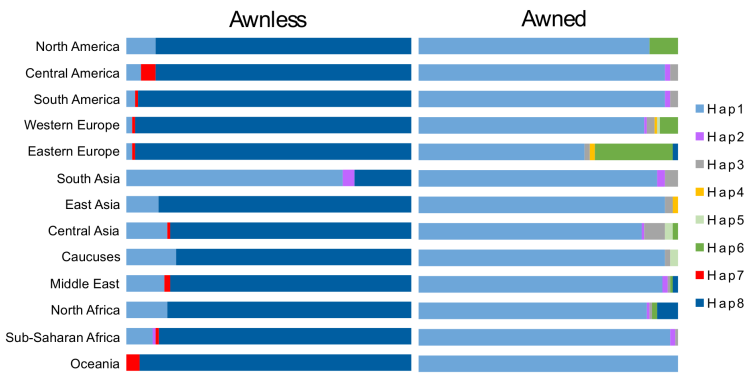

Figure 4.8: **Distribution of haplotypes in the *B1* region in 2439 global wheat accessions.** The majority of awn inhibited wheats possess *B1*-associated Hap8 and Hap7, implicating *B1* as the predominant determinant of awn suppression in all regions except South Asia. Hap1 that are awnless lines likely possess other awn-inhibiting genes. Although Hap1 is the predominant haplotype associated with the *b1* allele in awned accessions, rare haplotypes associated with geographical origin are observed. In Eastern Europe, Hap6 is a major haplotype associated with the presence of awns that is rare outside of Europe and North America. Lines from Central Asia contain a diverse set of haplotypes associated with the presence of awns. Bars are sized based on proportion awnless and awned accessions evaluated from each region.

all haplotypes except Hap4 were observed. Although all eight haplotypes were present in accessions from Western Europe, 95% of these accessions had either Hap1 or Hap8.

4.5 Discussion

Global variation for awn length in wheat suggests that the presence or absence of awns may be differentially adaptive to varying environments and production systems. Awns are known to contribute to yield in warmer, drier environments, cooling the wheat spike and supplying carbohydrates to developing grain (Grundbacher 1963; Kjack and Witters 1974; Motzo and Giunta 2002; Li et al. 2006; Tambussi et al. 2007; Amjad Ali et al. 2010; Maydup et al. 2014). A significant association of the *B1* awn suppressor with reduced test weight in our association panel suggests that awns influence kernel size in winter wheat grown in the southeastern United States, although there remains the possibility that a tightly linked variant co-segregating with *B1* is also impacting kernel morphology. In the bi-parental LM RIL population, QTL analysis confirmed that the presence of awns

was associated with increased kernel weight and greater kernel length, but with decreased spikelets per spike. In rice, which has non-photosynthetic awns, disabling GAD1 suppresses awn elongation, but also decreases grain length while increasing grains per panicle (Jin et al. 2016). Thus, a current hypothesis is that decreased grain size in awnless wheats may also be due to changes in floret development associated with awn inhibition, rather than simply decreased availability of photosynthate. Except in forage cultivars, the most economically important organ in wheat is the spike. As such, developing a better understanding of the gene networks that control spike development is critical for wheat breeders. Consistent with other studies, we observed that awn suppression was associated with an increase in the number of spikelets per spike (Rebetzke and Richards 2000). A major hindrance to increasing yield of modern wheat cultivars is the availability of sink tissues to fill with carbohydrates – increasing spikelet number is therefore one strategy to increase the number of grains harvested per unit area (Miralles and Slafer 2007). The final number of spikelets in wheat is correlated with the duration of the reproductive growth period, and genes associated with flowering time in wheat can influence the number of spikelets per spike (Guo et al. 2018). In our QTL mapping study, the segregating *Ppd-D1* flowering time allele was associated with QTL for both spikelets per spike and flowering time, but the *B1* region was associated only with spikelets per spike, indicating that *B1* influences the number of rachis nodes through a different mechanism.

Association of *B1* awn suppression with the up-regulation of candidate gene *TraesCS5A02G5426800* was identified in parallel by Huang et al. (2020) using RNA-sequencing of bulked awned and awnless lines. In addition, they find that constitutive over-expression of *TraesCS5A02G5426800* produces an awnletted phenotype in transgenic plants produced from an awned wild type. The closest characterized genes to *TraesCS5A02G5426800* as identified through protein BLAST are a set of zinc finger transcription factors in *Arabidopsis* , all of which contain conserved zinc finger and EAR-like motif domains. This family of transcription factors are usually repressors; the zinc finger domain binds to the target sequence, and the EAR-like motif recruits histones which down-regulate the target gene (Kagale and Rozwadowski 2011). The best characterized member of this family, KNUCKLES (KNU), disrupts stem cell maintenance in *Arabidopsis* (Payne et al. 2004). The floral homeotic protein AGAMOUS (AG) up-regulates KNU by binding to its promoter, with KNU in turn repressing the homeodomain protein WUSCHEL (WUS) until a specific developmental stage (Sun et al. 2009). WUS is responsible for the maintenance of stem cells in *Arabidopsis* , and over-expression of KNU preemptively terminates floral meristem

development (Sun et al. 2009). In this way, KNUCKLES represses growth of certain floral tissues to allow other floral tissues to develop at the same pace. If *B1* plays a similar role in wheat, its over-expression suppressing developing awn tissue, this suggests a potential explanation for the dominance of the *B1* allele. This is examined in our *B1* companion paper, where constitutive over-expression of *B1* suggests a role as a transcriptional repressor (Huang et al. 2020).

Haplotype analysis of global wheat germplasm confirms that *B1* is the predominant source of awn suppression in hexaploid wheat. The *B1* companion paper in this issue observed similar results in common wheat and identifies this gene as a source of awn suppression in tetraploid *T. turgidum* ssp. *durum* (Huang et al. 2019). Using six selected KASP assays, we identified two dominant haplotypes in the region of non-repetitive sequence flanking *B1*, with a smaller number of lines having haplotypes associated with geographic regions and the greatest number of haplotypes being present in central Asia and Europe. In our companion paper, Huang et al. (2020) identified six *B1*-like haplotypes when sequencing a 1961 bp region flanking the *B1* coding sequence and using a different set of markers near *B1*. They report a 25 bp deletion upstream of the candidate gene identified as being linked to *B1* but not entirely predictive, and as this deletion is found in *Triticum urartu* it is unlikely to be causative. We identified a 30 bp deletion approximately 4 kb downstream of the candidate gene that was not assayed by Huang et al. to be predictive of the *B1* awn suppressor in nearly all lines in this study. The marker 5A32641 assaying this deletion distinguished haplotypes 1-6 associated with *B1* from haplotypes 7-8 associated with *B1*. Alignment of the regions in the B and D genomes syntenic to *B1* reveal major variation in transposable elements both upstream and downstream of the candidate gene with expansion observed in the A genome. In addition, assemblies of the region in wheat lines carrying the *B1* allele reveal further TE expansion in the region compared to lines carrying the *B1* allele. Determining if the observed up-regulation of *B1* is a product of a binding site mutation, a by-product of TE insertions, or some other causative polymorphism will require further work.

Wheat's large genome size, hexaploidy, and long-range linkage disequilibrium can make association mapping more challenging than in other species. Using next-generation sequencing technologies, wheat geneticists have created communal resources that facilitate fine mapping and identification of candidate genes. The newly published wheat reference genome, tools that make it accessible to other researchers, and technologies that take

advantage of the reference genome (such as gene expression and exome capture data sets) were all used in this study to reduce the time and cost of positional cloning. In parallel, Huang et al. (2020) also identified overexpression of *TraesCS5A02G5426800* as underlying *B1*, through a combination of bulked segregant RNA-seq and fine-mapping in durum wheat, deletion mapping in UK bread wheats, and transgenic validation. Cloning of *B1* should allow for further experiments to identify downstream targets and effects of this gene, and to explore its interaction with other developmental genes. Improved understanding of the relationship between awn development and control of spikelets per spike, grain number and grain size in wheat will provide insight in to pathways that can be manipulated to increase grain yield. Characterization of *B1* adds to our growing understanding of the gene networks underlying spike development in wheat and may help breeders produce varieties better adapted to local climate and end-use.

4.6 Acknowledgements

The authors thank staff of the USDA-ARS Plant Science Unit for assistance in genotyping and growing field trials, collaborating breeding programs for data from evaluation of GAWN and SunWheat yield trials, and the J. Dubcovsky lab at the University of California Davis for use of instrumentation for phenotyping grain traits. Support was provided by the Agriculture and Food Research Initiative Competitive Grant 67007-25939 (WheatCAP-IWYP) from the USDA NIF.

4.7 Author Contributions

ND, MG, and GBG planned and designed the research. ND, MG, EL, and MS performed experiments and collected phenotypes. ND, EL, and PT analyzed genome-wide sequencing data. QH and DF developed and phenotyped the mutant population. JPM and DM planted and maintained field experiments. AA, KJ and EA performed and analyzed exome capture data of parents. ND, MG, and GBG wrote the manuscript.

References

Amjad Ali, M., Ali, Z., Zulkiffal, M., and Zeeshan AARI, M. (2010). Source-Sink Relationship between Photosynthetic Organs and Grain Yield Attributes during Grain Filling Stage in Spring Wheat (Triticum aestivum). Technical report.

Broman, K. W., Wu, H., Sen, S., and Churchill, G. A. (2003). R/qtl: QTL mapping in experimental crosses. *Bioinformatics*, 19(7):889–890.

Cao, L., Hayashi, K., Tokui, M., Mori, M., Miura, H., and Onishi, K. (2016). Detection of QTLs for traits associated with pre-harvest sprouting resistance in bread wheat (<i>Triticum aestivum</i> L.). *Breeding Science*, 66(2):260–270.

Cash, S. D., Bruckner, P. L., Wichman, D. M., Kephart, K. D., Berg, J. E., Hybner, R., Hafla, A. N., Surber, L. M. M., Boss, D. L., Carlson, G. R., Eckhoff, J. L., Stougaard, R. N., Kushnak, G. D., and Riveland, N. R. (2009). Registration of 'Willow Creek' Forage Wheat. *Journal of Plant Registrations*, 3(2):185–190.

Elbaum, R., Zaltzman, L., Burgert, I., and Fratzl, P. (2007). The role of wheat awns in the seed dispersal unit. *Science*, 316(5826):884–886.

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., and Kawamoto, K. (2011). A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. *PLoS ONE*, 6(5):19379.

Evans, L. T., Bingham, J., Jackson, P., and Sutherland, J. (1972). Effect of awns and drought on the supply of photosynthate and its distribution within wheat ears. *Annals of Applied Biology*, 70:67–76.

Glaubitz, J. C., Casstevens, T. M., Lu, F., Harriman, J., Elshire, R. J., Sun, Q., and Buckler, E. S. (2014). TASSEL-GBS: A High Capacity Genotyping by Sequencing Analysis Pipeline. *PLoS ONE*, 9(2):e90346.

Grundbacher, F. J. (1963). The Physiological Function of the Cereal Awn. pages 366–381.

Guo, Z., Chen, D., Röder, M. S., Ganal, M. W., and Schnurbusch, T. (2018). Genetic dissection of pre-anthesis sub-phase durations during the reproductive spike development of wheat. *The Plant Journal*, 95(5):909–918.

Huang, D., Zheng, Q., Melchkart, T., Bekkaoui, Y., Konkin, D. J. F., Kagale, S., Martucci, M., You, F. M., Clarke, M., Adamski, N. M., Chinoy, C., Steed, A., McCartney, C. A., Cutler, A. J., Nicholson, P., and Feurtado, J. A. (2020). Dominant inhibition of awn development by a putative zinc-finger transcriptional repressor expressed at the B1 locus in wheat. *New Phytologist*, 225(1):340–355.

Jin, J., Hua, L., Zhu, Z., Tan, L., Zhao, X., Zhang, W., Liu, F., Fu, Y., Cai, H., Sun, X., Gu, P., Xie, D., and Sun, C. (2016). GAD1 Encodes a Secreted Peptide That Regulates Grain Number, Grain Length, and Awn Development in Rice Domestication. *The Plant Cell*, 28(10):2453–2463.

Kagale, S. and Rozwadowski, K. (2011). EAR motif-mediated transcriptional repression in plants. *Epigenetics*, 6(2):141–146.

Kato, K., Miura, H., Akiyama, M., Kuroshima, M., and Sawada, S. (1998). RFLP mapping of the three major genes, Vrn1, Q and B1, on the long arm of chromosome 5A of wheat. Technical report.

King, R. and Richards, R. (1984). Water uptake in relation to pre-harvest sprouting damage in wheat: ear characteristics. *Australian Journal of Agricultural Research*, 35(3):327.

Kjack, J. and Witters, R. (1974). Physiological activity of awns in isolines of Atlas barley. *Crop Science*, 14(2):243–248.

Krasileva, K. V., Vasquez-Gross, H. A., Howell, T., Bailey, P., Paraiso, F., Clissold, L., Simmonds, J., Ramirez-Gonzalez, R. H., Wang, X., Borrill, P., Fosker, C., Ayling, S., Phillips, A. L., Uauy, C., and Dubcovsky, J. (2017). Uncovering hidden variation in polyploid wheat. *Proceedings of the National Academy of Sciences*, 114(6):E913–E921.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16):2078–2079.

Li, X., Wang, H., Li, H., Zhang, L., Teng, N., Lin, Q., Wang, J., Kuang, T., Li, Z., Li, B., Zhang, A., and Lin, J. (2006). Awns play a dominant role in carbohydrate production during the grain-filling stages in wheat (Triticum aestivum). *Physiologia Plantarum*, 127(4):701–709.

Lipka, A. E., Tian, F., Wang, Q., Peiffer, J., Li, M., Bradbury, P. J., Gore, M. A., Buckler, E. S., and Zhang, Z. (2012). GAPIT: Genome association and prediction integrated tool. *Bioinformatics*, 28(18):2397–2399.

Ma, C.-Y., Gao, L.-Y., Li, N., Li, X.-H., Ma, W.-J., Appels, R., and Yan, Y.-M. (2012). Proteomic Analysis of Albumins and Globulins from Wheat Variety Chinese Spring and Its Fine Deletion Line 3BS-8. *International Journal of Molecular Sciences*, 13(12):13398–13413.

Mackay, I. J., Bansept-Basler, P., Bentley, A. R., Cockram, J., Gosman, N., Greenland, A. J., Horsnell, R., Howells, R., O'Sullivan, D. M., Rose, G. A., and Howell, P. J. (2014). An eight-parent multiparent advanced generation inter-cross population for winter-sown wheat: Creation, properties, and validation. *G3: Genes, Genomes, Genetics*, 4(9):1603–1610.

Maydup, M. L., Antonietta, M., Graciano, C., Guiamet, J. J., and Tambussi, E. A. (2014). The contribution of the awns of bread wheat (Triticum aestivum L.) to grain filling: Responses

to water deficit and the effects of awns on ear temperature and hydraulic conductance. *Field Crops Research*, 167:102–111.

Miralles, D. and Slafer, G. (2007). Sink limitations to yield in wheat: how could it be reduced? *The Journal of Agricultural Science*, 145(02):139.

Motzo, R. and Giunta, F. (2002). Awnedness affects grain yield and kernel weight in near-isogenic lines of durum wheat. *Australian Journal of Agricultural Research*, 53(12):1285–1293.

Nishijima, R., Ikeda, T. M., and Takumi, S. (2018). Genetic mapping reveals a dominant awn-inhibiting gene related to differentiation of the variety anathera in the wild diploid wheat Aegilops tauschii. *Genetica*, 146(1):75–84.

Payne, T., Johnson, S. D., and Koltunow, A. M. (2004). KNUCKLES (KNU) encodes a C2H2 zinc-finger protein that regulates development of basal pattern elements of the Arabidopsis gynoecium. *Development*, 131(15):3737–3749.

Peleg, Z., Saranga, Y., Fahima, T., Aharoni, A., and Elbaum, R. (2010). Genetic control over silica deposition in wheat awns. *Physiologia Plantarum*, 140(1):10–20.

Poland, J., Endelman, J., Dawson, J., Rutkoski, J., Wu, S., Manes, Y., Dreisigacker, S., Crossa, J., Sánchez-Villeda, H., Sorrells, M., and Jannink, J. L. (2012). Genomic selection in wheat breeding using genotyping-by-sequencing. *Plant Genome*, 5(3):103–113.

Ramírez-González, R. H., Borrill, P., Lang, D., Harrington, S. A., Brinton, J., Venturini, L., Davey, M., Jacobs, J., van Ex, F., Pasha, A., Khedikar, Y., Robinson, S. J., Cory, A. T., Florio, T., Concia, L., Juery, C., Schoonbeek, H., Steuernagel, B., Xiang, D., Ridout, C. J., Chalhoub, B., Mayer, K. F. X., Benhamed, M., Latrasse, D., Bendahmane, A., Wulff, B. B. H., Appels, R., Tiwari, V., Datla, R., Choulet, F., Pozniak, C. J., Provart, N. J., Sharpe, A. G., Paux, E., Spannagl, M., Bräutigam, A., Uauy, C., Korol, A., Sharpe, A. G., Juhász, A., Rohde, A., Bellec, A., Distelfeld, A., Akpinar, B. A., Keller, B., Darrier, B., Gill, B., Chalhoub, B., Steuernagel, B., Feuillet, C., Chaudhary, C., Uauy, C., Pozniak, C., Ormanbekova, D., Xiang, D., Latrasse, D., Swarbreck, D., Barabaschi, D., Raats, D., Sergeeva, E., Salina, E., Paux, E., Cattonaro, F., Choulet, F., Kobayashi, F., Keeble-Gagnere, G., Kaur, G., Muehlbauer, G., Kettleborough, G., Yu, G., Šimková, H., Gundlach, H., Berges, H., Rimbert, H., Budak, H., Handa, H., Small, I., Bartoš, J., Rogers, J., Doležel, J., Keilwagen, J., Poland, J., Melonek, J., Jacobs, J., Wright, J., Jones, J. D. G., Gutierrez-Gonzalez, J., Eversole, K., Nilsen, K., Mayer, K. F., Kanyuka, K., Singh, K., Gao, L., Concia, L., Venturini, L., Cattivelli, L., Spannagl, M., Mascher, M., Hayden, M., Abrouk, M., Alaux, M., Luo, M., Valárik, M., Benhamed, M., Singh, N. K., Sharma, N., Guilhot, N., Ravin, N., Stein, N., Olsen, O.-A., Gupta, O. P., Khurana, P., Chhuneja, P., Bayer, P. E., Borrill, P., Leroy, P., Rigault, P., Sourdille, P., Hernandez, P., Flores, R., Ramirez-Gonzalez, R. H., King, R., Knox, R., Appels, R., Zhou, R., Walkowiak, S., Galvez, S., Biyiklioglu, S., Nasuda, S., Sandve, S., Chalabi, S., Weining, S., Sehgal, S., Jindal, S., Belova, T., Letellier, T., Wicker, T., Tanaka, T., Fahima, T., Barbe, V., Tiwari, V.,

Kumar, V., and Tan, Y. (2018). The transcriptional landscape of polyploid wheat. *Science*, 361(6403).

Rebetzke, G. J., Bonnett, D. G., and Reynolds, M. P. (2016). Awns reduce grain number to increase grain size and harvestable yield in irrigated and rainfed spring wheat. *Journal of Experimental Botany*, 67(9):2573–2586.

Rebetzke, G. J. and Richards, R. A. (2000). Gibberellic acid-sensitive dwarfing genes reduce plant height to increase kernel number and grain yield of wheat. *Australian Journal of Agricultural Research*, 51(2):235–245.

Sourdille, P., Cadalen, T., Gay, G., Gill, B., and Bernard, M. (2002). Molecular and physical mapping of genes affecting awning in wheat. *Plant Breeding*, 121:320–324.

Sun, B., Xu, Y., Ng, K.-H., and Ito, T. (2009). A timing mechanism for stem cell maintenance and differentiation in the Arabidopsis floral meristem. *Genes & Development*, 23(15):1791–1804.

Tambussi, E. A., Bort, J., Guiamet, J. J., Nogués, S., and Araus, J. L. (2007). The Photosynthetic Role of Ears in C3 Cereals: Metabolism, Water Use Efficiency and Contribution to Grain Yield. *Critical Reviews in Plant Sciences*, 26(1):1–16.

Tambussi, E. A., Nogués, S., and Araus, J. L. (2005). Ear of durum wheat under water stress: Water relations and photosynthetic metabolism. *Planta*, 221(3):446–458.

Taylor, J. and Butler, D. (2017). R package ASMap: Efficient genetic linkage map construction and diagnosis. *Journal of Statistical Software*, 79.

Toriba, T. and Hirano, H.-Y. (2014). The DROOPING LEAF and OsETTIN2 genes promote awn development in rice. *The Plant Journal*, 77(4):616–626.

Watkins, A. E. and Ellerton, S. (1940). Variation and genetics of the awn in Triticum. *Journal of Genetics*, 40(1-2):243–270.

Winter, D., Vinegar, B., Nahal, H., Ammar, R., Wilson, G. V., and Provart, N. J. (2007). An "Electronic Fluorescent Pictograph" Browser for Exploring and Analyzing Large-Scale Biological Data Sets. *PLoS ONE*, 2(8):e718.

Yoshioka, M., Iehisa, J. C., Ohno, R., Kimura, T., Enoki, H., Nishimura, S., Nasuda, S., and Takumi, S. (2017). Three dominant awnless genes in common wheat: Fine mapping, interaction and contribution to diversity in awn shape and length. *PLoS ONE*, 12(4).

Chapter 5

Bearded or smooth? Awns improve yield when wheat experiences heat stress during grain fill

Noah DeWitt[1, 2*] Mohammed Guedira[1] Jeanette Lyerly[1] Brian P. Ward[3] J. Paul Murphy[1] David Marshall[2] Nicholas Santantonio[4] Carl Griffey[4] Richard E. Boyles[5] Mohamed Mergoum[6] Stephen Harrison[7] Md Ali Babar[8] Richard E. Mason[9] Amir Ibrahim[10] David A. Van Sanford[11] Vijay Tiwari[12] Gina Brown-Guedira[1, 2]

1 Department of Crop and Soil Sciences, North Carolina State University, Raleigh, NC 27695
2 USDA-ARS SEA, Plant Science Research, Raleigh, NC 27695
3 Ohio Agricultural Research and Development Center, The Ohio State University, Wooster, OH
4 Department Of Crop and Soil Environmental Sciences, Virginia Tech, Blacksburg, VA
5 Pee Dee Research and Education Center, Clemson University, Florence, SC 29506
6 Department of Crop and Soil Sciences, University of Georgia, Athens, GA 30602
7 Department of Agronomy, Louisiana State University, Baton Rouge, LA 70803
8 Agronomy Department, University of Florida, Gainesville, FL 32611
9 Soil and Crop Sciences Department, Colorado State University, Fort Collins, CO 80523
10 AgriLife Research, Texas A&M University, College Station, TX 77843
11 Department of Plant and Soil Sciences, University of Kentucky, Lexington, KY

12 Department of Plant Science and Landscape Architecture, University of Maryland, College Park, MD 20742

Prepared for submission to *Crop Science*

5.1 Abstract

The presence or absence of awns – whether a wheat line is ”bearded” or not – is the most visible phenotype distinguishing wheat cultivars, but the contribution of awns to yield differences remains unclear. Previous studies suggest the potential for awns to improve yields in heat or water-stressed environments. Here we leverage historical phenotypic, genotypic, and climate data to estimate the yield effects of awns under different environmental conditions over a 12 year period in the Southeast US. Lines were classified as awned or awnless and observed heading dates were used to associate grain fill period of each line with climatic data and grain yield. Evidence was found for a positive yield effect of awns, but only in heat and water stressed environments more common at more southern locations. Breeders in environments where awns are only beneficial in some years may consider selection for awned cultivars to both reduce year-to-year yield variability, and with an eye towards future warming climates.

5.2 Introduction

Humans rely on wheat (*Triticum aestivum L.*) more than any other crop species for both direct calories and protein (FAO 2020). Despite its global range, wheat is more sensitive to heat and drought stress than other grains like maize (*Zea mays* L.) and sorghum (*Sorghum bicolor* L.), in part due to a lower water use efficiency compared to its C4 relatives (Aggarwal and Sinha 1983). Heat stress hampers plants' ability to fill grain, through decreasing sucrose levels and the overall duration of grain fill, and through deactivating the wheat starch synthase enzyme responsible for assimilating carbohydrates into developing grain (Bhullar and Jenner 1985; Hawker and Jenner 1993; Guedira and Paulsen 2002). The simplest way to breed for drought and heat tolerance in areas where these stresses are common is to plant yield trials in those environments. However, as climate change threatens higher temperatures and altered precipitation patterns in areas without those stresses, an understanding of the genetic and physiological basis of drought and heat stress response will be necessary to breed for future conditions.

A principal morphological trait implicated in drought and heat stress response is the presence or absence of awns or "beards". Awns are thin, rigid extensions of the lemma that

give wheat heads a brush-like appearance. In US germplasm, this trait is also a genetic marker – the presence of absence of awns is determined by the *B1* awn suppressor gene at the end of chromosome 5A (Kato et al. 1998; DeWitt et al. 2020). Awn suppression results from over-expression of the *B-A1* C2H2 zinc finger (Huang et al. 2020; DeWitt et al. 2020). Suppression of awns by *B1* produces awn lengths ranging from a totally awnless phenotype to an awnletted phenotype associated with short awns present near the tip of the spike (Watkins and Ellerton 1940). For simplicity, we will hereafter refer to all these phenotypes as "awnless".

As a typically monogenic and visually striking trait, studies on awns constitute some of the earliest in genetics research. In Ohio field trials, Hickman (1889) found that "bearded" cultivars out-yielded "smooth" cultivars 40.5 to 37.5 bu/acre. As early as 1937, it was "generally agreed that in certain cultivars awned cultivars out-yield awnletted or awnless ones", and that "awned cultivars have been selected and grown in regions of limited rainfall, for awnless types are generally not well-adapted to such conditions" (Gauch 1937). The distribution of awned and awnless types in wheat cultivars globally supports this observation (Börner et al. 2005; DeWitt et al. 2020). Anecdotally, breeders targeting environments warmer and dryer during the late growing season are disinclined to select awnless cultivars, while breeders in cooler climates may prefer to select awnless cultivars.

Research into the yield effects of awns, and the role of heat and drought stress in those effects, has centered on the photosynthetic potential of awn tissue (Grundbacher 1963; Tambussi et al. 2005; Amjad Ali et al. 2010; Motzo and Giunta 2002; Li et al. 2006). The movement of carbohydrates from awns into developing grain is facilitated by their physical proximity, and awns may consist of up to half the photosynthetic area of late-season wheat plants. As might be expected for a trait which has an environmentally conditional effect on yield, estimates of the yield effects of awns have shown some cases where awned lines outperform their awnless counterpart (Vervelde 1953; Weyhrich et al. 1994; Motzo and Giunta 2002; Martin et al. 2003), but also some where no significant differences are identified (Foulkes et al. 2007; Rebetzke et al. 2016; Sanchez-Bragado et al. 2020). Awns are more effective at photosynthesizing in water-limited conditions, using less water per quantity of CO_2 assimilated, and contribute a greater portion of carbohydrates relative to other tissues in water-limited and heat-stressed conditions (Evans et al. 1972; Blum 1986; Weyhrich et al. 1995; Li et al. 2006; Maydup et al. 2014). Awn tissue often senesces last, and has a higher temperature optimum for photosynthesis than flag leaf and spike tissue, which may

contribute to grain yield in heat-stressed conditions (Blum 1986). At the same time, adaptive advantage of awns in heat-stress environments may relate directly to the ability of awns to help cool ears by transferring heat into the atmosphere. In studies of canopy temperature, awns cool the head during grain fill (Ferguson 1975; Motzo and Giunta 2002; Rebetzke et al. 2016), which may reduce stress in a crop sensitive to heat during grain fill. The ability of awns to cool canopy temperatures and more efficiently photosynthesize in water-limited conditions may explain the perceived yield benefits of awned cultivars in warmer and drier climates. In general, awns tend to be associated with an increase in grain weight at the expense of grain number (Rebetzke et al. 2016; DeWitt et al. 2020; Sanchez-Bragado et al. 2020). *B1* increasing total seed number may explain the prevalence of awnless varieties in some environments, as increased seed number may lead to greater yields in environments where plants are not source-limited. *B1* increases seed number through increasing the number of spikelets per spike, an effect that may be related to resource allocation towards awns in the developing spike, or that may result from the effect on developmental pathways of the increase in *B1* expression also associated awn suppression.

Previous studies on the effect of awns on grain yield have focused on comparisons of near-isogenic lines or a small number of lines in a limited sample of environments. To this point no studies have directly estimated reaction norms for awn status in interaction with environmental conditions. Here we use a large historic data set consisting of real breeding lines where the frequency of *B1* is close to 50% to estimate the allele by environment effect of *B1*. Previous genotyping of these lines and identification of a SNP near *B-A1* predictive of awn status allows us to both classify historic lines into awned and awnless, and to control for population structure with genotypic data when estimating allele effects. Observed heading date was used to estimate grain fill periods for each line in each environment, and weather data collected from public databases was averaged across this period to test the interaction of *B1* with environmental variables. We use the estimated effect of awns on yield in interaction with maximum grain fill temperatures to investigate the utility of selection for or againsted awned cultivars in different environments.

5.3 Materials and Methods

5.3.1 Phenotypic data

Data on yield and heading date was collected for plots in cooperative nurseries representing 1,376 late-stage wheat lines submitted by participating public breeding programs in the Southern and Mid-Atlantic winter wheat growing region of the United States. Participating breeding programs were located at the University of Arkansas, Clemson University, the University of Florida, the University of Kentucky, Louisiana State University, the University of Maryland, North Carolina State University, Virginia Polytechnic Institute and State University, and the USDA-ARS SEA in Raleigh. Data from three cooperative nurseries (the Mason-Dixon, Gulf Atlantic Wheat Nursery, and SunWheat) were used, to which individual programs annually submitted entries that were planted once at all sites, but typically within only a single year, excluding check cultivars that were planted across years to estimate year effects (Table **??**). Individual trials consisted of two to four replications of each line, and analyzed data consisted of entry-means calculated by collaborators for each line in each site-year combination. Trials were planted in the October or November prior to the harvest year and were at minimum 1.3 m wide and 3.1 m long. Management of plots varied by collaborator depending on state-specific recommendations. Yield was reported in bu/acre based on whole-plot harvesting and weighing of grain, assuming a consistent test weight of 60 lb/bu and adjusting for moisture content when appropriate. Heading date values for lines were recorded as the date on which approximately half of wheat heads had emerged.

Nursery	Years	Mean # Sites	Avg # Lines	Entry Means	Awned
Mason-Dixon	2012-2018	3.6	76.94	1,385	46%
GAWN	2008-2020	5.5	58.82	4,235	52%
SunWheat	2014-2020	4.6	85.03	2,721	64%

Table 5.1: **Distribution of entry means among lines.** Year ranges, mean sites per year, average lines per year, max and minimum lines per year, total number of entry years, and percentage of tested lines with awns are given.

5.3.2 Genotypic data

Tissue samples from nursery lines were submitted to the USDA-ARS Eastern Regional Small Grains Genotyping Lab for genotyping via genome-wide single nucleotide polymorphism (SNP) markers. A genotype by sequencing approach using the PstI and MstI enzymes was performed per Poland et al. (2012) to allow for joint discovery and calling of SNPs. SNP discovery was performed through alignment of GBS reads from both the nursery lines phenotyped in this study and additional soft red winter wheat (SRWW) lines to capture additional variants that may be at low frequency in the study set. Reads were aligned to version 1.0 of the International Wheat Genome Sequencing Consortium RefSeq assembly (https://wheat-urgi.versailles.inra.fr/Seq-Repository/Assemblies) using version 0.7.12 of the Burrows-Wheeler algorithm (Li et al. 2009) via Tassel 5GBSv2 pipeline version 5.2.35 (Glaubitz et al. 2014). The resulting Tassel discovery database was used to call SNPs using reads from the nursery lines, and called SNPs were filtered to exclude markers with more than 50% total missing data, less than 5% minor allele frequency, and greater than 10% heterozygosity given the expected inbreeding of genotyped lines. Beagle was used to impute missing SNPs (Browning and Browning 2016).

DeWitt et al. (2020) found that in US SRWW germplasm, awn status is a monogenic trait controlled by the *B1* gene. They identified a single GBS marker on chromosome 5A at 598628417 bp (Refseq v1.0 coordinates) predictive of *B1* in 98% of 640 SRWW lines that constitute a subset of the lines in this data set. This SNP was also identified in the SNP discovery database of the full set of SRWW lines, and used to determine allelic status of *B1* in genotypes for which data on awn status was not available.

Covariance between allele frequency of *B1* and population structure could bias estimates of interactions between awns and environmental factors. GBS data was used to estimate realized relationships between nursery lines for later population structure analysis. Markers were first thinned based on LD ($r^2 < 0.8$), before being coded in terms of the allele dosage of the minor allele. The realized relationship matrix **G** was computed via the Van Raden method (VanRaden 2008) as:

$$\mathbf{G} = \frac{\mathbf{Z}\mathbf{Z}'}{\Sigma 2p(1-p)}$$

Where p represents a vector of allele frequencies of all SNP in the (-1, 0, 1) coded marker matrix **M**, and **Z** = **M** - **P**, where **P** is a matrix consisting of the vector p repeated rowwise to match the dimensions of **M**. Principal components of **G** were computed using the *eigen* function in R (R Core Team 2021).

5.3.3 Environmental data

Weather data for the GPS coordinates associated with each site in each year were downloaded from NASA POWER using the nasaPower R package (Sparks 2018). Within each site-year, principal grain fill periods were defined for each observed heading date by taking the time from the day after heading through twenty five days. To test optimum grain filling periods for each weather variable, additional variables were collected from twenty to thirty days after heading. Literature suggests an effect of awns in improving photosynthesis during grain fill in drought and heat stress conditions as a result of better heat conductance, increasing water use efficiency with less transpiration, and later senescence. Therefore, data on max temperature, precipitation, wind speed, and relative humidity were collected. Within each grain fill period associated with a heading date in a site-year, collected weather variables were averaged.

5.3.4 Testing and predicting weather effects

To estimate effects of *B1* on grain yield, a mixed linear model was fit to allow for the estimate of interactions between *B1* and environmental variables while controlling for effects of population structure and its interaction with environmental variables. Weather variables are relative to heading date to allow for a range of variables within each site year associated with each genotype's mean heading date. Mean yield values for each line were computed within each site-year as y_{ijk} by collaborators prior to data aggregation. For awn status a of line i, matrix of principal components **Q** with a vector of values for each line i, and matrix of weather variables **W** with a vector of values for each line i within trial j in site-year k, the data was analyzed using the model:

$$y_{ijk} \sim a_i + \mathbf{Q}_i + \mathbf{W}_{ijk} + a_i\mathbf{W}_{ijk} + \mathbf{Q}_i\mathbf{W}_{ijk} + g_i + l_k + t(l)_{jk} + \varepsilon_{ijk}$$

Where g_i represents the genotype effect of line i fit with the realized relationship matrix ($g_i \sim \mathcal{N}(0, \mathbf{G}\sigma^2_g)$), l_k represents the random effect of the kth site-year with $l_k \sim \mathcal{N}(0, \sigma^2_l)$, and $t(l)_{jk}$ represented the nested random effect of the jth trial within the kth site-year with $t_{jk} \sim \mathcal{N}(0, \sigma^2_t)$. The residual ε_{ijk} is drawn from $\varepsilon_{ijk} \sim \mathcal{N}(0, \mathbf{R}\sigma^2_\varepsilon)$, where the diagonal of **R** contains separate variances for different trials to accommodate heterogenous variances across trials ($\mathbf{R} = \sigma^2_{r1}\mathbf{I} \oplus \sigma^2_{r2}\mathbf{I} \oplus ... \sigma^2_{rn}\mathbf{I}$ for n trials). The model was fit in Asreml-R, and significance of main fixed effects was computed via the Wald test (Butler et al. 2017).

Estimates of the fixed effects were used to derive reaction norms for allele effects of *B1* in interaction with significant environmental variables. Reaction norm equations were used to estimate allele effects of awns in each site-year, based on values for environmental variables observed in those site-years. Assuming that trial locations within states are representative of breeders' target environments, site-year effects were grouped by state and averaged to estimate mean allele effects of awns within each state.

5.4 Results and Discussion

5.4.1 Population Structure of Programs

To be useful in modeling the relative performance of lines from different programs under different environmental conditions, the principal components of the realized relationship matrix should capture genetic differences between programs resulting from patterns of parental selection. At the same time, severe population stratification may hinder accurate estimation of allele effects. The first four PCs of **G** together captured 14% of the total variation in relationships between 1,371 genotypes. Public programs were separated by these first four components (Fig. **??**), but without severe stratification. This may be expected from the structure of programs, which are distinct but cooperative and regularly share germplasm. Averaging the values of all individuals within each program shows that the first PC separated out the mid-Atlantic programs (MD, VA, KY, NC) from southern programs (FL, LA, TX, GA, SC), while the second PC separated programs within the mid-Atlantic and southern groups (Fig. **??**). The fourth PC was also effective at separating the most southern programs that may be more heat-stress resistant (FL, LA, and TX) from other germplasm. Discrepancies from previous studies using this germplasm (e.g. Sarinelli et al. (2019)), which showed the

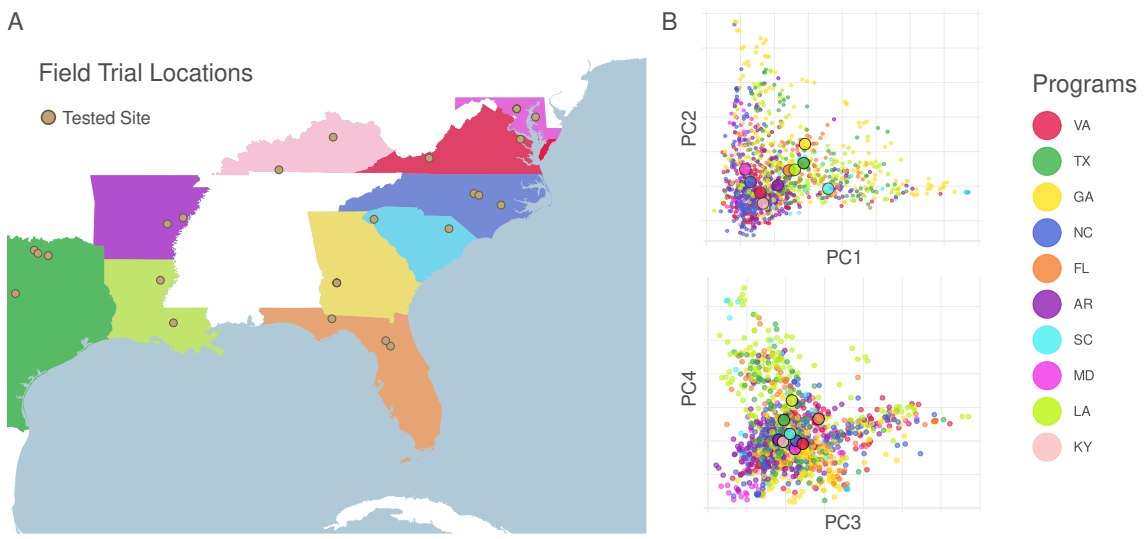

Figure 5.1: **Geographic and genetic distance between public programs.** Public-sector breeding programs contributing germplasm and phenotypes to tested nurseries are colored, and field site locations labeled (A). The distance between individuals in the space of the first four principal components of the realized relationship matrix **G** reflect geographic separation of the programs (B).

principal components being largely reflective of the presence or absence of SNP-dense translocations, result from LD thinning of the marker matrix performed prior to estimation of the realized relationships between lines.

5.4.2 Environmental factors relevant to awn yield effects

Weather variables were selected based off of their role in generating heat stress or altering transpiration rates: maximum temperature, precipitation, specific humidity, and wind speed information was obtained for the growing season of each site year. An initial grain fill window of 22 days after heading was assumed for initial testing of the importance of variables, and the heading date of each plot in each site year was used to average weather variables over that window. Of the tested variables, only maximum temperature was found to have a significant interaction with awn status after controlling for population structure. Testing possible alternative grain fill periods longer and shorter than 22 days, the grain fill period of 23 days produced the strongest interaction between average daily maximum temperature and awn status ($p < 2.269E-09$). The first four principal components of

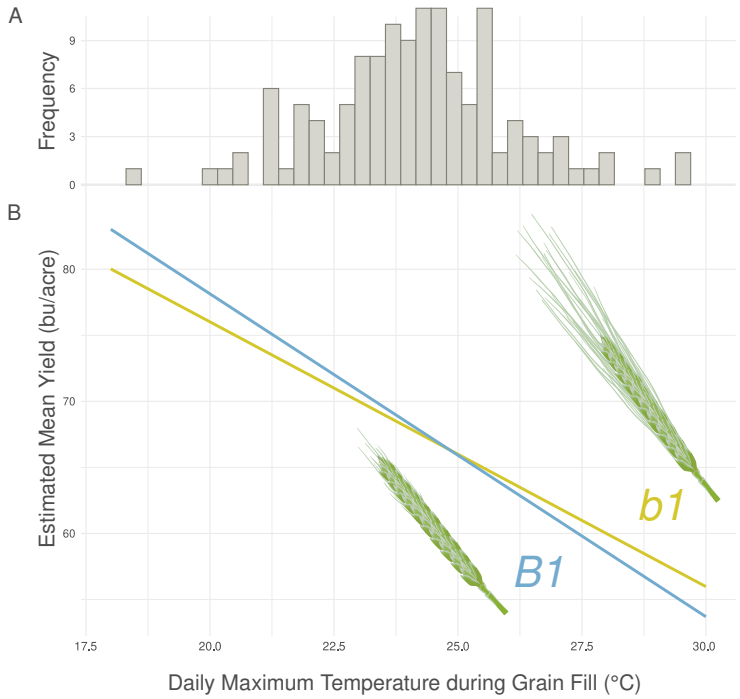

Figure 5.2: **Interaction betwee B1 and max grain fill temperature.** Average maximium daily temperature during grain fill at each site year was 24.1 $°C$ (A), with a high maximum grain fill temperature of 29.8$°C$ (Gainesville, FL 2017), and a low of 18.5$°C$ (Warsaw, VA 2020). Estimated reaction norms for the conditional yield effect of the *B1* awn suppressor (B). *B1* awn suppression was favorable or neutral in most environment, but had a substantial yield effect in heat-stressed environments.

G reflecting genetic distances between programs also interacted significantly with awn status(PC1, $p < 1.61E-03$; PC2, $p < 1.35E-04$; PC3, $p < 1.98E-07$; PC4, $p < 1.66E-04$).

Both awned *b1* and awnless *B1* lines were lower-yielding in environments with higher maximum temperatures during grain fill, but the decrease in yield was less severe for the awned lines than for awnless lines (Fig. 5.2). The lack of association between lower precipitation and improved yield for awn cultivars may suggest that direct cooling of the wheat head is the primary driver of improved yield for awned cultivars in heat-stressed conditions, but does not necessarily suggest that drought stress is unimportant physiologically for generating a yield effect of awns. Precipitation received during grain fill is only one of the drivers of drought stress experienced by plants during grain fill, as increasing temperatures may increase evapotranspiration and dry soil more rapidly. From the perspective of breeding for target environments, however, to what extent improved yield in heat-stress conditions

for awned cultivars results from heat stress per se or drought stress driven by increasing temperatures is less important than knowing that awned lines out-perform awnless in those conditions. Increasing performance of *B1* awnless lines in cooler, higher-yielding environments may relate to the increase grain number associated with awn suppression. In this case, while *B1* awn suppression may lower yields in poor environments, it may allow plants to take advantage of good growing conditions with less source limitations.

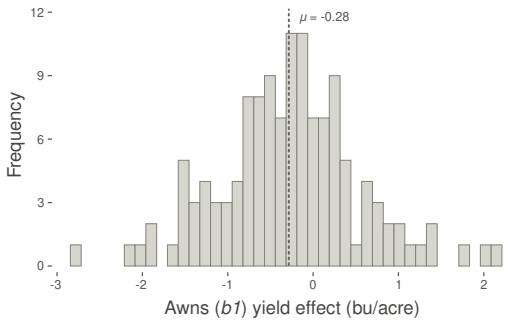

Figure 5.3: **Distribution of estimated per-site *b1* yield effects.** The mean estimated yield effect of awns is slightly negative in sampled environments, but environmental conditions associated strong positive and negative yield effects are also observed.

5.4.3 Overall effect of awns on yield varies by environment, year, and germplasm

Historic data on testing environments within each state were used to estimate the yield effect of awns in each site year. Reaction norms estimated in the mixed model were combined with the heading-date based grain fill mean daily maximum temperatures to estimate environment-specific yield effects (Fig. 5.3). While the average estimated effect of awns was slightly negative (-0.28 bu/acre), the effect within site years varied with temperatures from a decrease of 2.8 bu/acre (Warsaw, VA in 2020), to an increase of 2.1 bu/acre (Gainesville, FL 2017). Breeders generally select field sites representative of their target environments, but the proportions of warmer to cooler environments sampled by collaborators through the trials in this data set likely does not accurately represent the relative proportions of those

environments in real wheat acreage. Nevertheless, to get a sense of the within-state patterns of *b1* yield effects, effects on all field trials conducted within each state were averaged over years (Table 5.2). Overall positive effects of awns on yield were observed in only site in Florida and Texas. At the same time, the majority of states with an overall negative yield effect of awns had years with a positive awns effect more than a quarter of the time. While the states with the overall largest advantage for *b1* and the largest advantage for *B1* were the most southern (FL) and most northern (MD) states respectively, awn effects did not scale exactly with the overall climate of states. Despite being the state with the overall coldest climate, Kentucky had the third-highest yield effect of *b1*, and an overall positive effect of *b1* in 40% of years (Table 5.2). This likely results from the observation that Kentucky heading dates were on average the latest of any state (128 days), pushing grain fill into late May when temperatures rise. Conversely, South Carolina had one of the largest overall negative yield effects of awns, potentially relating to its average early heading date (97.8). Because heading date is itself under the control of environmental conditions, the effect of awns on yield may not reflect overall climactic differences between regions. At the same time, effect of awns on yield will vary within a population, reflecting differences in plants' flowering habits.

State	μ_{HD}	μ_{Yield}	μ_{GFMaxC}	*b1* Effect	% Positive Trials
FL	95.4	56.66	26.98	0.95	77
TX	99.5	60.36	24.98	0.07	60
KY	128.1	76.92	24.52	-0.13	40
LA	92.9	56.94	24.42	-0.18	20
GA	96.4	73.83	24.19	-0.28	25
AR	101.7	58.57	23.94	-0.39	08
NC	104.6	66.80	23.55	-0.56	33
SC	97.8	81.13	23.30	-0.67	33
VA	118.8	77.05	23.24	-0.70	30
MD	125.9	81.04	22.66	-0.95	00

Table 5.2: **Mean observations and estimated effects**

Breeders may consider the effect of awns on both overall yield and on year-to-year variability in yield. While in a majority of sampled environments the *b1* allele for awn presence was associated with negative yield, yields in these environments were overall higher than environments where *b1* had a positive effect. If awnless cultivars have greater

yield potential in favorable conditions, but suffer in stressed environments, then their overall variability will be higher. In environments where the yield effect of awns is sometimes negative and sometimes positive, growers may prefer awned cultivars with a higher yield "floor", even if they have lower overall yield potential.

Variability in the yield of *B1* versus *b1* lines within each state was estimated using the site year estimates of awn effects (Table 5.3). While Virginia was the state with the second lowest overall yield advantage for awned cultivars, high variability in grain fill temperatures resulted in a much higher estimated yield variability for awned lines compared to awnless. This contrasts with Arkansas, where *B1* awnless lines were overall higher-yielding and were only marginally more variable. The year to year heat stress variability underlying these differences in estimates of yield variability is difficult to estimate. Variation in heat stress during grain fill will result from both variation in weather conditions during the winter and spring that influence the heading dates of wheat lines, and then variation in weather conditions during grain fill itself. As climate change increases both overall temperatures and temperature variability, it may be valuable to select for traits that help produce consistent yields under a range of conditions.

State	σ^2_{HD}	σ^2_{Heat}	$\sigma^2_{\mu_{B1Yield}}$	$\sigma^2_{\mu_{b1Yield}}$	Diff. σ^2
VA	77.5	5.06	30.3	20.3	10.0
FL	60.3	3.32	19.8	13.3	6.5
NC	57.7	2.89	17.3	11.6	5.7
MD	130.7	2.92	17.4	11.7	5.7
TX	147.2	2.74	16.4	11.0	5.4
SC	72.5	2.53	15.1	10.2	4.9
KY	17.2	1.64	9.8	6.6	3.2
LA	50.9	1.58	9.5	6.4	3.1
GA	36.0	0.96	5.7	3.8	1.9
AR	147.6	0.73	4.4	2.9	1.5

Table 5.3: **Variance in observations and estimated effects.** States ranked according to the benefits of developing awned cultivars in terms of reduction of year to year yield variance.

5.5 Conclusions

The striking visual difference between awned and awnless cultivars has made the *B1* gene one of the best-studied in crop genetics. Conflicting results on the importance of this gene in contributing to yield variation, potentially in an environment-specific manner, has been limited by geneticts' ability to sample a sufficient number of diverse environments and genotypes to make real inferences on its yield effect. Here, we present the largest study yet on this gene, and find that its effect on yield varies in relation to heat stress during grain fill. We estimate that while awns have a slight negative effect across years for all states in our data set except Texas and Florida, on a year-to-year basis the effect of awns may be positive or negative, depending on heat stress during grain fill. Overall estimated effects suggest that *B1* awn suppression provides a slight yield improvement in high-yielding environments, but will lower yields relative to awned lines in lower-yielding, heat stressed environments. Breeders targeting environments where wheat plants experience heat stress during grain fill should select awned b1 cultivars, while those targeting environments that rarely experience grain fill heat stress should select awnless *B1* cultivars. At the same time, breeders targeting environments where awns have a small average negative yield effect may still consider selecting awned cultivars for their higher yield "floor" and lower year-to-year yield variability. As both overall temperature and year-to-year variability in heat stress increase, breeders in these marginal environments may also consider preferentially selecting awned lines for future climates.

5.6 Competing interests

The authors declare that they have no competing interests.

5.7 Author's contributions

ND prepared genotype data, performed all analyses, and wrote the initial draft of the manuscript. ND, MG, JPM, and GBG conceived of the analyses. ND, MG, JPM, and GBG edited the manuscript. JL and BPW organized and prepared the phenotypic data. JPM, DM,

NS, CG, RB, MM, SH, MAB, RM, AI, DVS, and VT planted and managed yield trials, collected phenotypes, and performed initial within-site analyses.

5.8 Acknowledgements

The authors thank the staff of the USDA-ARS Eastern Regional Small Grains Genotyping lab, particularly Kim Howell and Jared Smith, for DNA extraction and library preparation for the genotype data used in this study. Thanks to Zachary Winn for feedback on some analyses early in the drafting of the manuscript. The authors also thank staff of cooperating breeding programs for planting and phenotyping of cooperative yield trials. Weather data was obtained from the NASA Langley Research Center (LaRC) POWER Project funded through the NASA Earth Science/Applied Science Program. Support was provided by the Agriculture and Food Research Initiative Competitive Grant 67007-25939 (WheatCAP-IWYP) from the USDA NIFA.

References

Aggarwal, P. and Sinha, S. (1983). Water stress and water-use efficiency in field grown wheat: A comparison of its efficiency with that of C4 plants. *Agricultural Meteorology*, 29(3):159–167.

Amjad Ali, M., Ali, Z., Zulkiffal, M., and Zeeshan AARI, M. (2010). Source-Sink Relationship between Photosynthetic Organs and Grain Yield Attributes during Grain Filling Stage in Spring Wheat (Triticum aestivum). Technical report.

Bhullar, S. and Jenner, C. (1985). Differential Responses to High Temperatures of Starch and Nitrogen Accumulation in the Grain of Four Cultivars of Wheat. *Functional Plant Biology*, 12(4):363.

Blum, A. (1986). The Effect of Heat Stress on Wheat Leaf and Ear Photosynthesis. Technical Report 174.

Börner, A., Schäfer, M., Schmidt, A., Grau, M., and Vorwald, J. (2005). Associations between geographical origin and morphological characters in bread wheat (Triticum aestivum L.) . *Plant Genetic Resources*, 3(3):360–372.

Browning, B. and Browning, S. (2016). Genotype Imputation with Millions of Reference Samples. *The American Journal of Human Genetics*, 98(1):116–126.

Butler, D. G., Cullis, B. R., Gilmour, A. R., Gogel, B. J., and Thompson, R. (2017). ASReml-R Reference Manual Version 4.

DeWitt, N., Guedira, M., Lauer, E., Sarinelli, M., Tyagi, P., Fu, D., Hao, Q., Murphy, J. P., Marshall, D., Akhunova, A., Jordan, K., Akhunov, E., and Brown-Guedira, G. (2020). Sequence-based mapping identifies a candidate transcription repressor underlying awn suppression at the B1 locus in wheat. *New Phytologist*, 225(1):326–339.

Evans, L. T., Bingham, J., Jackson, P., and Sutherland, J. (1972). Effect of awns and drought on the supply of photosynthate and its distribution within wheat ears. *Annals of Applied Biology*, 70:67–76.

FAO (2020). *Crop Prospects and Food Situation #1, March 2020*. FAO.

Ferguson, H. (1975). Use of Variety Isogenes in Plant Water-Use Efficiency Studies. pages 25–29.

Foulkes, M. J., Sylvester-Bradley, R., Weightman, R., and Snape, J. W. (2007). Identifying physiological traits associated with improved drought resistance in winter wheat. *Field Crops Research*, 103(1):11–24.

Gauch, H. G. (1937). *A Physiological Study of the Awns of Red Winter Wheat*. PhD thesis, Kansas State University.

Glaubitz, J. C., Casstevens, T. M., Lu, F., Harriman, J., Elshire, R. J., Sun, Q., and Buckler, E. S. (2014). TASSEL-GBS: A High Capacity Genotyping by Sequencing Analysis Pipeline. *PLoS ONE*, 9(2):e90346.

Grundbacher, F. J. (1963). The Physiological Function of the Cereal Awn. pages 366–381.

Guedira, M. and Paulsen, G. M. (2002). Accumulation of starch in wheat grain under different shoot/root temperatures during maturation. *Functional Plant Biology*, 29(4):495.

Hawker, J. and Jenner, C. (1993). High Temperature Affects the Activity of Enzymes in the Committed Pathway of Starch Synthesis in Developing Wheat Endosperm. *Functional Plant Biology*, 20(2):197.

Hickman, J. F. (1889). Experiments in wheat seeding; Comparative test of varieties of wheat. Technical report, Ohio Agricultural Experiment Station.

Huang, D., Zheng, Q., Melchkart, T., Bekkaoui, Y., Konkin, D. J. F., Kagale, S., Martucci, M., You, F. M., Clarke, M., Adamski, N. M., Chinoy, C., Steed, A., McCartney, C. A., Cutler, A. J., Nicholson, P., and Feurtado, J. A. (2020). Dominant inhibition of awn development by a putative zinc-finger transcriptional repressor expressed at the B1 locus in wheat. *New Phytologist*, 225(1):340–355.

Kato, K., Miura, H., Akiyama, M., Kuroshima, M., and Sawada, S. (1998). RFLP mapping of the three major genes, Vrn1, Q and B1, on the long arm of chromosome 5A of wheat. Technical report.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16):2078–2079.

Li, X., Wang, H., Li, H., Zhang, L., Teng, N., Lin, Q., Wang, J., Kuang, T., Li, Z., Li, B., Zhang, A., and Lin, J. (2006). Awns play a dominant role in carbohydrate production during the grain-filling stages in wheat (Triticum aestivum). *Physiologia Plantarum*, 127(4):701–709.

Martin, J. N., Carver, B. F., Hunger, R. M., and Cox, T. S. (2003). Contributions of Leaf Rust Resistance and Awns to Agronomic and Grain Quality Performance in Winter Wheat. *Crop Science*, 43:1712–1717.

Maydup, M. L., Antonietta, M., Graciano, C., Guiamet, J. J., and Tambussi, E. A. (2014). The contribution of the awns of bread wheat (Triticum aestivum L.) to grain filling: Responses to water deficit and the effects of awns on ear temperature and hydraulic conductance. *Field Crops Research*, 167:102–111.

Motzo, R. and Giunta, F. (2002). Awnedness affects grain yield and kernel weight in near-isogenic lines of durum wheat. *Australian Journal of Agricultural Research*, 53(12):1285–1293.

Poland, J., Endelman, J., Dawson, J., Rutkoski, J., Wu, S., Manes, Y., Dreisigacker, S., Crossa, J., Sánchez-Villeda, H., Sorrells, M., and Jannink, J. L. (2012). Genomic selection in wheat breeding using genotyping-by-sequencing. *Plant Genome*, 5(3):103–113.

R Core Team (2021). R: A Language and Environment for Statistical Computing.

Rebetzke, G. J., Bonnett, D. G., and Reynolds, M. P. (2016). Awns reduce grain number to increase grain size and harvestable yield in irrigated and rainfed spring wheat. *Journal of Experimental Botany*, 67(9):2573–2586.

Sanchez-Bragado, R., Kim, J. W., Rivera-Amado, C., Molero, G., Araus, J. L., Savin, R., and Slafer, G. A. (2020). Are awns truly relevant for wheat yields? A study of performance of awned/awnless isogenic lines and their response to source–sink manipulations. *Field Crops Research*, 254.

Sarinelli, J. M., Murphy, J. P., Tyagi, P., Holland, J. B., Johnson, J. W., Mergoum, M., Mason, R. E., Babar, A., Harrison, S., Sutton, R., Griffey, C. A., and Brown-Guedira, G. (2019). Training population selection and use of fixed effects to optimize genomic predictions in a historical USA winter wheat panel. *Theoretical and Applied Genetics*, 132(4):1247–1261.

Sparks, A. (2018). nasapower: A NASA POWER Global Meteorology, Surface Solar Energy and Climatology Data Client for R. *Journal of Open Source Software*, 3(30):1035.

Tambussi, E. A., Nogués, S., and Araus, J. L. (2005). Ear of durum wheat under water stress: Water relations and photosynthetic metabolism. *Planta*, 221(3):446–458.

VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. *Journal of Dairy Science*, 91(11):4414–4423.

Vervelde, G. J. (1953). The agricultural value of awns in cereals. *Netherlands Journal of Agricultural Science*, 1(1):2–10.

Watkins, A. E. and Ellerton, S. (1940). Variation and genetics of the awn in Triticum. *Journal of Genetics*, 40(1-2):243–270.

Weyhrich, R. A., Carver, B. F., and Martin, B. C. (1995). Photosynthesis and Water-Use Efficiency of Awned and Awnletted Near-Isogenic Lines of Hard Winter Wheat. *Crop Science*, 35(1):172–176.

Weyhrich, R. A., Carver, B. F., and Smith, E. L. (1994). Effects of awn suppression on grain yield and agronomic traits in hard red winter wheat. *Crop Science*, 34(4):965–969.