THE GENETIC BASIS OF GRAIN TRAITS IN SORGHUM

A Dissertation

by

ASHLEY JEAN MATTISON

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	John E. Mullet
Committee Members,	William L. Rooney
	Dorothy E. Shippen
	James C. Womack
Head of Department,	Dorothy E. Shippen

August 2017

Major Subject: Genetics

ABSTRACT

Crop production must increase by 70% by 2050 in order to meet the expected requirements for population growth. While previous research has resulted in genetic improvement of the grain output in major cereal crops such as *Zea mays* and *Oryza sativa*, much less investment has been made to improve the grain yield of *Sorghum bicolor,* a drought-tolerant grain and forage crop native to Africa*.* Multiple genetic and environmental factors influence grain yield making it challenging to identify the pathways determining the amount of grain produced. In this dissertation, variation in grain yield, grain weight and grain number per plant, panicle architecture traits, and grain composition were measured in three different RIL populations. Several quantitative trait loci (QTL) for grain number and weight that act independently to modify grain yield were identified. QTL for for grain composition, stem hollowing, and nutrient remobilization were identified, a subset of which aligned with QTL for grain weight. A QTL on LG01 was identified for both grain weight and grain number in the BTx642 x Tx7000 recombinant inbred lines (RIL) population; however, the tradeoff between grain weight and grain number resulted in no increase in grain yield. In the BTx623 x IS3620c population, a QTL was identified in the same region on LG01 for grain yield and grain number. These results suggest that two tightly linked genes influencing grain weight and grain number are located in this region of LG01.

Sorghum is a photoperiod-sensitive short-day plant with critical photoperiods ranging from 11-14 hours. When grain sorghum was introduced into more temperate climates, breeders selected for earlier flowering times to optimize grain yield. Six maturity loci, *Ma*$_1$-*Ma*$_6$, were identified that modify flowering time. When recessive, *Ma*$_1$ – *Ma*$_6$ result in an earlier flowering phenotype in long days. Four of the six maturity genes, *Ma*$_1$, *Ma*$_3$, *Ma*$_5$ and *Ma*$_6$, have previously been map-based cloned and identified. In this study, *Ma*$_2$ was fine mapped, sequenced, and identified as Sobic.002G302700, a gene encoding a zinc finger transcription factor. The identified SNP in 80M creates a stop codon causing a loss of function of *Ma*$_2$.

DEDICATION

I dedicate this dissertation to my daughter Kaylee, who inspires me to reach for the stars so I can show her that anything is possible.

ACKNOWLEDGEMENTS

I would like to thank my committee chair, Dr. Mullet, and my committee members, Dr. Rooney, Dr. Shippen, and Dr. Womack, for their advice and support during graduate school.

Thanks also goes to my friends, colleagues and the department faculty and staff for making my time at Texas A&M University a great experience.

Finally, thanks to my husband Scott for his patience and encouragement.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by a dissertation committee consisting of Professor John Mullet and Professor Dorothy Shippen of the Department of Biochemistry and Biophysics and Professor William Rooney of the Department of Soil and Crop Science and Professor James Womack of the Department of Veterinary Medicine.

Field studies were planted and maintained by the Rooney lab that were used in research described in Chapter II and Chapter III. Student workers and Susan Hall aided in phenotyping of traits described in Chapter II. The *Ma_2* QTL was identified by a previous graduate student, Dr. Sara Olson. *Ma_2* gene sequence analysis was aided by graduate student Manish Thakran. MQM penalty calculations were performed by Sandra Truong for the grain traits in Chapters II and III and by Anna Casto for flowering time in Chapter IV.

All other work conducted for the dissertation was completed independently by Ashley Mattison.

Funding Sources

The graduate study of grain traits was supported by the Adkisson Chair in Agricultural Biology. Partial funding for research leading to the identification of *Ma_2* identification was provided by grants from the Department of Energy.

NOMENCLATURE

RIL	Recombinant Inbred Line
QTL	Quantitative Trait Locus
CIM	Composite Interval Mapping
MQM	Multi-QTL Mapping
DG	Digital Genotyping

TABLE OF CONTENTS

Page

Page

Page

LIST OF FIGURES

Page

Page

LIST OF TABLES

Page

Page

CHAPTER I

INTRODUCTION AND LITERATURE REVIEW

Sorghum Background

Sorghum bicolor is the 5th most economically important cereal crop worldwide and has uses in human consumption, livestock feed, and biofuels (1–5). In the United States, sorghum is used primarily as livestock feed; however, in many areas in the world, sorghum is used as a primary food source for humans (6, 7). *Sorghum bicolor* subsp. bicolor contains all of the cultivated sorghums. There are 5 major races of *Sorghum bicolor* (L.) Moench: bicolor, caudatum, durra, guinea, and kafir (7). Each race of sorghum is adapted to a different region of Africa. For example, guinea is from West Africa and has been selected to grow in regions of higher rainfall compared to sorghum of the race durra.

In their 2050 mandate, the Food and Agriculture Organization prioritized improvement in crop grain yield (8, 9). This mandate came as a response to reports from the United Nations Department of Economic and Social Affairs indicating that the world's population will exceed 9.1 billion people by the year 2050 (10). To meet the demand for crops required to feed a population of 9.1 billion, it is estimated that crop production must increase by 60 – 110% (8, 11). This estimate includes both crop products directly consumed by humans and crops used for meat production. Crop production must increase from 2.1 billion tonnes to 3 billion tonnes (2.1 billion to 3 billion Mg) during this time period (12). Additionally, this estimate only considers crop production for grain and forage

and does not include the expected increase in production of crops used for bioenergy (2, 3, 8). As countries become more developed, urbanization increases further increasing the need for improved grain yield. Typically, as the standard of living increases the demand for protein in human diets increases (13). The expected increases in the consumption of protein in developing countries will significantly impact the demand for forage and feed. Unfortunately, the rate of improvement in the grain yield of maize, rice and wheat, the major sources of calories for humans, has been declining and is insufficient to meet the projected demand (14).

Sorghum has been under much less intensive selection than other grain crops (15). Due to its relatively small, sequenced genome of 730 Mbp, sorghum is an excellent system for genetic studies (15, 16). Additionally, sorghum is a diploid, largely inbreeding species with only 10 chromosomes making it ideal for genetic studies. Therefore, research on sorghum could enhance its grain yield, contributing to the 2050 mandate to increase agricultural productivity in order to feed the world.

Sorghum Evolution

Dicots, such as Arabidopsis, first developed on Earth between 120 and 150 million years ago (17). When an asteroid impacted the Yucatán Peninsula roughly 66 million years ago, the environmental impact provided an opportunity for grasses to proliferate (18). *Sorghum bicolor*, a C4 grass, and rice, a C3 grass, diverged from a common ancestor about 50 million years ago. Twenty-

five to thirty million years ago, *Sorghum bicolor* developed the capacity for C4 photosynthesis. Due to its more efficient use of CO_2, C4 photosynthesis is hypothesized to have developed in response to declining atmospheric CO_2 levels and improved heat, drought and salinity tolerance (19). There have been several attempts to engineer C4 photosynthesis into C3 plants; however, there has been limited success (20, 21). Despite the limited success of these experiments, there have been reports of these attempts increasing both photosynthetic capacity and grain yield (22).

Grain Yield Improvement History

Grain yield is a complex trait that is influenced by many different genes and genetic pathways (23). Due to multiple genetic and environmental influences, it has been challenging to identify all of the pathways involved in determining the amount of grain produced. Grain yield was increased significantly with hybrid crop development, enabled by the discovery of cytoplasmic male sterility (24). Cytoplasmic male sterility causes pollen infertility making the plant an obligate out-crosser. Aided by the discovery of cytoplasmic male sterility in *Sorghum bicolor*, the breeding of hybrids led to an increase in grain yield (25). This increase in yield was due to the phenomenon of hybrid vigor, which occurs when the heterozygote offspring outperforms either inbred parent. This concept had previously been exploited in many different crops to increase grain yield (26, 27). There are several different theories to explain why hybrid vigor leads to these distinct phenotypes. The dominance hypothesis

postulates that both parents have different, slightly deleterious alleles in their genome; therefore, the complementation of these alleles creates a hybrid with the expression of favorable alleles at these loci. An alternative theory is the over-dominance hypothesis (28). The over-dominance hypothesis states that the heterozygous genotype has phenotypic superiority to either homozygous genotype (29). The over-dominance hypothesis has been observed at individual loci; however, both hypotheses hold merit and it is widely accepted that both contribute to hybrid vigor (30). While hybrid vigor has led to a sizable increase in grain yields, an understanding of the specific genetic/trait basis of increases in grain yield per plant has advanced more slowly (31).

Grain yield is the product of grain number and grain weight. Grain yield can be analyzed on the basis of yield per panicle, plant, or m^2 of land. Grain yield can be increased by higher availability of nutrients that are required for growth or grain filling. In fact, Sinclair and Jamieson (2006) discuss the possibility that grain yield is controlled by nitrogen availability and uptake. They noted that grain number increased when environmental conditions were favorable for nitrogen uptake which suggested that nitrogen is the main determinant of grain yield and grain number was a consequence of environmental conditions (32). Fischer (2008) offered a rebuttal to this concept based on studies of nitrogen availability and its impact on grain number. He concluded that grain number is an important aspect of grain yield and is not just a consequence of nitrogen availability. Fischer hypothesized that the best way to

increase grain yield is by studying the physiological mechanisms involved in grain yield pathways (33). He thought that an understanding of the physiological mechanisms that contribute to grain yield could help identify bottlenecks in these pathways can be relieved, thereby increasing yield.

Sorghum Development

To understand the genetic/biochemical pathways that affect grain yield, it is helpful to first characterize the developmental phases and phenotypes that lead to grain production. There are four main stages of *Sorghum bicolor* development; juvenile, vegetative, floral initiation to anthesis (booting), and grain filling (34). These stages are outlined in Figure 1.

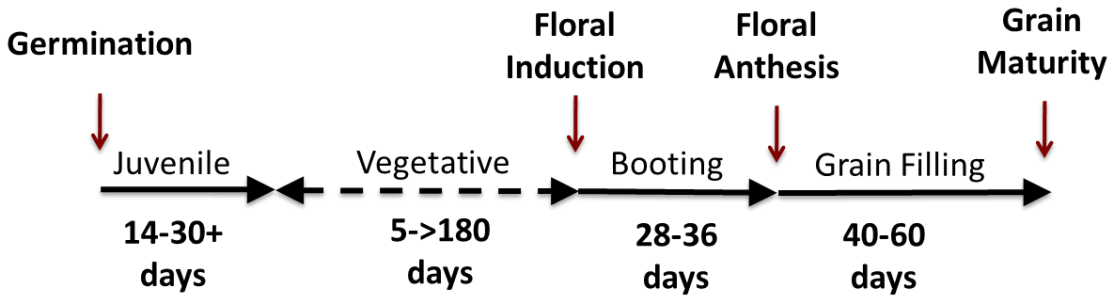

Figure 1. The four main developmental stages of *Sorghum bicolor*. Key transitions and typical length of each stage are labeled.

The first stage of *Sorghum bicolor* development is the juvenile stage. The juvenile stage begins at germination and continues for ~14-28 days depending on genotype and environmental conditions until ~5-7 juvenile leaves are fully expanded. *Sorghum bicolor* transitions to adult phase vegetative growth

following exit from the juvenile phase. Genetic reprogramming has been reported to occur at the end of the juvenile stage, suggesting that the events of the juvenile stage may only have indirect effects on grain yield (35). In the vegetative stage, the length and diameter of the stem increase, larger leaves develop and the leaf canopy is created (34). The size and angle of the leaves that develop in the leaf canopy and plant density determines how much light is intercepted by the plant (36). The duration of the vegetative stage in *Sorghum bicolor* is controlled by maturity loci that were identified in part by Quinby and others (37–39). When *Sorghum bicolor* grain production was extended to temperate climates, breeders selected earlier flowering genotypes that were better adapted to these regions (37). These early flowering genotypes were due to mutations primarily in genes that conferred photoperiod sensitivity that delay flowering in long days. These genetic loci were named maturity loci, Ma_1-Ma_6 (37–39).

The reproductive stage begins at floral initiation. The interval from floral initiation to anthesis is termed the booting stage. This phase is termed the booting stage due to the panicle being “in boot” during the final days prior to emergence. The booting stage contributes significantly to grain yield as it plays a key role in determining panicle size and grain number (40). During this stage, the panicle architecture is determined, setting the maximum number of seed possible. *Sorghum bicolor* has a significantly more branched panicle than other cereal crops (41). A high degree of branching generates increased variation in

panicle architecture. As the maximum number of seed is set during the booting stage (34), events that occur following the onset of floral initiation through anthesis would be useful to examine and assess for their potential impact on grain yield. I hypothesize that panicle architecture is a main genetic determinant of grain number per panicle because panicle architecture dictates the number of sites on the panicle where grain can be formed. Another important determinant of grain number is floret termination (42). After panicle architecture is determined, florets on the panicle are then terminated to prevent more seed from being produced than the plant can support. During the booting stage, an increase in floret termination has been linked with nutrient limitations (43, 44). Another factor that can affect grain number is the number of grain produced at each potential seed site. Sorghum is known for the phenomenon of twin seeding, which occurs when two grains are produced per site (45). Sorghum can produce up to 3 seeds per site, but the plant does not do this under normal conditions. Unlike the wild type plant which produces one or two seeds per site, the mutant *msd-1* produces three seed independent of nutrient availability (45, 46). Despite the dramatic increase in grain number, the overall grain yield of the mutant *msd-1* line is not increased due to a reduction in weight per grain. This is an example of the tradeoff between grain number and weight that creates a challenge when attempting to increase grain yield. Analysis of the genetic basis of the variation in grain yield and its subcomponents through QTL mapping may help identify genetic determinants that are useful for increasing yield.

The grain filling stage begins after fertilization at anthesis. Grain filling affects grain weight and composition. While grain number has previously been given more attention as a factor that modulates grain yield (31, 32, 40, 47, 48), grain weight can be equally important in some circumstances (49). Grain number and grain weight are negatively correlated, supporting the relationship seen in tradeoff scenarios. The rate of grain filling greatly affects the weight of the grain. Nutrient allocation depends on source activity and sink strength. During seed development, the plant is the source of nutrients (sugars, amino acids, etc.) whereas the developing seed is a sink for these compounds. Once the grain filling stage starts, the “sink strength” of the panicle is given priority over the other sink-related activities of the plant (49–51).Therefore, grain filling is often associated with leaf senescence and stem hollowing resulting from mobilization of carbon and nitrogen from vegetative tissues to the grain (49). If carbon or nitrogen is limiting, an increase in grain number will lead to a decrease in grain weight due to an inadequate supply of nutrients. Figure 2 shows nitrogen and carbon resources and how they are impacting the booting and grain filling stages.

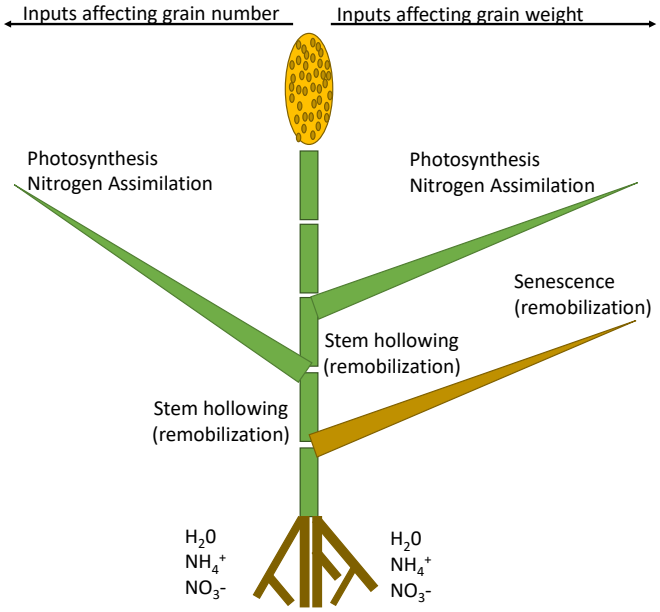

Figure 2. Sources of nitrogen, carbon and water during the booting and grain filling stage. Photosynthesis, nitrogen assimilation and root uptake of nutrients are ongoing processes that will increase the available amount of nutrients. Remobilization starts at the onset of floral induction but increases in the rate of remobilization during the grain filling stage. Remobilization occurs due to the sink strength of the panicle.

In addition to optimizing the genetic regulation of grain number and grain weight, it has been suggested that an increase in the duration of the booting stage or increases in the duration of the grain filling stage could potentially lead to an increase in grain yield (52). Moreover, increasing the duration of grain filling would potentially increase grain weight without affecting the rate of filling. An increased amount of time for photosynthesis would increase carbon availability for the seed development (53). Understanding the mechanisms that control the duration of these stages could enable the identification of a mechanism to alleviate the negative correlation that exists between grain

number and grain weight allowing both factors to be increased in parallel (54–56). Independent control of the duration of these phases of development could enable grain number and grain weight to become independent variables in the grain yield equation, potentially resulting in increased grain yield.

QTL Mapping

Genetic loci controlling phenotypes associated with grain yield and traits that modulate grain yield can be identified via quantitative trait locus (QTL) mapping. QTL mapping can be used to identify the location and nature of genes/alleles that modulate grain yield components by correlating the segregation of phenotypic traits with genetic markers in the genome (57). Favorable alleles of main effect QTL can be selected for in breeding programs to increase yield. QTL also enable the identification of pathways and genes that modulate grain yield.

In order to create QTL maps, phenotypic data on grain yield traits and genotype information must be collected (58). In this dissertation, only bi-parental QTL mapping is used. Other forms of QTL mapping that do not require a set family structure include genome wide association studies (GWAS) (57). In order to create a bi-parental family structure, two parents are selected that differ in alleles that modulate the trait of interest. These parental lines are crossed to create an F1 plant that is subsequently selfed to generate an F2 population. These F2 plants and progeny from additional generations are selfed to reduce heterozygosity in the derived progeny, thereby generating a recombinant inbred

line (RIL) population. RILs contain different genetic combinations of the two parental genomes through the process of random association and recombination. Recombination occurs when homologous parental chromosomes exchange genetic material through the process of crossing over during meiosis. The new allelic combinations in the RILs can provide insight into how phenotypes are altered by genotypic background through QTL mapping. Populations comprised of higher numbers of RILs provide more statistical power in QTL mapping and have the potential to fine map regions containing alleles underlying QTL to a greater extent. By definition, RILs have low amounts of heterozygosity as lines are not typically classified as RILs until the F7 or F8 generation.

The populations used in this research were genotyped using Digital Genotyping (DG) (59). DG identifies single nucleotide polymorphisms (SNPs) that distinguish parental lines. DNA is extracted from leaf tissue and digested with restriction enzymes that recognize and cut specific sequences in the genome. For example, FseI recognizes and cuts non-methylated DNA having the 8 base sequence GGCCGGvCC (59). Digestion of the *Sorghum bicolor* genome with FseI results in 23,000 DNA fragments containing unique sequences in the genome that can be subjected to sequencing. NgoMIV is a restriction enzyme that recognizes the 6-base sequence GvCCGGC resulting in approximately 120,000 digestion fragments containing unique sequences useful for genotyping from the *Sorghum bicolor* genome. A DNA template is generated

from the digestion fragments and 100 base sequences are obtained adjacent to sites of digestion using Illumina sequencing platforms. DG sequences derived from both parental lines of a mapping population that are unique in the *Sorghum bicolor* genome sequence are compared and scanned for polymorphisms that distinguish the two parents. For the FseI depth, around 900-1300 markers are typically identified in bi-parental crosses used for QTL mapping and up to 10,000 when using NgoMIV analysis.

Mapmaker (60) was used to convert the collected DG data into genetic maps for the analyzed populations. QTL analysis was performed using QTL Cartographer (58). The maps, genotypic data, and phenotypic data from each population analyzed were loaded into QTL Cartographer and used to identify QTL for each phenotype of interest. Composite interval mapping was utilized to search for correlations between markers and phenotypes. In some instances, single-marker mapping can be used in place of CIM. In single marker QTL mapping, each marker is evaluated independently with the phenotype to see if there is a correlation of that marker and the phenotypic trait. An advantage of this method is the speed at which this mapping code can run. However, this speed comes at the cost of increased sensitivity provided by CIM, and interval mapping (61). Interval mapping considers two markers and the region between them when identifying QTL. As it is statistically unlikely that a SNP used for genotyping will be the cause of phenotypic change that is observed, considering

two markers increases the sensitivity of interval mapping when identifying QTL located between two markers.

Interval mapping uses all of the RIL phenotypes at two markers to estimate the location of a causative SNP between two markers through the different recombination sites of the RILs between the two markers. Interval mapping utilizes a sliding window of two markers during each calculation; therefore, markers not at the ends of chromosomes will be considered twice.

CIM is a useful mapping method because it adds sensitivity to interval mapping through the consideration of background markers. Considering markers that are not within the window of analysis enables a greater understanding of genetic background. By considering background markers, CIM is capable of identifying QTL may are only present under certain genetic conditions. CIM enabled the development of multi-QTL mapping (MQM). MQM is capable of identifying more QTL by taking into account the genetic background. Additionally, MQM can use the information from CIM to identify epistatic interactions amongst QTL. However, the increased information and sensitivity comes at the cost of increased running time and computation power. It takes days of run time on a super computer to complete the required penalty calculations for each phenotypic trait being processed via MQM; whereas, other methods take minutes.

To identify QTL, a threshold for a significant relationship between genotypic data and phenotypic data is determined via a permutation test. For

CIM, a p-value of 0.05 at 1000 permutations was used for all analyses. This permutation test utilized all phenotypes and genotypes that were entered into the program and randomly assign them to each other. This provides a base line of what can be considered by pure chance correlations. Any peaks over this base line are considered to be true correlations between the genotypic and phenotypic data. This test eliminates many false positives and provides a 95% confidence that the QTL over the threshold line are true QTLs.

R/qtl was used as an alternative to mapmaker for map creation of the BTx642 x Tx7000 population as well as the MQM data. MQM enables further analysis of the data collected from CIM (61). MQM's higher statistical power allows it to identify ghost QTL, additional QTL that CIM did have the statistical power to identify, and epistatic interactions. MQM requires 10,000 permutations rather than 1,000 like traditional QTL mapping and is plotted along a probability of the logarithm of the odds (pLOD) instead of a logarithm of the odds (LOD) score. The initial QTL map can identify additional QTL not found by traditional QTL mapping as the additional permutations allow QTL that were under the threshold line in CIM to become significant. The different allelic combinations at each QTL are analyzed to identify interactions between QTL. These tools are extremely important if QTL mapping methods are to be applied to pathway identification as they enable the identification of epistatic interactions. Understanding how different QTL interact with each other for the phenotype of interest help to determine their location in the genetic pathway. The modeling

aspect of MQM is also an additional tool that can be very beneficial by providing an estimate of the amount of variation that is being explained by the given model. This provides insight into how many other QTL with minor effects may or may not still be left to be discovered for this particular phenotype and population.

Flowering Time

When grain sorghum was introduced into more temperate climates, breeders selected for earlier flowering times to optimize grain yield in the new environment (38, 39, 62). In the U.S., this led to the identification of six maturity loci, *Ma*$_1$-*Ma*$_6$, which when recessive, resulted in an earlier flowering phenotype in long days. In optimal conditions, overall grain yield is positively correlated with longer flowering times (39). In temperate climates, grain maturation needs to occur before frost or onset of poor growing seasons and as a consequence, earlier flowering times were often better suited for agricultural purposes (38).

Earlier flowering times were introduced into sorghum lines following the identification of mutations in six different loci that control flowering time (*Ma*$_1$ – *Ma*$_6$). These maturity genes enabled breeders to select for the optimal days to flowering (DTF) in a given environment (39). By crossing different varieties of Milo, Quinby and Karper created 4 different phenotype classes of flowering: early, intermediate, late, and ultra-late. The DTF of these phenotype classes ranged from ~40 to 100 days. Genetic analysis showed that 3 maturity loci contributed to the variation in flowering time, *Ma*$_1$, *Ma*$_2$, and *Ma*$_3$ (38). The phenotypic impacts of *Ma*$_1$ – *Ma*$_3$ were observed in long, 14-hour photoperiod

day conditions and no phenotypic impacts were shown under short, 10-hour photoperiod day conditions. *Ma*$_1$ – *Ma*$_3$ were named in order of their phenotypic effect; with *Ma*$_1$ having the largest impact. Notably, the phenotypic effects of *Ma*$_2$ and *Ma*$_3$ were attenuated in *ma*$_1$ genetic backgrounds. Due to this epistatic interaction, it has been hypothesized that *Ma*$_2$ and *Ma*$_3$ were downstream of *Ma*$_1$ in the flowering time pathway.

Quinby later identified a 4th maturity locus (*Ma*$_4$) in Hegari *Ma*$_1$*Ma*$_2$*Ma*$_3$*ma*$_4$ (62). *Ma*$_4$ was found by crossing Hegari with different Milo lines, including 60M which is *ma*$_1$*ma*$_2$*ma*$_3$*Ma*$_4$. The flowering time distribution in this cross showed segregation of 4 distinct maturity genes. *Ma*$_1$ – *Ma*$_4$ are widely considered the four classic maturity genes in sorghum; however breeders continued to see variation in DTF that could not be explained by only *Ma*$_1$ – *Ma*$_4$ (39). For example, *Ma*$_5$ and *Ma*$_6$ were identified in populations that showed even higher photoperiod sensitivity and delayed flowering in long days (>12.4 hours) (39).

There has been considerable effort to identify the causative genes/alleles that correspond to the 6 maturity loci. *Ma*$_3$ was identified to be *PHYTOCHROME B* (*PHYB*) (63). The recessive *ma*$_3$ was reported as the result of a frameshift mutation resulting in a stop codon in 58M. The phenotypes displayed by the *PHYB* mutant sorghum included very early flowering, a reduction of chlorophyll content, and lack of sensitivity to red light (63). This result is similar to what is observed in *PHYB* mutants in Arabidopsis (64–66).

Ma_1 was identified as the *PSEUDO RESPONSE REGULATOR PROTEIN 37 (PRR37)* (67). Under long days, *SbPRR37* represses *CONSTANS* and EHD1 which encode activators of flowering time. The overall effect on flowering time caused by *SbPRR37* is influenced by both the circadian clock and photoperiod (67). *Ma_6* has been identified as *GRAIN NUMBER, PLANT HEIGHT AND HEADING DATE (Ghd7)* (67).*Ghd7* works with *PRR37* to repress floral initiation by repressing the expression of EHD1. *Ma_5* has been identified as *PHYTOCHROME C (PHYC)* through genetic analysis and sequence variants of PHYC found in R07007 (*ma_5*), 90M (*$ma_{5\text{-}1}$*), 100M (*$ma_{5\text{-}1}$*), and BTx623 (*Ma_5*) (68). The identification of 4 of the 6 sorghum maturity genes has enabled great progress towards an understanding the flowering time pathway of sorghum.

Flowering Time in Arabidopsis

Due to its status as a genetic model, the factors that contribute to floral initiation have been well studied in *Arabidopsis thaliana*. The biochemical signals from many distinct pathways influence the transition from vegetative growth to floral initiation (69–71). Several pathways converge to impact the overall flowering time pathway; however, when considering the core flowering time pathway, there are two dominant factors, photoperiod and the circadian clock (72, 73).

Developmental responses of plants are dependent upon the ratio and relative length of the light and dark photoperiod (73, 74).There are three typical classifications of photoperiod: long-day, short-day, and day neutral (75).

Arabidopsis is a long-day plant because it shows earlier flowering time under long-day conditions, whereas sorghum is a short-day plant because long days delay flowering and short days hasten flowering (75–77). Photoperiod responses are regulated by photoreceptors such as phytochromes that absorb red/far-red light, as well as cryptochromes, which are strong absorbers of the blue/ultraviolet region of the electromagnetic spectrum (77, 78). Photosensitive molecules play many roles within the organism, including regulating gene expression and stabilizing proteins. The regulation of biological responses mediated by interactions between the products of the photoperiod pathway and outputs of the circadian clock are often consistent with the external coincidence model. First proposed by Edwin Bunning in 1936, the external coincidence model proposes that a circadian rhythm of photoperiod sensitivity occurs within an organism and gene expression responds to these signals (79). The circadian rhythm influenced by the external photoperiod results in an internal oscillation of evening expressed genes, that are light sensitive, and morning expressed genes, that are light insensitive (80).

As the circadian clock defines an organism's response to environmental changes within a regular cycle, it is of great interest in many disciplines of biology (81–84). In plants, the circadian clock plays a role in the response to light, stomatal opening, and chloroplast movement in addition to many other regulatory processes (85–87). Initial studies of the circadian clock in plants can be traced back to 400 B.C. when Androsthenes recognized diurnal rhythms by

describing the leaf movements of the tamarind tree (88–90). In 1729, French astronomer de Mairan noted that daily leaf movements in a heliotrope plant persisted even in no photoperiod, he hypothesized that the movements indicated that circadian rhythms are of endogenous origin. De Mairan further hypothesized that these rhythms may be related to the sleep rhythms of bedridden humans (89–91). In 1832, de Candolle used *mimosa pudica* to accurately measure the period of these leaf movements to be 24 hrs. De Candolle further demonstrated that a 90 degree phase shift (12 hour time shift) could be introduced into these rhythms by offsetting the alternation of light and dark by 12 hours (89, 90). This suggested that the 24 hour rhythms of these leaf movements were endogenous and not simple responses to environmental time cues(89, 90).

Leaf movements being of an endogenous origin was widely disputed amongst scientists. In 1873, Pfeffer suggested that light leakage invalidated any claims that the leaf movement rhythms had endogenous origins. However, Pfeffer found that the free running periods of leaf movement rhythms differed from 24hrs, showing that these rhythms were endogenous and not driven by an undetected environmental cue (89, 90).

In the 1880s, Charles and Francis Darwin suggested that circadian rhythms were a heritable trait instead of an imprint of a 24 hour period caused by exposure to diurnal cycles during development (89, 90, 92). They tested this hypothesis by looking at seedlings that were exposed to light/dark cycles that differed from the normal 24 hour period. When these plants were released into

continuous conditions, the endogenous circadian period was restored. This showed that the inheritance of period length is genetic (89, 90).

The circadian clock has been well studied in plants, with Arabidopsis serving as a model for other higher plants (75, 76, 84). The core oscillator of the circadian clock is composed of three main genes: *TIMING OF CAB EXPRESSION 1* (*TOC1*), *CIRCADIAN CLOCK ASSOCIATED 1* (*CCA1*), and *LATE ELONGATED HYPOCOTYL* (*LHY*) (93). The core oscillator, depicted in Figure 3, influences flowering time. *TOC1* is an evening phase gene that promotes expression of morning phase transcription factors *CCA1* and *LHY*. *CCA1* and *LHY* in turn negatively regulate the expression of *TOC1* (94). This negative feedback loop is the basis of the circadian clock model and *TOC1*, *CCA1*, and *LHY* are the core genes required for circadian function. In Arabidopsis, the initial screen of clock mutants led to the discovery of *toc1*; indicating it is needed for expression in clock functions. The phenotype of the *toc1* mutant resulted in very short period lengths and loss of proper oscillation of gene expression within these periods (95). A similar phenotype, with arrhythmic expressions of genes regulated by the clock, presented itself within *lhy* and *cca1* mutants (96).

The rhythmic expression of these circadian genes leads to the rhythmic expression of *GIGANTEA* (*GI*), a gene that mediates output from the clock. *GI* works in conjunction with *FLAVIN-binding KELCH DOMAIN F BOX PROTEIN1* (*FKF1*) to induce the degradation of *CYCLING DOF FACTOR1* (*CDF1*). *CDF1* is

a repressor of *CONSTANS* (*CO*) transcription; therefore, activity of *GI* leads to increased expression of *CO* (97, 98). The protein produced by *CO* leads to the upregulation of *FLOWERING LOCUS T* (*FT*). *FT*. FT is a gene that produces florigen, a peptide that moves from leaves to the shoot apical meristem (SAM) where it induces floral initiation (99). Florigen interacts with FD, *SUPRESSOR OF OVEREXPRESSION OF CONSTANS* (*SOC1*) to promote the expression of *LEAFY* (*LFY*) and *APETALA1* (*AP1*) leading to floral initiation (99, 100).

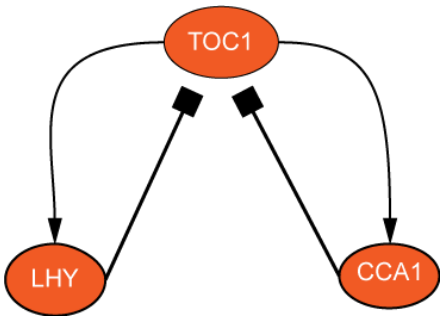

Figure 3. Illustration of the negative feedback loop that exists between TOC1, LHY, and CCA1 in the circadian cycle. This represents the core oscillator. TOC1 activates LHY and CCA1. LHY and CCA1 repress TOC1.

CO is quickly degraded by *PHYTOCHROME B* (*PHYB*) and *CONSTITUTIVE PHOTOMORPHOGENIC 1* (*COP1*). In the presence of light, cryptochromes and *PHYTOCHROME A* (*PHYA*) counteract the degradation of *CO* by blocking the activity of COP1 (101). In long days, the increased expression of *CO* by the circadian clock occurs during the light phase when CO stability is elevated. Thus, CO accumulates in Arabidopsis in long days, leading to activation of *FT* and flowering (101, 102).

Flowering Time in Sorghum

Figure 4 provides a basic model of the sorghum flowering time pathway. The gene ultimately responsible for a plant entering the flowering stages is *FT*. For a plant to transition from the vegetative stage to the flowering stages, a threshold of expression of *FT* must be met. Until the expression level of *FT* crosses this threshold, the plant will remain in the vegetative state (99, 103). There are different pathways that lead to increased *FT* expression depending on whether the photoperiod is short or long days (103). Both photoperiod and the clock are known to impact flowering in sorghum. Figure 4 denotes both of these influences. The maturity genes that impact the flowering time of sorghum affect flowering when the plant is grown under long days, thus the flowering time pathway under long day conditions is of greater interest which is the opposite of Arabidopsis (38, 39, 62).

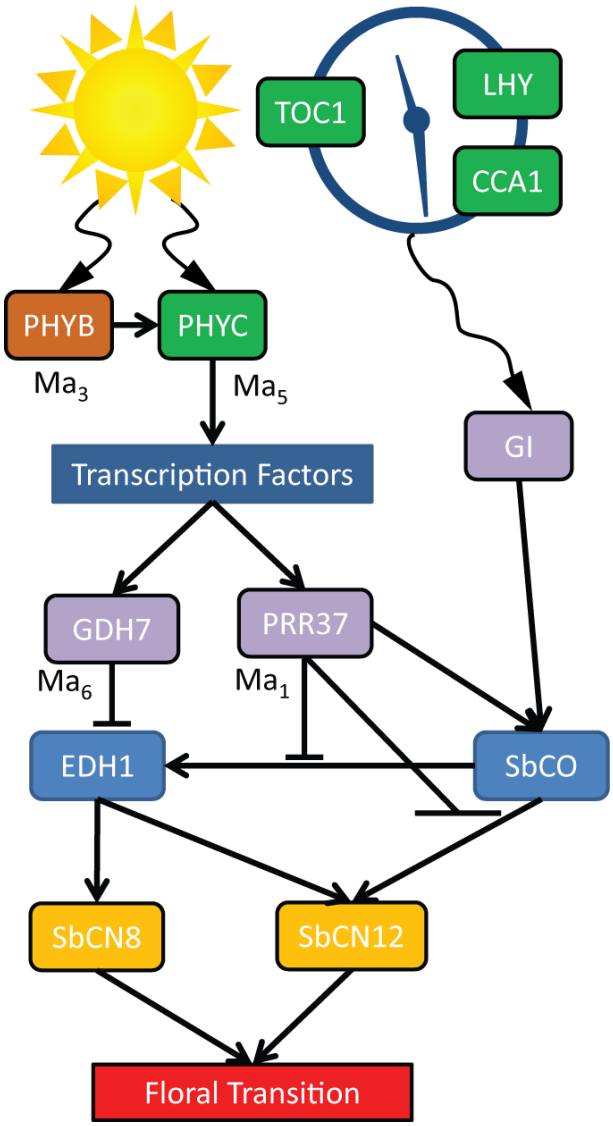

Figure 4. Simplified illustration of the flowering time pathway in *Sorghum bicolor.* Day length and the clock are denoted at the top as the two main pathways that converge to determine when the plant enters floral transition. Known maturity genes are denoted as well.

In sorghum and maize, FT is produced by homologs of Arabidopsis *FT*, and up-regulated by *CENTRORADIALIS* 12 (*CN12*), CN8 and CN15 (67, 104) which are up-regulated by Ehd1 and CONSTANS, thus regulators of these genes are integral to the flowering time pathway . However, *PRR37* represses

CO activity post-transcriptionally causing a delay in flowering under long day conditions. *PRR37* also represses *CO*'s activation of *Ehd1*. Additionally, *Ghd7* represses *Ehd1*, which activates *CN8* and *CN12,* thus delaying flowering. *PHYB* activates *PRR37* and *GHD7* expression thereby repressing *CN8, CN12* and most likely *CN15* and delaying flowering (67, 68, 104).

Inputs and Components of Grain Yield

Figure 5 denotes many of the subcomponents that influence the overall grain yield that each panicle produces (25, 31, 43, 48). For the purposes of these studies, grain yield is defined on a per plant basis, where plants contain only one panicle bearing grain. The grain sorghum populations were bred to produce only one panicle and the tillers were removed from populations that still had the potential to produce multiple panicles. This better enables the analysis of nutrient usage in the source vs. sink strength analysis since there are not multiple sinks to create additional variable factors. The goal in these studies is to eliminate as many external variables as possible so the variation that is found can be related to the given phenotype being studied rather than the other factors such as multiple sinks, shading, etcetera.

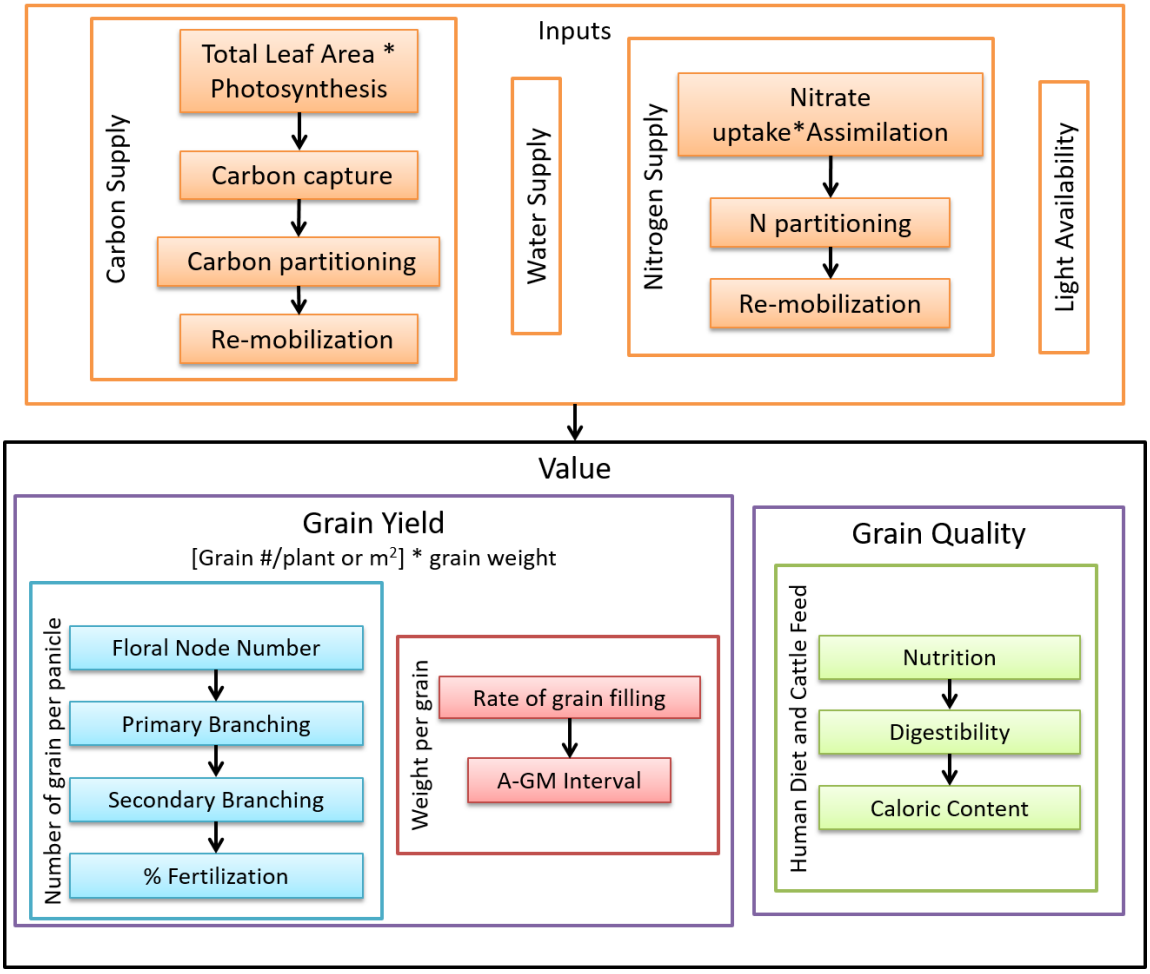

Figure 5. Inputs and components of grain yield. The orange box shows input factors and their subcomponents that influence the grain's value. The black box encompasses traits that determine the value of the grain including grain yield and quality and their subcomponents.

As previously discussed, grain yield can be divided into many different subcomponents. As stated in Equation 1, below, the main two sub-components of grain yield are grain number and grain weight. Both of these sub-components can be divided further into other phenotypes of interest. By studying these sub-components, a better understanding of the genetic basis of grain yield can be gained. Using this information, genes that have a more direct effect on

increasing the genetic components of grain yield may be identified. The logic behind this approach is that the identification of QTL for grain yield, grain weight, and grain number could lead to a better understanding of genetic factors that modify grain yield.

$$\frac{grain\ yield}{panicle} = \frac{grain\ number}{panicle} * \frac{grain\ weight}{panicle} \quad (1)$$

Grain yield and grain quality (composition) affect the value of the grain as shown in Figure 5. Many environmental factors impact the yield and composition of grain (33, 42, 49, 53, 105–107). The main contributors include the availability of light, nitrogen, and water (108–110). The potential grain yield is defined to be the upper limit on the amount of grain an individual plant can produce under optimal environmental conditions with a given genetic background; whereas the actual grain yield is the grain that is produced in an environment that is less than optimal (55, 108).

Grain Number

When analyzing grain yield, grain number is an excellent starting point due to the potential of this sub-component having a higher genetic control in many populations (32, 47). Grain number can be further broken down into subcomponents that may be phenotyped. The subcomponents of grain number include panicle architecture traits and percent fertilization (111–113). As can be observed in Figure 6, there is a large amount of variation in panicle architecture, even amongst grain sorghums. A typical grain sorghum panicle is shown in the

second panicle from the left in Figure 6; however, some grain sorghum's have a more open panicle with a higher degree of branching.

Figure 6. Variations of panicle architecture amongst different lines of *Sorghum bicolor*. The left picture is a diagram showing panicle anatomy. The left shows various phenotypes of grain sorghum panicles. Grain sorghum, typically has a tight packed sorghum as seen in the two panicles on the left. The panicles become more open from left to right. The ones on the right have an open panicle, long branches, and space between floral nodes.

The sorghum panicle has many different traits that can be phenotyped to gain a better understanding of how panicle architecture can affect overall grain yield. The panicle's "stem" is called the racchis. In the populations studied, there were typically between 6-10 nodes on an individual panicle. Each node, instead

of producing a leaf like is found on the stem, produces primary branches in a ring around the racchis; therefore, the total node number is a panicle architectural trait that would be expected to have a large impact on overall grain yield. Another trait of interest is the number of primary branches. The number of primary branches per node varies significantly depending on node number (development) and genotype. The first node on the panicle, the one closest to the peduncle, has the most primary branches. The number of primary braches per node decreases the farther the node is from the start of the panicle. A third phenotype is the degree of secondary branching. Secondary branches are derived from primary branches. Panicle branching nomenclature varies depending on the group studying the traits. In some studies, secondary branching is defined as any branch coming off the primary branch and tertiary branching is any branching coming off the secondary branch (114). Sorghum is one of the few cereals that have tertiary branching. In other studies, a secondary branch is defined as one that leads to further branching. A branch that comes off the primary branch and terminates in a floret is called a lateral branch (48). In this dissertation, the second nomenclature will be used. Node number, primary branching, and secondary branching phenotypes have an important impact on panicle architecture (48). QTL for these panicle architecture traits can be aligned with QTL for grain number to assess the influence of panicle architecture on grain number.

Analysis of percent fertilization traits can also be performed to determine the impact of percent fertilization on grain number. While panicle architecture sets the total number of florets, not all of the florets on the panicle end up producing a seed (115–117). In barley only about 40% of florets end up producing a seed (118). However, studying % fertilization is much more complicated than studying panicle architecture as aborted florets must be identified and compared to the total number of grain produced. Therefore, the work presented in this dissertation focuses on the impact of panicle architecture on grain number.

Grain Weight

Grain weight can be influenced by the rate of grain filling and the duration of the grain filling stage that occurs between anthesis and grain maturity (49, 55, 56). Typically, grain filling stage is ~30 days; however, there is variation of +/- 5 days in the duration of grain filling (seed development) (34). More extreme examples of variations in the duration of grain filling also exist (56). It is important to have the grain mature before the end of the season in a ‘good’ environment so the grain is viable for use. Assuming a constant rate of grain filling, the longer the grain has time to fill, the higher the grain weight. However, while the length of time from anthesis to grain maturity affects grain weight, it has been found in other cereals that the primary source of variation in grain weight is linked to the rate of the grain filling (119, 120).

The grain-filling rate (GFR) is the second main sub-component that impacts grain weight. Basic grain filling activities can be divided into three different phases: the lag phase, the grain filling phase, and the drying phase (34). The lag phase is when the grain is starting to develop and actively undergoing cell division and differentiation to setup the grain structure. During this period, the GFR is very low. The grain filling stage, as the name suggests, is when grain filling occurs at the highest rate during the grain-filling interval (121, 122). At the end of this phase, the grain weight is at the maximum. The grain weight decreases during the drying phase as the grain matures and a black layer is formed between the seed and the plant to eliminate nutrient flow (7). The black layer is composed of compressed cells that form an abscission layer that appears visually black and has been used as a sign of grain maturity in many different crops. It has been documented in other cereal crops, particularly maize, that the filling stages have a high genetic control but are affected by the input factors such as temperature and day length (56, 105).

Grain Quality

While grain quality does not necessarily impact grain yield directly, it still plays an important role in the viability of the seed and seed value. The overall value of grain is determined by grain yield and grain quality (6, 123). Grain factors include nutrition content, digestibility, and caloric content. Nutritional content is a very important aspect if grain is going to be used as food or feed (124). Digestibility refers to the amount of the nutrients that are absorbed by

animals compared to calories/nutrients that are not absorbed (125). Low digestibility decreases the value of the seed. QTL can be identified for low digestible proteins, for example, and selection can and has occurred for an increase in digestibility (126, 127). The percent of the protein that is absorbed by the body is defined to be the digestibility coefficient. The digestibility coefficient varies amongst different animals so the composition of seeds can be improved by modifications that improve digestibility by the target animal (128, 129).

Caloric content is a very important grain quality factor. Typically, sorghum is utilized for human consumption in third world countries where malnutrition is one of the leading causes of death (130). Increasing the available calories to the people who grow sorghum as a primary food source would aid in the battle against malnutrition and directly contribute to the 2050 mandate to increase grain production to feed the growing global population.

Inputs

While the subcomponents of grain number and grain weight are important to consider, it is also important to consider environmental input factors that impact grain yield and all of its subcomponents. As mentioned in relation to GFR, light availability is an important factor in photosynthesis, yield, and nutrient accumulation. Light availability can be influenced by day length, intensity, and shading by other crops (36).

Water limitation is a major factor that can limit grain yield for all crops (122, 131, 132). Since sorghum is drought tolerant, it is not impacted by drought

as severely as other cereal crops such as maize and rice; however, sorghum is still dependent on water availability (133). While water availability is not a current constraint in all parts of the world, in regions where irrigation is not available water supply is often a limiting factor when growing crops (134). As the need to increase grain production continues to rise, the increased amount of water being devoted to irrigation needs will cause water to be one of the most limiting factors in grain production (8, 11). This is just one of the many reasons to work on increasing grain yield in drought tolerant crops.

Two of the most important input factors to consider when discussing environmental conditions that influence grain yield are carbon and nitrogen assimilation. Nitrogen and water have a close relationship as it has been found that nitrogen uptake and water supply are positively correlated (135). Nitrogen is important for plant growth due to, in part, to its involvement in photosynthesis. Chlorophyll has 4 nitrogen atoms making nitrogen availability critical to photosynthesis as well as further down in the photosynthesis pathway through ribulose-1,5-bisphosphate carboxylase oxygenase (RuBisCO) (136). Rubisco has the potential to contain as much as 30% of the total N content of the leaves (136).

The energy from photosynthesis is used for the assimilation of carbon dioxide and nitrate ions to form carbohydrates and amino acids. Nitrogen plays an important role in the synthesis of amino acids which in turn, is important in many of the composition components of the grain and plants itself, specifically

protein content. Sorghum is a C4 crop and compared to C3 crops, is more efficient in terms of photosynthesis and nitrogen use (19, 137).

Depending on the genotype, anywhere between 28%-70% of N is partitioned in grain (138, 139). This is highly variable due to the source vs sink needs of the plant. In order to limit the need for N and C, nutrients are remobilized from the stem to the panicle (140). Understanding of this remobilization processing has become a focus of furthering the understanding of grain yield and decreasing fertilizer need.

Overview of Dissertation

Each chapter reports on a different phase in the developmental cycle of sorghum grain and its influence on grain yield. Chapter II analyzes grain yield and its sub-components grain number and grain weight as well as sub-components of grain number. These phenotypes were analyzed via bi-parental QTL mapping in RIL populations. Three populations were used in this study, BTx642 x Tx7000, BTx623 x IS3620c, and SC170 x M35-1. In Chapter II, QTL that affect panicle architecture were identified. These phenotypes identify important QTL that affect panicle architecture as well as grain number.

Chapter III focuses on phenotypes that influence grain weight during the grain filling stage of grain development. While grain weight is often seen as less genetically and more environmentally controlled, Chapter III discusses the population dependence of the genetic control of grain weight. Phenotyping was performed to analyze grain composition, stem hollowing, and stem biomass

composition. Grain composition is an important subcomponent of not only grain weight but is an important factor into grain quality as well. Monitoring of the stem hollowing phenotype enables identification of genes which are actively involved in the reallocation process of nutrients from the stem to the panicle during the grain filling stage. Alternatively, by looking at stem biomass composition, specifically the change in biomass that occurs during grain filing, nutrients that are being reallocated to the panicle from the stem may be identified.

In Chapter IV, genes that control the onset of floral induction are discussed and reported. Maturity gene 2, Ma_2, is identified via QTL and then map-based cloned to identify the gene. To further confirm gene identification, whole genome sequences were evaluated to determine additional ma_2 allelic variants. Ma_2's role in the flowering time pathway will be characterized via qPCR in 100M vs 80M which are isogenic for flowering time genes with the exception of Ma_2. Additional flowering time genes are characterized in the same parental lines to identify Ma_2's location in the flowering time pathway.

Through analysis of traits at different stages of development, the optimal conditions and genetic backgrounds can be found for each stage. Applying this knowledge can have an important impact on grain yield production that will aid in the 2050 mandate to help feed the growing population by having significant gains in an individual plant that will scale up to an increase in grain yield across hectares.

CHAPTER II

THE GENETIC BASIS OF VARIATION IN GRAIN YIELD

Introduction

Crop production must increase by 70% by 2050 in order to meet the expected requirements of an increasing world population that is undergoing rapid development (10). Sorghum [*Sorghum bicolor* (L.) Moench], a drought tolerant crop used for grain, forage and bioenergy production is native to Africa, and now used throughout the world (7). Over the past century, there has been significant investment in breeding efforts to improve sorghum crop fitness for new regions of production. However, compared to other major cereal crops, much less investment has been made in crop breeding and genetic improvement of grain yield in sorghum (1, 28, 32, 33, 48, 54, 119, 141). There are many genetic and environmental factors that influence grain yield, making the improvement of grain yield a significant challenge (107, 122, 142).

To understand the genetic/biochemical pathways that affect grain yield it is helpful to characterize the individual traits that contribute to overall grain production. Grain yield is the product of grain number and grain weight. Grain yield can be defined per panicle, per plant, or per unit of land. In the current study, grain yield per panicle of a single stem plant will be analyzed. Equation 1, in Chapter I, defines the equation for grain yield per panicle.

Grain sorghum has been under selection for reduced tillering, and as a consequence, many grain sorghum genotypes and management systems result in plants that produce a single stem with one panicle (123). Analysis of grain yield on a per plant (and panicle) basis simplifies the analysis of genetic control of the traits that affect grain yield. In sorghum, grain number is largely determined during the booting stage that occurs between floral initiation and anthesis (34, 42). During the booting stage, panicle architecture is established and this sets the maximum number of seed possible for each panicle. The number and types of panicle branches formed impacts the number of florets and potential grain number. However, many of the nascent florets are aborted during this stage for a variety of reasons. The environment is the largest factor that influences floret abortion. During the booting phase, the plant will abort potential seed sites ensuring seeds that are produced have enough nutrients available to survive (143). Other factors that influence floret abortions include genotypes with poor fitness that exhibits low levels of photosynthesis or nitrogen assimilation, and proximity to other floret sites (144). In Scot pines, the floret abortion rate has been found to be as high as 76% in self-pollinated lines due to bad genotypes (145, 146). Once the floret number is determined, the next factor that limits seed production is percent fertilization. Percent fertilization varies greatly across environments and genotypes and is influenced by external factors such as temperature, wind, and humidity and pollen viability (147).

Several terminologies are used to describe branching and panicle architecture (48, 114). For the purpose of this study, a primary branch arises from the rachis (stalk of the panicle), and a secondary branch is a branch that originates from a primary branch and leads to further branching. A lateral branch that forms on a primary or secondary branch terminates in a lateral flower and does not lead to further branching. This nomenclature is illustrated in Figure 7, and is consistent with the rice nomenclature for panicle architecture (48). To investigate the genetic basis of variation in branch number and its impact on grain number, several panicle phenotypes were analyzed in sorghum. Floral node number, the number of primary branches, the number of secondary branches, and branch length were all investigated as potential contributors to the variation of grain number.

Unlike many other cereal crops, sorghum is known to produce multiple seed at the end of branches. This occurs when branch termini contain two seeds, leading to the term “twin seeding” in sorghum. A sorghum *multiseed* (*msd*) mutant has also been reported to inhibit sorghum’s ability to terminate florets, thereby causing triseeding to occur (45). Triseeding occurs when sorghum does not terminate any florets formed on a terminal branch. However, environment was found to affect grain number even in the *msd* background (46).

Increasing grain number may cause a decrease in grain weight due to allocation of limiting carbon and nitrogen resources among developing seeds. For example, grain produced by the *msd*-mutant is often very small in size due

to a lack of required nutrients available to compensate for the increase in grain number (46). To ensure production of some viable seed, sorghum terminates a subset of its florets during the booting stage; therefore, nutrient limitations during the booting phase are detrimental to overall grain yield (105, 107, 112, 144).

Panicle architecture varies significantly amongst parental lines and populations. Four populations, derived from four parental crosses, were analyzed for grain yield traits, with a focus on panicle architecture. The first population analyzed was derived from a cross of BTx642 x Tx7000. BTx642 is a stay-green genotype selected for its drought tolerance, while Tx7000 is an R-line that was originally developed for grain production prior to the use of hybrids. Both of these parents have the tight, compact panicle with short branch lengths that is typical of grain sorghum. A second population, BTx623 x IS3620c population, was selected because the parental lines and RILs varied extensively in panicle architecture. BTx623 is the genotype used for the sorghum reference genome (148), and IS3620c is a guinea race sorghum that has an open panicle and exhibits the seed shattering phenotype. In wild sorghum, seed shattering aids in seed dispersal as this is the process of seeds detaching from the panicle making the plant self-sufficient in beginning the next generation (149–151). In grain sorghum populations, the seed shattering phenotype hinders the harvesting of grain, thus, seed shattering has been strongly selected against in most grain sorghum lines. The third population used in this study was a cross of SC170 x M35-1. SC170 is a caudatum race sorghum with wide short stems and

is one of the grain sorghum lines used in the United States that was generated by the sorghum conversion program (16, 148, 152). Maldandi 35-1 (M35-1) is a grain sorghum used in India and is taller and thinner stemmed than SC170. The fourth population used to study the genetic basis of secondary branching was derived from a cross of Standard broomcorn x SC170. Broomcorn panicles have long branches ideal for creating brooms (6). Broomcorns have low levels of secondary branching compared to most other grain sorghum lines. All four of these populations were analyzed to identify quantitative trait loci that affect grain yield traits.

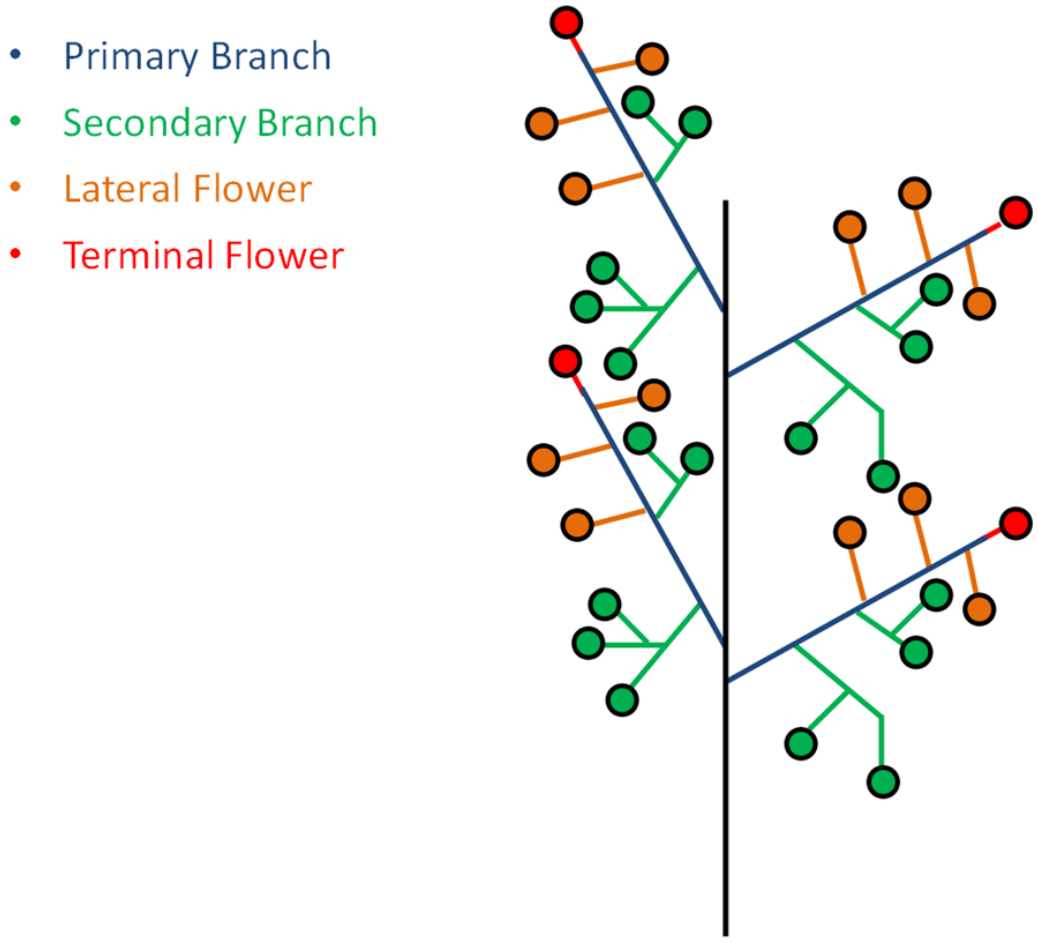

Figure 7. Diagram of panicle architecture terminology used for phenotyping traits that impact overall variation in grain number. Primary branches are shown in blue and are the branches that arise from the rachis. Secondary branches are shown in green and are branches that arise from primary branches that lead to further branching. Lateral flowers are branches that arise from the primary branch that lead directly to florets and are shown in yellow. Terminal flowers are branches that lead directly to florets at the end of a primary branch and are shown in red.

Results

BTx642 x Tx7000

Analysis of the BTx642 x Tx7000 RIL population in 2012 for grain yield per panicle revealed a single QTL peak on LG03 (Figure 8). The grain yield QTL on LG03 was also aligned with a grain weight QTL (Figure 9). A second QTL affecting grain weight was mapped on LG01. The QTL for grain weight on LG01 aligned with a QTL for grain number per panicle. Figure 10 shows that the causative allele for increased grain number per panicle is from Tx7000 whereas, Figure 9 shows the causative allele for increased grain weight was derived from BTx642. A QTL was also identified for grain number on LG07. This QTL does not align with any QTL for grain weight. This suggests that the QTL on LG07 is an independent modulator of grain number like the QTL on LG03 is an independent modulator of grain weight.

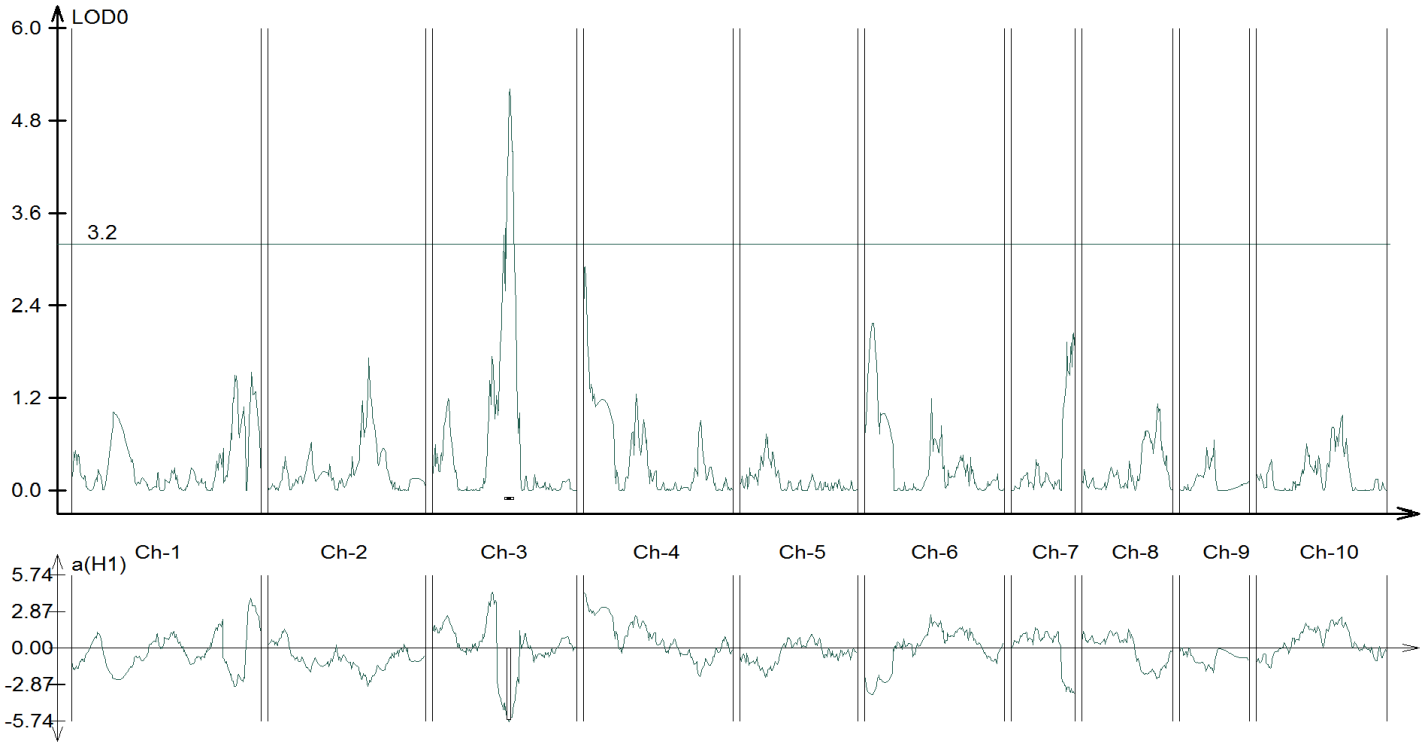

Figure 8. A QTL map showing significant peaks of grain yield trait in BTx642 x Tx7000 (2012). The additive graph (lower panel) is also displayed indicating parental influence at a given loci. Positive is BTx642 and Negative is Tx7000. Only one QTL is identified on LG03.

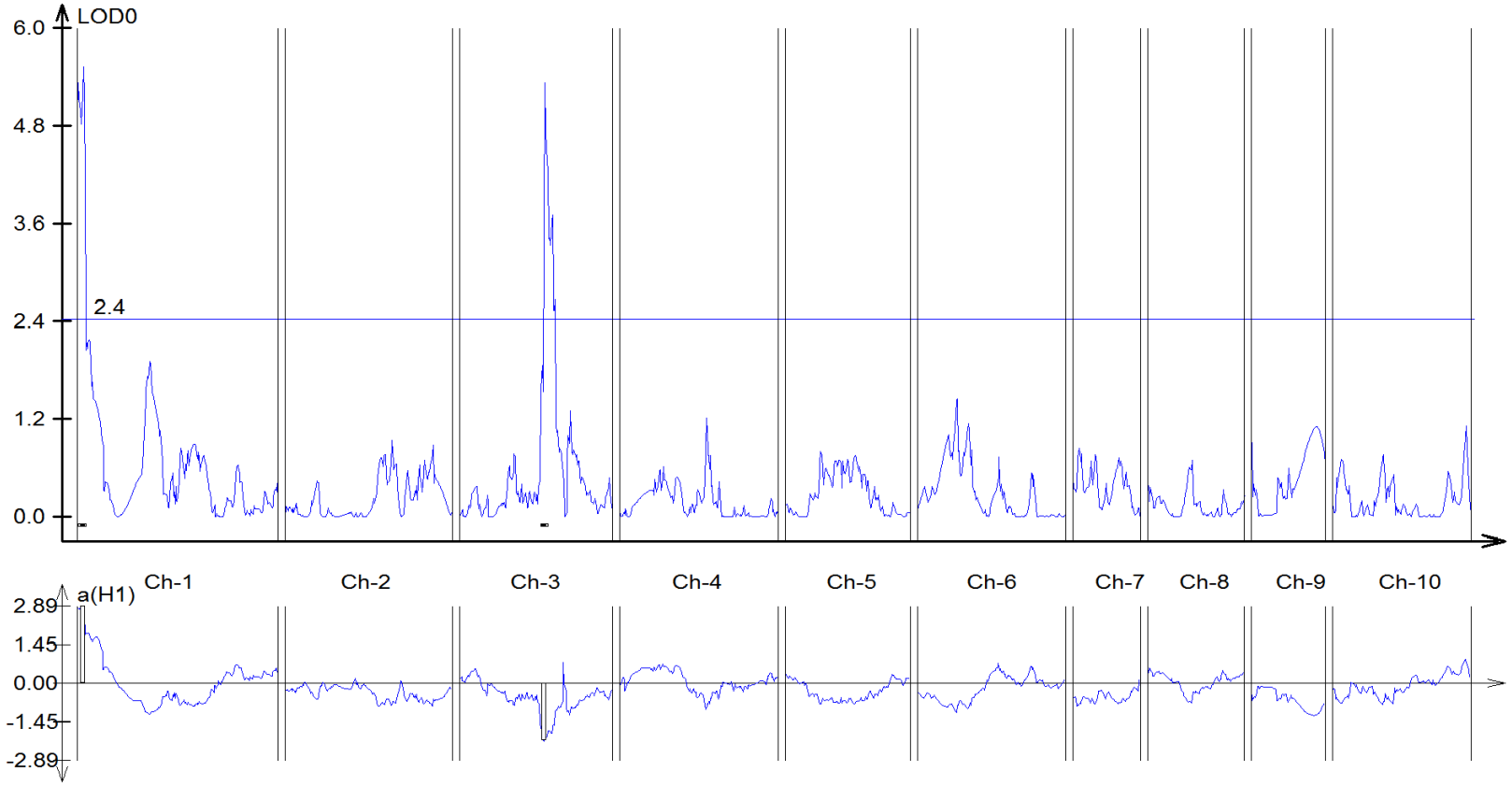

Figure 9. A QTL map showing significant peaks of grain weight trait in BTx642 x Tx7000 2012 College Station. The additive graph (lower panel) is also displayed indicating parental influence at a given loci. Positive is BTx642 and Negative is Tx7000. Two QTL were identified. One QTL is the same QTL identified in grain yield on LG03 and a new QTL was identified on LG01.

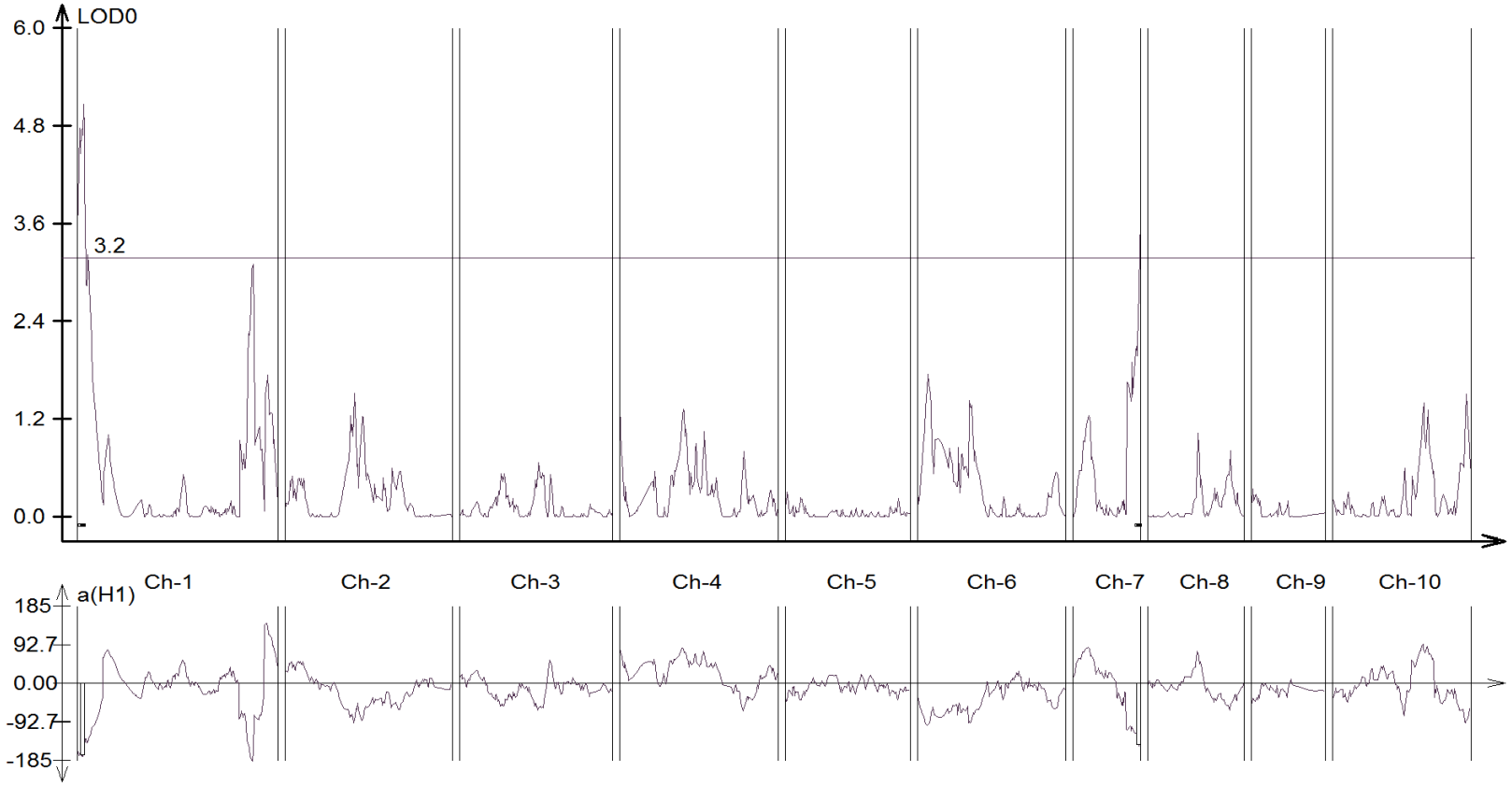

Figure 10. A QTL map showing significant peaks of grain number trait in BTx642 x Tx7000 2012 College Station. The additive graph (lower panel) is also displayed indicating parental influence at a given loci. Positive is BTx642 and Negative is Tx7000. Two QTL were identified. One QTL is the same QTL identified in grain weight on LG01 and a new QTL was identified on LG07.

The QTL identified for grain yield per panicle, grain weight, and grain number for all years and environments are illustrated in Figure 11 and Table 1. It has been proposed that grain number has a significant impact on grain yield (47). Since panicle architecture affects potential grain number the genetic basis of variation in panicle architecture was analyzed. The QTL identified for number of primary branches per node, total primary branching, total number of nodes, and total secondary branches on nodes 3 and node 6 are reported in Figure 12. Panicle branch analysis was performed for populations grown in CS (2012), Lubbock (2012), and CS (2013). A QTL with the highest LOD score was identified on LG01. This QTL was identified when analyzing grain yield, grain weight, grain number, and both secondary branching phenotypes. On LG03, a QTL for grain yield and grain weight was observed consistently. An additional primary branch number QTL was identified on the end of LG03 for the number of primary branches on node 1, node 4, total floral node number, and total primary branch number. Another branching QTL of interest is on LG10. This QTL was identified in all traits relating to primary branch number as well as floral node number and secondary branching in the Lubbock (2012) trial around ~55 Mb. On LG06, a secondary branching QTL was identified in only the Lubbock (2012) population. Many other QTL were identified in only one of the populations suggesting these QTL identified are particular to the environment in which the population was grown. Detailed analysis of all QTL labeled in Figure 9 is available in Table 1.

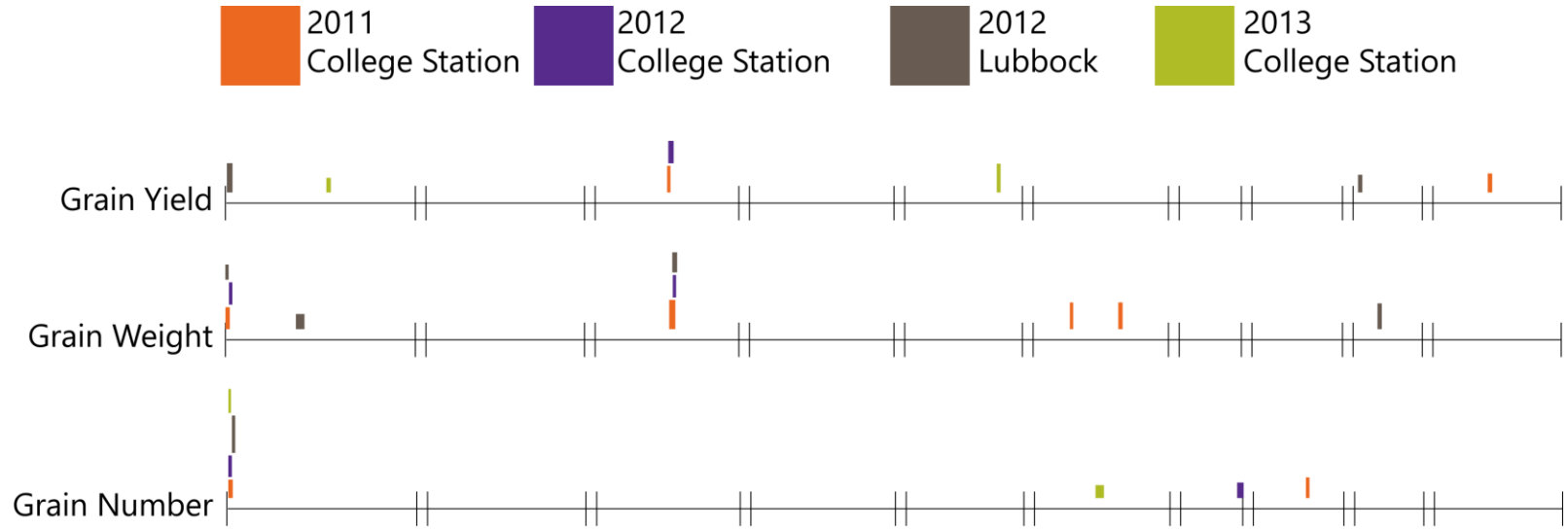

Figure 11**.** All QTL identified in grain traits for the BTx642 x Tx7000 population. Results are color coded by the year and environment the population was grown. The height of the box representing a QTL corresponds to its relative LOD score while the width of each box indicates the 1 LOD interval of the identified QTL. Two consistent QTL were identified. One QTL was on LG01 and the second QTL was on LG03.

Table 1**.** Information describing QTL identified for grain yield, grain weight, and grain number through analysis of the College Station BTx642 x Tx7000 RIL populations. Chr. indicates chromosome, R^2 is the amount of the additive variance that is explained and Add Var. is the additive variance.

Population	Trait	Chr.	Add. Var.	R^2	Peak (cM)	Peak (bp)	Left (cM)	Right (cM)	Left (bp)	Right (bp)
2011 CS	Grain Weight	1	1.4929	0.09425	2.01	1,827,819	0.1	2.9	456,101	1,937,629
2011 CS	Grain Number	1	-139.2	0.12119	4.41	2,166,013	2.8	5	1,936,111	2,336,220
2011 CS	Grain Yield	3	-4.6691	0.19237	68.11	54,415,446	67.5	69.1	53,322,260	54,934,613
2011 CS	Grain Weight	3	-1.5564	0.13092	72.21	56,113,862	69.4	75	54,468,087	57,060,786
2011 CS	Grain Weight	6	-1.6529	0.12661	35.11	37,278,450	34.5	35.5	32,053,994	37,521,479
2011 CS	Grain Weight	6	1.1868	0.07586	81.01	51,812,524	79.9	82.1	51,221,632	51,840,141
2011 CS	Grain Number	8	126.572	0.1305	52.21	54,958,748	50.6	53.8	54,664,489	55,378,910
2011 CS	Grain Weight	9	-1.3287	0.09318	25.21	3,233,660	24	27.6	3,078,286	3,513,606
2011 CS	Grain Yield	10	5.1946	0.13396	52.81	9,672,775	51.4	53.9	9,551,948	10,205,350
2012 CS	Grain Weight	1	2.8928	0.22046	5.01	2,337,436	3.3	6.5	1,943,704	2,569,406
2012 CS	Grain Number	1	-174.8	0.17485	5.01	2,337,436	2.8	6	1,936,111	2,507,430
2012 CS	Grain Yield	3	-5.7374	0.18891	70.41	54,818,723	68.6	73.5	53,762,236	56,418,535
2012 CS	Grain Weight	3	-2.1678	0.12263	73.41	56,413,248	72.8	75.8	56,130,664	57,096,942

Table 1. Continued.

Population	Trait	Chr.	Add. Var.	R^2	Peak (cM)	Peak (bp)	Left (cM)	Right (cM)	Left (bp)	Right (bp)
2012 CS	Grain Number	7	-151.53	0.12193	57.71	53,849,060	56.3	58.3	49,425,392	55,545,623
2012 Lubbock	Grain Yield	1	-4.6113	0.26278	7.21	3,363,641	6.6	11.8	2,580,440	5,968,511
2012 Lubbock	Grain Weight	1	1.4012	0.08059	0.61	1,171,958	0	2	42,185	1,827,352
2012 Lubbock	Grain Weight	1	-1.53	0.09654	71.61	23,044,203	66.1	74	19,761,932	24,800,046
2012 Lubbock	Grain Number	1	-217.36	0.28847	6.61	2,581,544	6.1	7.5	2,525,271	3,381,181
2012 Lubbock	Grain Weight	3	-1.7281	0.12682	75.61	57,088,355	72.3	76.7	56,114,974	57,200,018
2012 Lubbock	Grain Yield	9	-3.5382	0.14797	4.71	1,214,104	4.6	8.6	1,207,089	1,718,209
2012 Lubbock	Grain Weight	9	-1.9766	0.156	25.21	3,233,660	22.8	26.8	2,892,559	3,489,597
2013 CS	Grain Yield	1	-2.9051	0.16566	95.71	58,853,599	94.7	98.7	58,638,217	59,524,030
2013 CS	Grain Number	1	-192.81	0.32959	3.31	1,943,961	2.9	5.2	1,937,629	2,360,545
2013 CS	Grain Yield	5	-3.9368	0.34109	87.61	67,436,610	86.3	89.8	67,274,233	67,777,547
2013 CS	Grain Number	6	-126.48	0.14574	61.51	45,771,472	58.5	66.1	44,856,542	46,475,664

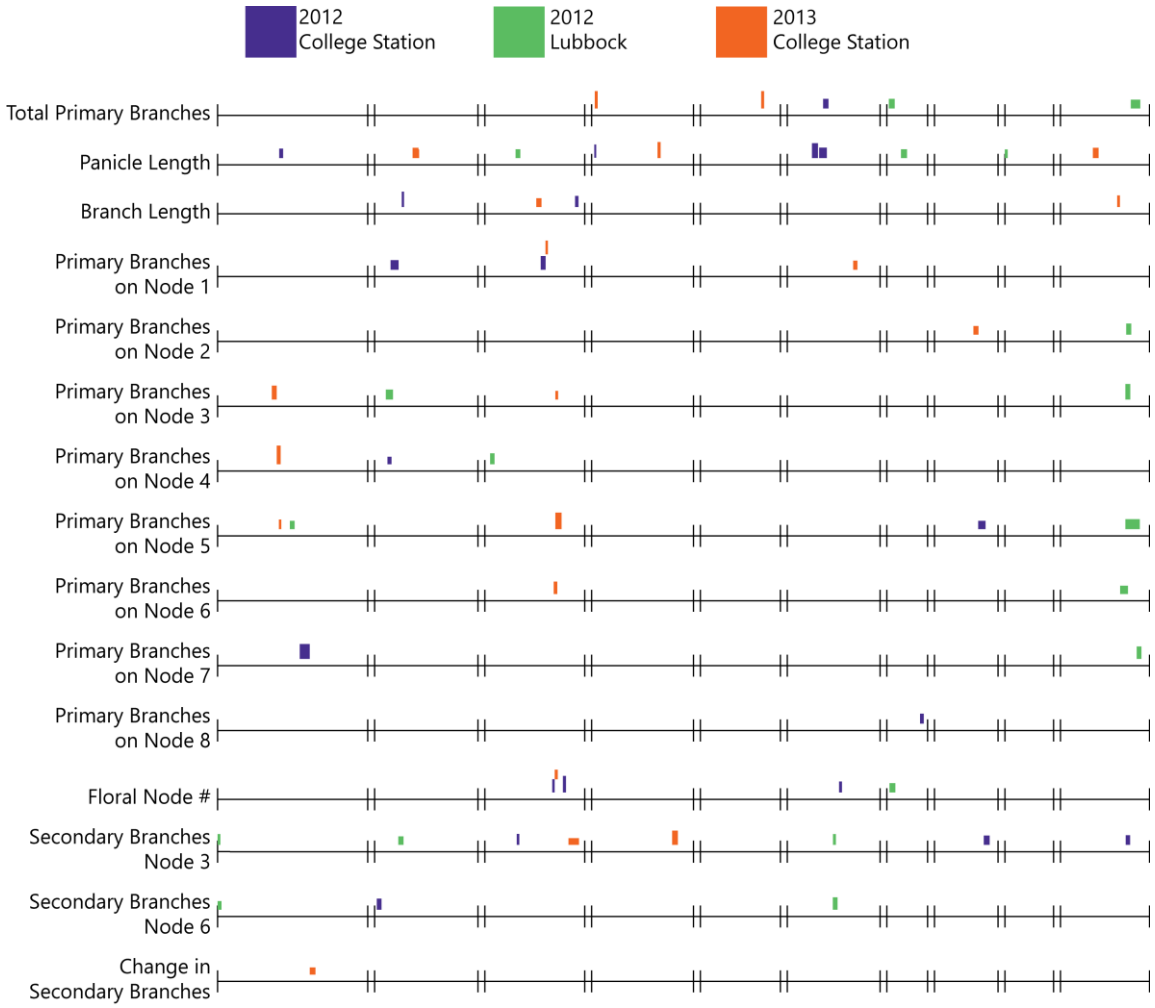

Figure 12. All QTL identified in branching traits for the BTx642 x Tx7000 population. Results are color coded by the year and environment the population was grown. The height of the box representing a QTL corresponds to its relative LOD score while the width of each box indicates the 1 LOD interval of the identified QTL. Branching traits had 3 consistent QTL of interest. These QTL are on LG03, LG06 and LG10.

Table 2. Information describing QTL identified for panicle architecture through analysis of the College Station BTx642 x Tx7000 RIL populations. Chr. indicates chromosome, R^2 is the amount of the additive variance that is explained and Add Var. is the additive variance.

Pop.	Trait	Chr.	R^2	Add. Var.	Peak (cM)	Peak (bp)	Left (cM)	Right (cM)	Left (bp)	Right (bp)
2012 CS	Panicle Length	1	0.119	-1.221	93.31	48,632,804	90.1	95.7	46,114,704	50,404,250
	Panicle Length	4	0.171	1.471	5.41	1,072,471	4.1	6.6	822,462	1,253,605
	Panicle Length	6	0.193	1.604	39.01	41,856,240	36.3	44.8	40,120,199	44,708,620
	Panicle Length	6	0.180	1.483	52.81	45,000,000	46.8	57.8	44,708,620	45,730,642
	Branch Length	2	0.196	1.012	41.41	7,866,848	39.7	42.8	6,873,304	8,404,730
	Branch Length	3	0.139	-0.878	135.01	72,000,000	132	136.6	71,832,221	72,711,436
	Node 1	2	0.126	0.673	33.21	6,411,540	23.5	34.9	4,899,014	6,873,304
	Node 1	3	0.202	-0.831	83.61	57,500,000	81.7	88.8	57,226,675	58,969,358
	Node 4	2	0.121	0.764	22.91	4,899,014	18.9	30.2	4,720,368	6,411,540
	Node 4	3	0.137	0.852	2.01	1,000,000	0	5.8	0	1,638,809
	Node 5	8	0.129	-0.605	68.91	51,000,000	64	74.5	50,116,222	52,075,617
	Node 7	1	0.131	0.615	125.61	55,779,723	120.7	127.4	54,650,601	56,245,572
	Node 7	7	0.211	0.773	54.11	9,662,976	52.3	55.1	7,115,441	52,727,152
	Node 7	8	0.194	-0.728	73.01	51,620,632	69.5	84	50,817,616	54,322,182
	Node 8	7	0.169	0.788	52.41	8,240,709	48.5	53.8	6,702,491	9,662,975
	Floral Node #	3	0.162	-0.833	100.01	63,522,585	98.6	102.5	61,551,692	64,376,683

Table 2. Continued.

Pop.	Trait	Chr.	R^2	Add. Var.	Peak (cM)	Peak (bp)	Left (cM)	Right (cM)	Left (bp)	Right (bp)
2012 CS	Floral Node #	3	0.219	0.894	114.71	69,116,135	114.1	118.5	68,982,966	69,728,620
	Floral Node #	6	0.137	-0.569	78.01	52,094,273	75.8	79.8	51,499,148	52,338,107
	Total Branch #	6	0.128	-3.753	58.01	45,730,642	52.4	60.1	44,708,620	46,247,880
	Secondary Node 3	3	0.141	0.485	48.31	10,391,393	46.8	50.2	10,102,706	10,683,496
	Secondary Node 3	8	0.130	0.471	76.51	52,750,000	72.2	80.5	51,148,148	53,493,827
	Secondary Node 3	10	0.123	0.455	99.91	56,396,650	95.3	101.6	55,524,582	56,511,327
	Secondary Node 6	2	0.194	-0.537	5.11	2,000,000	3.2	9.9	1,588,403	3,687,906
2012 Lubbock	Floral Node #	7	0.126	-1.150	8.11	1,900,000	3.9	12.2	1,400,000	2,200,000
	Secondary Node 3	1	0.132	0.505	0.01	1,000,000	0	2.4	0	1,900,000
	Secondary Node 3	2	0.108	-0.458	39.21	6,873,000	34.6	42.1	6,600,000	8,200,000
	Secondary Node 3	6	0.127	0.513	69.71	48,297,000	67	70.9	49,500,000	51,200,000
	Secondary Node 6	1	0.127	0.305	0.01	1,000,000	0	2.2	0	1,900,000
	Secondary Node 6	3	0.084	-0.322	74.31	54,934,000	71.3	77	53,700,000	56,000,000

Table 2**.** Continued.

Pop.	Trait	Chr.	R^2	Add. Var.	Peak (cM)	Peak (bp)	Left (cM)	Right (cM)	Left (bp)	Right (bp)
2012 Lubbock	Secondary Node 6	6	0.153	0.332	70.31	48,499,000	66.6	73.3	47,500,000	50,000,000
	Node 2	10	0.147	-0.772	99.91	56,394,000	96.1	103.3	56,000,000	56,700,000
	Node 3	2	0.129	-0.737	20.51	4,800,000	16.5	26.7	4,720,000	5,400,000
	Node 4	10	0.204	-0.910	98.51	56,100,000	94.9	101.9	55,600,000	56,600,000
	Node 4	3	0.142	-0.851	9.21	1,991,940	7.8	10.7	1,900,000	2,000,000
	Node 5	3	0.141	-0.893	132.01	71,832,221	128.6	134.9	70,700,000	72,000,000
	Node 5	1	0.104	-0.710	108.91	52,647,802	105.6	112.5	51,650,000	53,200,000
	Node 6	10	0.125	-0.770	101.61	56,511,327	95	104.6	55,800,000	56,728,379
	Node 7	10	0.126	-0.787	107.61	56,900,000	104.6	115.7	56,728,379	58,500,000
	Node 6	10	0.109	-0.756	92.71	55,282,000	87.3	98.6	55,000,000	56,200,000
	Node 7	10	0.184	-1.127	117.21	57,800,000	111.4	118.2	57,000,000	59,063,000
	Total Primary	7	0.123	-6.748	7.11	1,744,990	3	11.4	1,300,000	2,100,000
	Total Primary	10	0.107	-6.256	108.21	56,922,188	102.9	116.5	56,600,000	58,500,000
	Panicle Length	3	0.119	1.016	48.31	10,391,393	45	51.7	10,000,000	11,000,000
	Panicle Length	7	0.137	0.939	24.41	3,350,000	20.6	29.3	2,900,000	4,100,000
	Panicle Length	9	0.124	-0.906	1.01	1,000,000	0	3	0	1,113,094
2013 CS	Node 1	3	0.209	-0.948	90.31	59,366,053	88.7	92	58,555,059	59,753,639
	Node 1	6	0.127	0.760	97.31	55,318,076	96.5	102.4	55,157,298	56,586,356
	Node 2	8	0.152	-0.589	61.61	50,116,222	57	64.3	49,587,492	50,817,616
	Node 3	1	0.224	-0.912	80.51	23,893,666	79.2	86.3	21,904,980	46,114,704

Table 2. Continued.

Pop.	Trait	Chr.	R^2	Add. Var.	Peak (cM)	Peak (bp)	Left (cM)	Right (cM)	Left (bp)	Right (bp)
2013 CS	Node 3	3	0.156	-0.719	104.71	65,525,393	103	105.8	64,376,683	66,434,503
	Node 4	1	0.307	-1.258	90.91	47,071,208	86.4	91.9	43,997,458	47,401,666
	Node 5	1	0.130	-0.678	90.91	47,071,208	89.6	93	46,114,704	48,632,804
	Node 5	3	0.244	-1.004	104.71	65,525,393	103.1	112	64,376,683	67,939,827
	Node 6	3	0.220	-1.272	104.71	65,525,393	100.6	105.7	63,522,585	66,434,503
	Total Primary	4	0.246	6.192	5.41	1,072,471	4.8	6.2	1,008,603	1,253,605
	Total Primary	5	0.223	6.094	90.11	55,598,516	88.9	91.4	55,308,844	56,083,187
	Floral Node #	3	0.160	0.673	104.11	64,646,936	102.1	106.5	63,522,585	66,434,503
	Secondary 3	3	0.122	-0.801	135.41	72,173,211	122.3	137.5	69,728,620	73,145,525
	Secondary 3	4	0.281	1.230	121.41	63,335,696	117.8	126	62,670,633	64,400,937
	Secondary Change	1	0.135	-0.821	142.31	58,108,968	134.6	142.9	57,158,862	59,563,845
	Panicle Length	2	0.144	-1.151	63.71	47,000,000	55.6	64.8	8,404,730	53,502,998
	Panicle Length	4	0.275	1.575	97.81	59,000,000	96.3	100.6	58,935,877	59,838,225

Table 2. Continued.

Pop.	Trait	Chr.	R^2	Add. Var.	Peak (cM)	Peak (bp)	Left (cM)	Right (cM)	Left (bp)	Right (bp)
2013 CS	Panicle Length	10	0.183	1.289	51.61	9,000,000	47.3	55.6	8,056,948	10,075,847
	Branch Length	3	0.127	-0.456	77.41	56,238,895	75.3	82.7	54,415,446	57,803,734
	Branch Length	10	0.218	0.586	83.91	53,832,398	82.8	84.8	53,551,604	53,832,398

Multiple QTL mapping (MQM) was utilized in the CS (2012) trial to identify potential interactions between QTL and additional QTL for grain number. MQM is an expansion on typical QTL mapping that allows for a deeper analysis of the trait of interest. MQM can identify additional QTL due to its higher statistical power and 10,000 penalty calculations to determine thresholds providing more confidence in identifying smaller effect QTL as well as linked QTL. Figure 13 shows the results for the MQM grain number analysis. Two additional grain number QTL were identified via MQM, one on LG04 and one on LG10. The QTL identified by MQM on LG10 is at 79cM or ~59.8Mb so it does not co-align with the QTL identified in the Lubbock (2012) study. MQM analysis also identified an epistatic interaction between QTL on LG01 and LG10. When the QTL on LG01 is fixed for the BTx642 parent (A) and LG10 QTL for Tx7000 (B), a significant increase in grain yield is observed that is much more than the additive effect of independent interactions, indicating there is an epistatic interaction between these 2 loci. The best model identified for grain number only explained 33% of the variance indicating many additional QTL with small effect have yet to be identified.

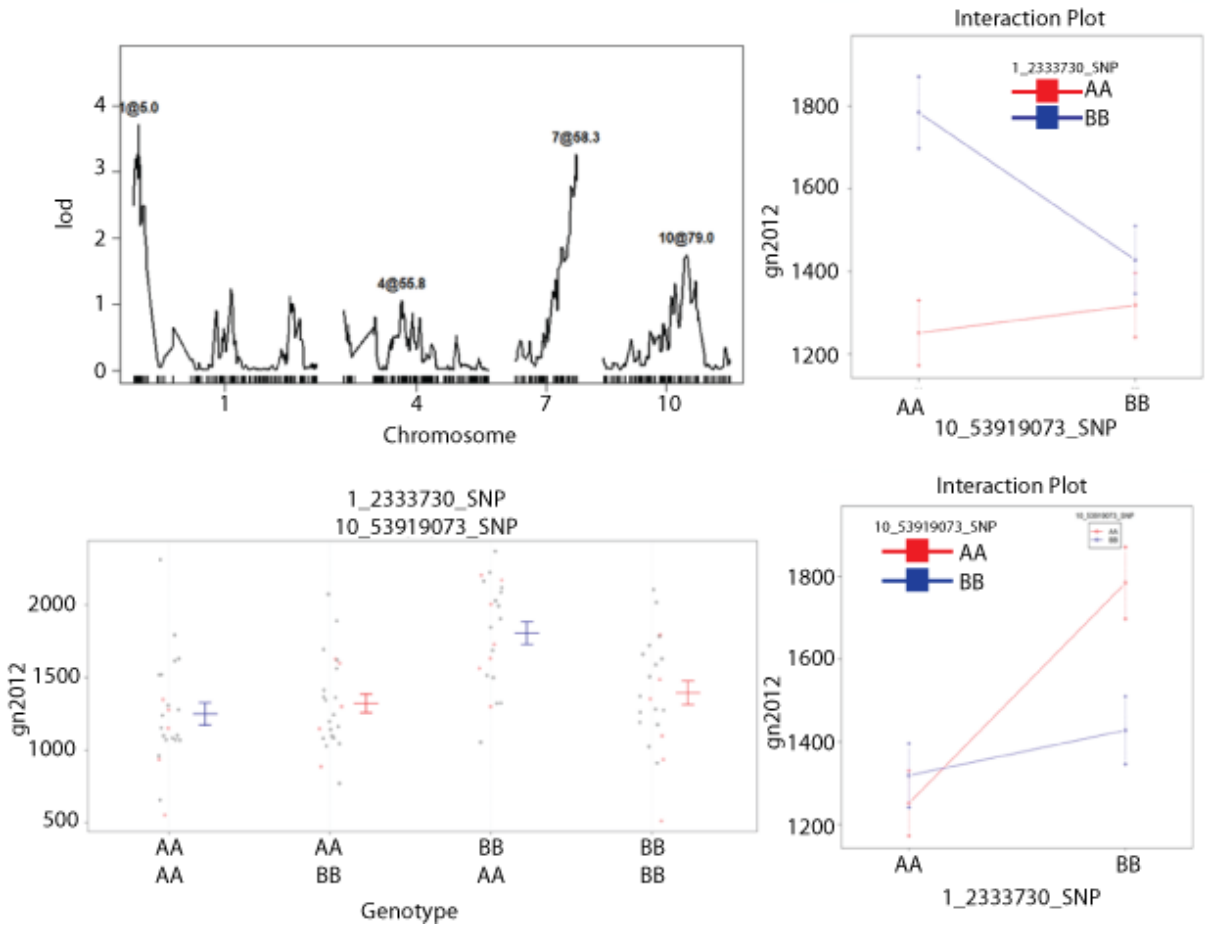

Figure 13. MQM results from 2012 grain number phenotype. Top left is QTL map based on MQM analysis with additional QTL identified. Four QTL were identified on LG01, LG04, LG07 and LG10. QTL on LG04 and LG10 were not previously identified in CIM. Top right and bottom left graphs show the epistatic interaction between the QTL on LG01 and LG10. When the QTL on LG01 is fixed for Tx7000 (parent B) and LG10 is fixed for BTx642 (parent A), an epistatic interaction is observed.

BTx623 x IS3620c

BTx623 x IS3620c was planted in College Station fields in both 2012 and 2013. Due to pest issues in 2012, grain trait phenotypes could not be collected although branching traits were analyzed. In 2013, grain trait phenotypes were collected and used for QTL analysis.

Figure 14 shows all QTL identified from the BTx623 x IS3620c population. In CS (2013), the same QTL identified in BTx642 xTx7000 on LG01 was also found in this population but only in grain number and grain yield, not grain weight. Another QTL on LG01 for grain yield was identified. This QTL was not observed when mapping QTL for grain weight or grain number but was observed in total node number. A QTL on LG03 was identified for total primary branch number, total floral node number, and secondary branch number on node 3 indicating this QTL plays an important role in modulating panicle architecture in the BTx623 x IS3620c population. The parent contributing to an increase in panicle branching is BTx623. Detailed information regarding the position of these QTL is found in Table 3.

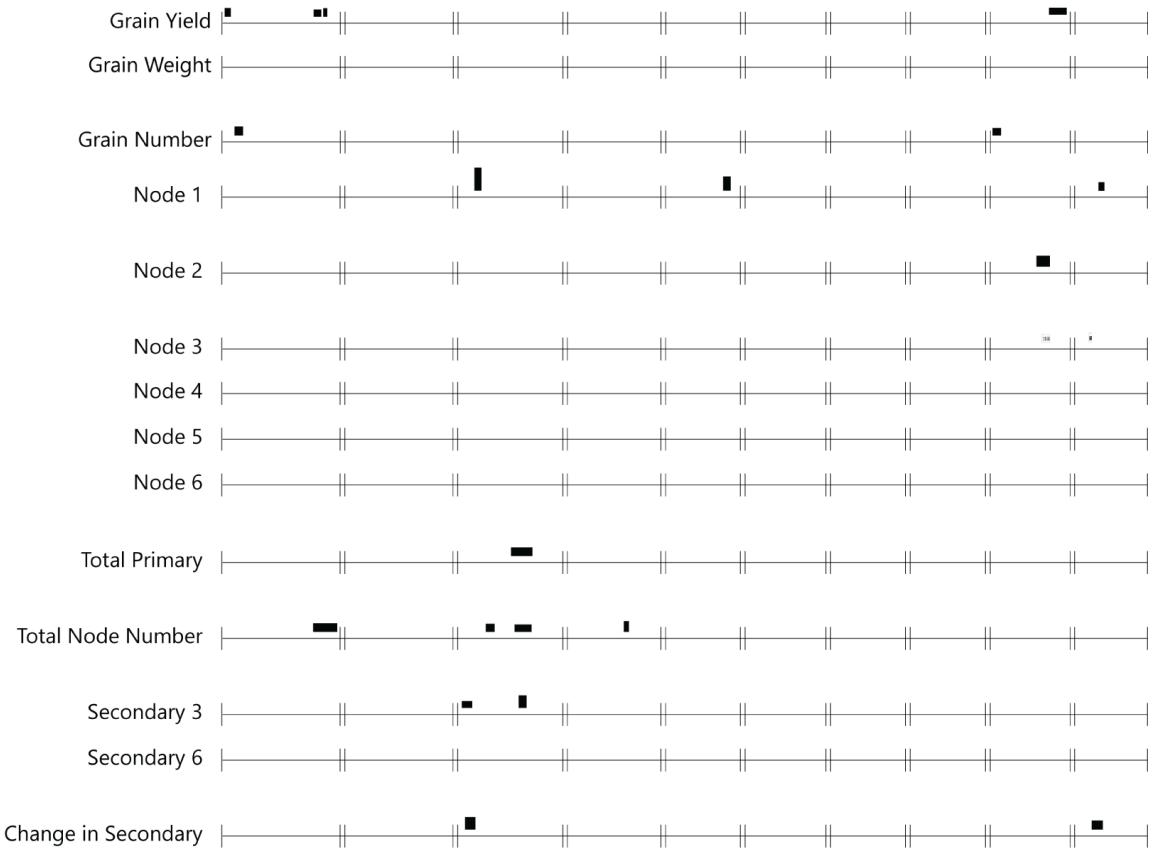

Figure 14**.** All QTL identified in grain traits and branching traits for the BTx623 x IS3620c RIL population. Results for the grain traits are found in the 2013 College Station population and the branching phenotypes are from the 2012 College Station population. The height of the box representing a QTL corresponds to its relative LOD score while the width of each box indicates the 1 LOD interval of the identified QTL. For the grain traits data, a QTL is identified on LG01 in both grain yield and grain number. In branching, a QTL on LG03 shows up consistently in total primary branch number, total node number, and total number of secondary branches off of node 3 primary branches.

Table 3. Information describing QTL identified for grain yield, grain weight, grain number, and panicle architecture through analysis of the BTx623 x IS3620c RILs from College Station 2012 (panicle architecture) and College Station 2013 (grain yield data). Chr. indicates chromosome, R^2 is the amount of the additive variance that is explained and Add Var. is the additive variance.

Trait	Chr	R2	AddVar	cM Peak	bp peak	cM left	cM right	bp left	bp right
Grain Yield	1	0.073011	2.013	7.01	1,824,846	6	8.2	1,591,175	2,058,866
Grain Yield	1	0.061273	1.7714	133.31	65,604,466	132.7	135.4	65,460,202	66,627,150
Grain Yield	1	0.06083	1.7723	144.41	68,300,000	143.2	144.6	68,069,506	68,348,460
Grain Yield	9	0.065597	-1.9073	103.81	58,819,314	98.9	105.1	56,136,496	59,016,935
Grain Number	1	0.08463	89.461	22.71	6,998,016	21.2	24.2	6,217,711	7,409,871
Grain Number	9	0.06654	70.8431	7.91	1,204,881	3	11.3	709,731	1,660,043
Node 1	3	0.321308	7.7859	26.11	5,817,827	23.7	26.8	4,910,642	5,858,349
Node 1	5	0.175884	6.0558	83.51	58,288,740	81.2	84.7	58,065,375	58,604,528
Node 1	10	0.10144	4.3573	53.41	12,743,989	51.9	54.7	9,979,434	45,710,415
Node 2	9	0.205362	-0.7603	87.41	53,5000,000	84.6	90.8	53,050,536	55,646,769
Node 3	9	0.139297	-0.5967	88.41	54,000,000	88.3	92.3	53,139,893	55,646,769
Node 3	10	0.155854	-0.6403	33.01	5,105,189	31.9	33.3	5,375,742	5,537,680
Total Node Number	1	0.13778	0.6455	151.91	69,596,809	145.6	156.8	68,636,828	71,099,652
Total Node Number	3	0.135151	0.6275	49.01	10,736,950	45.5	49.6	9,784,220	11,158,078
Total Node Number	3	0.12049	0.6185	95.01	60,589,079	91.9	99.8	59,753,655	61,872,366
Total Node Number	4	0.183107	-0.8055	84.91	55,031,052	82.9	85.2	54,318,003	55,243,393
Secondary 3	3	0.120864	0.6915	8.51	2,492,203	8.1	12.9	2,450,537	3,198,966

Table 3. Continued

Trait	Chr	R2	AddVar	cM Peak	bp peak	cM left	cM right	bp left	bp right
Secondary 3	3	0.216434	0.9786	91.71	59,714,172	91.3	95	59,320,166	60,614,760
Secondary Change	3	0.182662	0.7251	19.41	4,469,669	17.5	22.4	4,120,278	4,760,359
Secondary Change	10	0.15168	0.673	43.91	7,704,387	41.7	46.8	7,437,970	8,540,325

SC170 x M35

Many QTL were identified for panicle length in the SC170 x M35-1 population, including a QTL on LG03 that co-aligns with a QTL in the branching data of BTx623 x IS3620c. Another panicle length QTL on LG07 was found to co-align with the QTL found in BTx642 x Tx7000 for grain number. Additional panicle length QTL were identified on LG02 and LG06. This result is shown in Figure 15. Grain weight QTL were found on LG01, LG06, and LG09. The QTL on LG01 co-aligns with the known position of *Ehd1* while the QTL on LG09 co-aligns with a QTL identified in thousand grain weight in the BTx642 x Tx7000 population. The remaining QTL that were identified do not co-align with any previously identified QTL in this study. This result is shown in Figure 16. Detailed information regarding these QTL is outlined in Table 4.

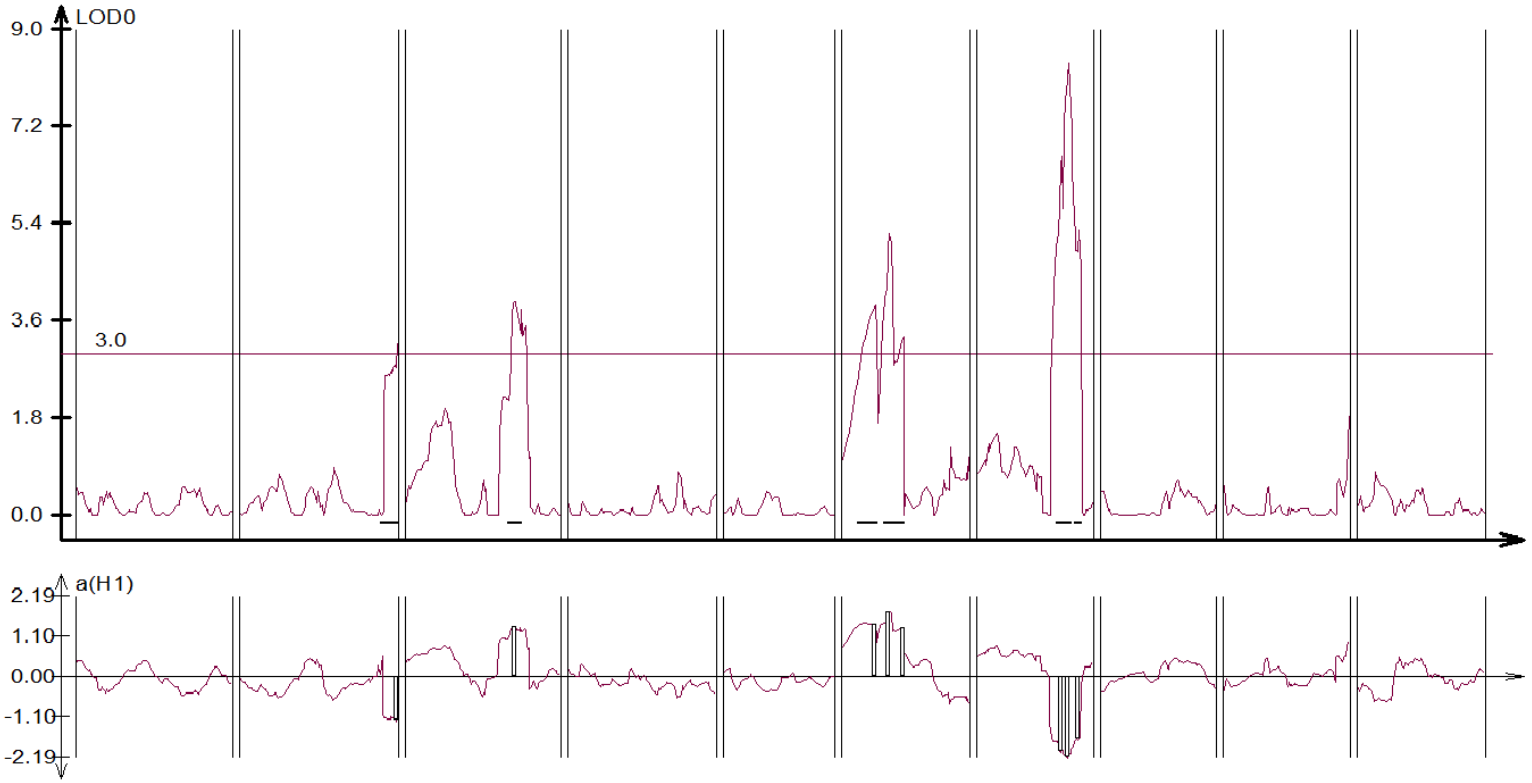

Figure 15. A QTL map showing significant peaks of panicle length in SC170 x M35-1 2011 College Station. The additive graph (lower panel) is also displayed indicating parental influence at a given loci. Positive is M35-1 and Negative is SC170. Four QTL were identified and these QTL were on LG02, LG03, LG06 and LG07.

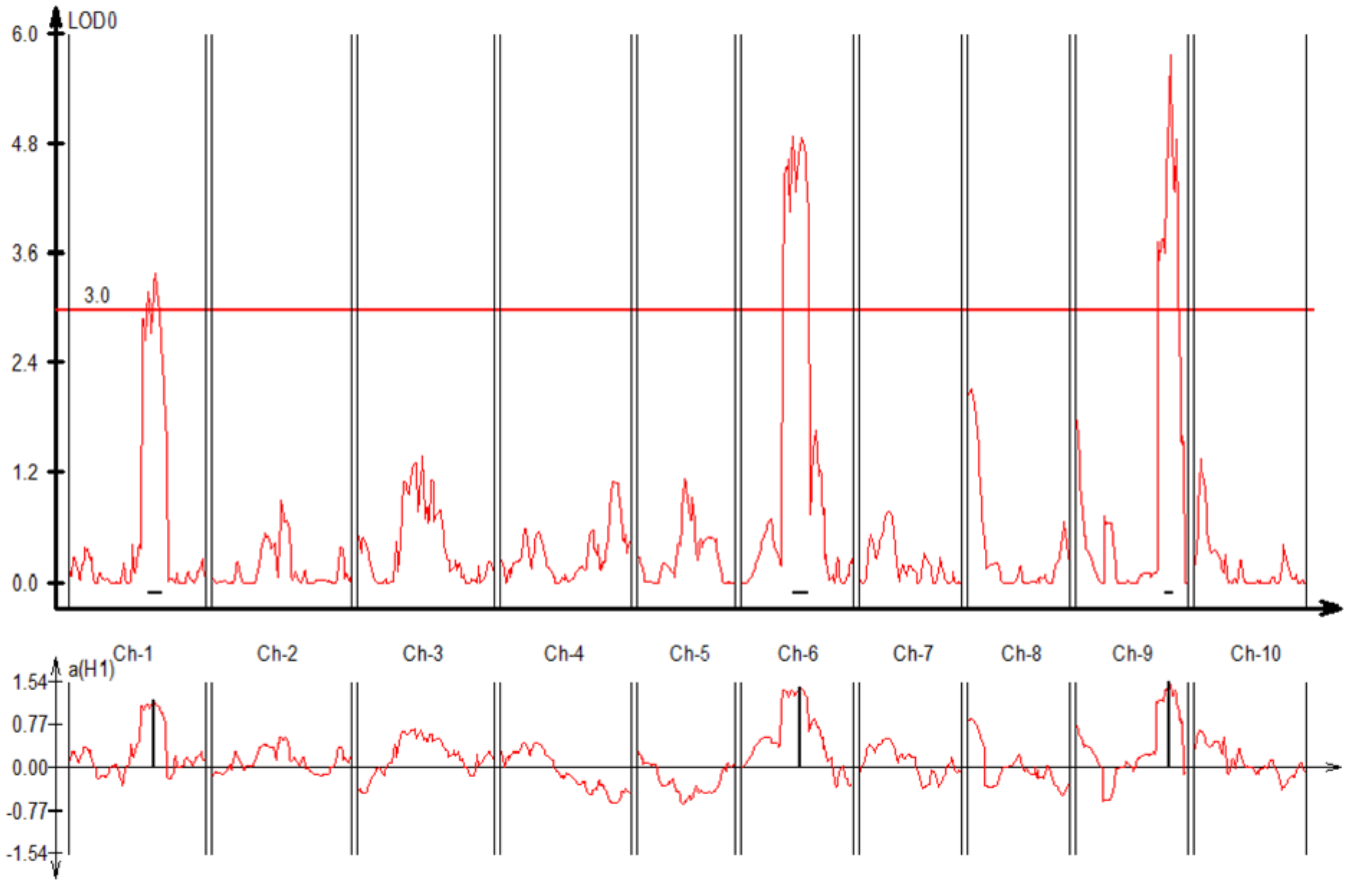

Figure 16. A QTL map showing significant peaks of grain weight in SC170 x M35-1 2011 College Station. The additive graph (lower panel) is also displayed indicating parental influence at a given loci. Positive is M35-1 and Negative is SC170. Three QTL were identified and these QTL were on LG01, LG06, and LG09.

Table 4. Information describing QTL identified for grain composition through analysis of the College Station SC170 x M35-1 population. Chr. indicates chromosome, R^2 is the amount of the additive variance that is explained and Add Var. is the additive variance.

Trait	Chr.	R2	Add. Var.	Peak (cM)	Peak (bp)	cM Left	cM Right	bp Left	bp Right
Panicle Length	2	0.0813	-1.204	118.4	77,206,520.00	108.4	119.4	74,500,000.00	77,619,949.00
Panicle Length	3	0.0992	1.3374	82.3	62,350,000.00	78.8	86.2	60,750,000.00	65,450,000.00
Panicle Length	6	0.1684	1.7345	35.7	4,000,000.00	33.3	37.7	3,300,000.00	6,000,000.00
Panicle Length	7	0.6281	-2.1946	69.1	60,400,000.00	61.8	78.1	58,450,000.00	62,516,000.00
1000 GW	1	0.1064	1.1993	74.11	55,200,000.00	69.8	79.3	54,600,000.00	56,875,000.00
1000 GW	6	0.1597	1.4174	51.61	42,500,000.00	46.3	57.3	39,106,180.00	46,250,000.00
1000 GW	9	0.1876	1.5349	79.91	55,247,312.00	78	82.2	54,900,000.00	55,800,000.00

Standard Broomcorn x SC170

Due to Standard Broomcorn's long primary branches and lack of secondary branching, the Standard Broomcorn x SC170 shows significant phenotypic differences in panicle architecture compared to other sorghum populations used in this study. This population was utilized to study the effect of the presence of secondary branching on the panicle. Broomcorns produce small seed and few in number (153). There was not a correlation between grain yield and secondary branching identified in this particular population; however, the distinct phenotype of no secondary branching does provide further insight into the level of variation in sorghum panicle architecture and the wide range of genetic variation in sorghum.

Discussion

Grain Trait QTL on LG01

The tradeoff between grain weight and grain number is clearly illustrated in results of the QTL analysis of the BTx642 x Tx7000 RIL population grown in College Station (2012). A QTL affecting grain weight was identified on LG01 with the allele for increased grain weight coming from the BTx642 parent. A QTL for grain number on LG01 was aligned with the QTL for grain weight; however, the allele increasing grain number was from the Tx7000 parent. The lack of a corresponding QTL for grain yield at this location on LG01 suggests that the QTL for grain number is causing a compensating change in grain weight since the causative alleles for grain weight and grain number came from opposite

parental lines and acted in opposition to one another; thus negating their impact on overall grain yield. The grain number/weight QTL on LG01 more consistently affected grain number than grain weight across years and environments suggesting stronger genetic control for the grain number at this locus. Therefore it may be beneficial to fix the allele for increased grain number in grain crop genotypes. The QTL on LG01 co-aligns with a QTL identified in the NIR data in Chapter III as well as a previously identified QTL for panicle shape (154). I hypothesize that these results could be generated by alleles of a single gene that affects grain number by anthesis and this change in grain number result in changes in grain weight that occur post anthesis. Alternatively, the results could be generated by closely linked genes that modulate grain number and grain weight respectively.

In order to identify the gene underlying this QTL, heterogeneous inbred families (HIFs) would need to be created that were heterozygous across this region. Depending on the environment and year, the QTL spans 600kb to 1.2 Mb. Through the analysis of breakpoints through phenotyping and genotyping cycles, the causative gene(s) could be identified.

Panicle architecture sets the maximum number of grain the panicle can produce, therefore branching traits can impact grain number. Secondary branching can increase grain number without increasing the number of floral nodes or length of the panicle. When the BTx642 x Tx7000 population was grown in Lubbock (2012) the amount of secondary branching was modulated by

a QTL on LG01 that aligned with the QTL for grain number. The identification of a secondary branching QTL on LG01 in Lubbock (2012) indicates that the grain number QTL on LG01 could result, in part, from variation in the extent of secondary branching.

In Lubbock (2012), irrigation was stopped 14 days before anthesis and the population was subjected to a typical post-anthesis water deficit treatment. Under these conditions the grain number QTL on LG01 aligned with a grain yield QTL rather than a QTL for grain weight that was observed under water sufficient conditions in CS (2012). Therefore, the environment in Lubbock (2012) was favorable during most of the booting phase when grain number is established which probably explains why the QTL for grain number on LG01 was detected. However, after anthesis, water limitation likely reduced photosynthesis and nitrogen assimilation. The corresponding reduction in carbon and nitrogen status would be expected to reduce grain filling. Based on this, I propose that the water limitation reduced grain filling to a sufficient extent that prevented expression of the tradeoff between grain number and grain weight that was observed under non water-limiting conditions. Alternatively, it is possible that there are several genes in this region of LG01 that have independent effects on grain number, weight and yield depending on environmental factors such as water limitation.

The QTL for grain number on LG01 identified through analysis of the BTx642 x T7000 population co-aligns with the location of *short panicle 1* ortholog (SP1) (154). SP1 is part of a peptide transporter family known to

regulate panicle morphology that in turn modulates grain yield (155). Map-based gene identification will be required to more fully understand the action or actions of genes/alleles underlying the QTL for grain number/weight on LG01.

Several populations with different genetic backgrounds were used to analyze QTL for panicle architecture, grain number and grain weight to increase understanding of QTL that modulate these traits. In addition to the BTx642 x Tx7000 population, the BTx623 x IS3620c population was analyzed for grain traits and branching traits. The QTL on LG01 that was observed to have a tradeoff between number and weight in the BTx642 x Tx7000 population was also found in this population. The BTx623 x IS3620c identified this QTL in overall grain yield and grain number, not grain weight. This eliminated the observed tradeoff indicating the QTL on LG01 is most likely two distinct QTL. This is an excellent example of the benefit of looking at traits across populations because without the information from the second population, it would be difficult to hypothesize if these two QTL are due to the same allele or due to two distinct genes located within the same region. However, with the additional insight provided by multiple populations, it is possible to hypothesize that the grain weight and grain number QTL on LG01 are two different QTL as the opposing action was not observed in all populations.

MQM

MQM analysis could provide additional insight into why the grain number QTL did not correspond with the panicle architecture results through identification of additional QTL that might align with panicle architecture traits. This would suggest that panicle architecture is affecting grain number, but is not the largest contributor to the variation found. The MQM data results reported a model for the College Station (2012) grain number trait that, out of the four identified grain number QTL, only the QTL on LG07 co-aligned with a branching QTL. This QTL is seen in the number of primary branches on upper panicle nodes. This result raises questions as to what other subcomponents of grain number are causing these additional QTL. It has been well documented that not all florets the panicle architecture supports become seeds (113). Many factors, including resource limitations and lodging issues, play a role in how many florets fully develop to grain maturity. If panicle architecture does not show a large contribution to the overall grain number, the percent fertilization would be the next phenotype to explore as the main modulator to grain number.

There are several QTL that show up consistently within a single population across multiple years, but not across populations. It is logical that not all QTL would be seen across all populations as not all populations have segregating alleles between the two parents when the population is constructed. The populations analyzed do not have the same allelic variants; therefore a QTL identified may not appear at the same location across multiple populations. It is

also possible that if allelic variants at the location of the QTL are present in all populations, there is potentially another required allele in the pathway this is not consistent across locations. This makes the epistatic interaction models found in the MQM analysis instrumental to pathway identification.

Environmental factors also play a significant role in determining grain yield; therefore, it is important to consider environmental conditions when analyzing the differences in QTL appearance across and within populations. For example, the BTx642 x Tx7000 grown in Lubbock, a drought simulation environment, compared to the same population grown in College Station, a more temperate environment, provides a different perspective on the same genetic information under different stresses. Further analysis of phenotypes of interest across different populations enables the larger picture grain yield QTL to be observed instead of one particular genetic background and environment.

Secondary Branching QTL Analysis

Another notable QTL is present on LG10 that appears consistently in branching data in the BTx642 x Tx7000 Lubbock population but not even in the other BTx642 x Tx7000 populations. This is not surprising since, as was mentioned earlier, Lubbock and College Station present different environments. In 2013 College Station, the BTx642 x Tx7000 population was not irrigated to attempt to simulate drought conditions in the College Station fields; however, even this population did not capture the QTL on LG10. This suggests that the QTL on LG10 only appears in a more extreme water deficit environment than the

2013 CS population, or another environmental factor more specific to Lubbock played a role in the presence of this QTL. Additional field analysis in other environments and years would aid in answering these questions.

In addition to QTL analysis, the study of standard broomcorn and notes from the BTx742 x Tx7000 Lubbock population showed a morphological change to the panicle in terms of secondary branching. Standard broomcorn has no secondary branching in the panicle and the BTx642 x Tx7000 Lubbock population showed reduced secondary branching. The potential hypothesis to these morphological differences between the two populations is different for each population. Standard broomcorn has been bred to have long branches that are used to make brooms. Thus, there has been significant selection against further branching. However, in the BTx642 x Tx7000 Lubbock population, environmental factors changed the panicle architecture from what was observed in the College Station field of the same population. By limiting the number of potential seed sites, there is a better chance of survival for the seed that are formed, thereby increasing the overall fitness of the plant.

Branching is an interesting trait to break down into distinct phenotypes as there are many different subcomponents of panicle architecture that may be explored. Lower nodes on the rachis were observed to have a higher number of QTL than the nodes closer to the top of the rachis for number of primary branches at each node. As there are a higher number of primary branches present on the lower nodes, this may be a compounding factor of primary branch

number. Alternatively, the higher number of QTL on lower nodes of the rachis may be indicative of greater potential for variation and genetic control in the lower half of the panicle. However, total primary branch number, floral node number, and some quantification of secondary branching provide a good overall presentation of the QTL identified through panicle architecture analysis. In a comparison of population panicle morphology, the BTx623 x IS3620c branching phenotypes had significantly more variation than what was found in BTx642 x Tx7000. Despite this increased variation, there was not a large difference in the number of QTL identified for branching. In fact, there was overlap between traits and QTL identified in these populations. The QTL on LG03 appears in total secondary branching in both populations, indicating its importance in the determination of these phenotypes across current sorghum populations and less domesticated lines. This QTL for secondary branching is also identified in the 2012 BTx623 X IS3620c CS population; however, QTL were also identified at the same location for total node number and primary branching in this population. This suggests the secondary branching QTL on LG03 potentially plays a larger role in the variation of panicle morphology within the BTx623 x IS3620c population than the BTx642 x Tx7000 population.

SC170 x M35-1 QTL Analysis

As the SC170 x M35-1 2011 College Station population was the first population used to study grain yield, only panicle length and thousand grain weight were investigated. Two of the four QTL identified within this population co-aligned with QTL found in other populations. These QTL are the ones on LG03 and LG07 in the BTx623 x IS3620c and BTx642 x Tx7000 populations respectively. As informative QTL were identified, the population was planted in Halfway to study grain traits under a different environment. Unfortunately, no QTL were identified in the Halfway trial for this population. The most likely explanation for no QTL being identified is to lack of thinning to a standard density rather than lack of variation in grain yield. Thinning to a standard density allows for many other compounding environmental affects which influence the amount of grain produced to be controlled. When thinning is not performed, the genetic control is still present; however, the effects are masked due to these other factors.

Conclusion and Future Directions

Many QTL and observations of interest that contribute to the variation of overall grain yield were identified in this study of four grain sorghum populations. Through these analyses, the basis is established so the biochemical pathways of grain yield in sorghum can be better understood. MQM provided information on additional QTL and interactions in the BTx642 x Tx7000 College Station 2012 RIL population. The interaction term in the best fit model that was identified by

MQM was between the QTL on LG01 and LG10. These interactions can provide further insight into the grain yield pathway as the genes underlying these QTL are identified. Further analyses would include fine mapping of QTL identified here that play a large role in grain yield determination and further exploration of epistatic interactions via MQM analysis.

Materials and Methods

Grain yield was quantified on a per plant basis. By observing grain yield on a per plant basis in fields that have been thinned to a standard plant density, factors such as plant density per se and tillering, which changes plant density, were minimized as modifiers of grain yield per plant. The exception to thinning is the SC170xM35-1 population that was planted in Halfway. This population was not thinned. All populations that are reported here were field trials from 2011 to 2014 in College Station, Halfway, and Lubbock. All populations, except the 2013 College Station and 2012 Lubbock populations of BTx642 x Tx700 population, were irrigated. Grain sorghum genotypes grown at typical density usually produce one panicle per plant. Grain trait data was collected on 5 panicles per plot (genotype) using a randomized replicated planting design. As a result, 2 sets of 5 panicles were analyzed for each genotype. Panicles were not collected from plants at the end of plots because these plants intercept more light and have greater access to water compared to plants within rows.

To measure total grain yield, all grain from one panicle was removed from the panicle and weighed. This provided total grain yield on a per plant basis so

variation amongst and across different genotypes could be observed. The standard for quantifying the individual weight per grain produced by a plant is to measure the weight of 1000 grains. A seed counter was utilized to count 1000 grains that were subsequently weighed.

Grain number can be quantified by removing all grain from a panicle and counting the seeds. Conversely, grain number per panicle can be calculated from the grain yield per panicle and the 1000 grain weight using equation 1. Grain number was quantified in 10 plants and compared to the calculated value. There was less than a 3% variation between observed and calculated values for grain number; therefore, following this experiment, it was determined that grain number per panicle could be obtained by calculating the values using equation 1.

For panicle architecture traits, the panicle length was measured from collar to the end of the panicle. The floral node number was determined by counting the number of circular groupings of primary branches. The number of primary branches per node was counted as well as the total number of primary branches. The secondary branch number was counted on 3 primary branches on either node 3 or node 6. Total branch length was calculated by randomly selecting 5 branches to measure. For secondary branching in Standard Broomcorn, observations were made in a qualitative way of there being secondary branching present or absent from the panicle.

BTx642 x Tx7000 and BTx623 x IS3620c are both RIL populations meaning that these populations have been selfed to the point that the heterozygosity in these populations has been eliminated. SC170 x M35-1 were in the F5 indicating that heterozygosity levels should be at approximately 6.25%. All QTL analyses were run under the assumptions of the populations being RILs as this best matched level of heterozygosity that was present. The Standard Broomcorn x SC170 population was in the F2s at the point of analysis. In this analysis, a quantitative approach was done to see the presence or lack thereof of secondary branching. No QTL analysis was done since this was a qualitative trait rather than quantitative so could not be mapped. Due to aphids and other pests in the field of the F3s, a further study of this population was not able to be completed.

Genotyping data for QTL mapping was prepared by Digital Genotyping (DG) and then run on the Illumina HiSeq for all populations (59). For BTx642 x Tx7000, over 10,000 unique markers were identified. Approximately 2,750 informative markers were identified. Informative markers are markers that provide new information to the refinement of QTL due to breakpoint constraints found in a given population. This was a necessary step for this population in order to perform the MQM analysis in a reasonable amount of time as the 10,000 permutation calculations take a significant amount of time for each additional marker so eliminating markers that were not informative saved time in

the analysis. SC170 x M35-1 had ~600 markers and BTx623 x IS3620c had ~850 markers in the maps used for QTL mapping.

All QTL mapping was done through QTL Cartographer (58). MQM was done via the R/qtl package (61, 156). All populations were run as RILs with the Kosambi mapping function. BTx642 x Tx7000 has ~90 RILs in its population. BTx623 x IS3620c has ~400 RILs in its population and SC170 x M35-1 F5s has ~200 lines. All lines were phenotyped for each trait 10x as described above. The only exception is branching of the BTx623 x IS3620c where a subset of the population (~150 RILS) was phenotyped rather than all 400+ lines and all branching was done for each plot in triplicate rather than 5x so the total panicles phenotyped per RIL was 6 instead of 10 for branching traits.

CHAPTER III

THE GENETIC BASIS OF VARIATION IN GRAIN COMPOSITION

Introduction

Sorghum [*Sorghum bicolor* (L.) Moench] is a drought tolerant grain crop native to Africa that serves as an important source of grain, forage, and bioenergy production throughout the world (6, 7). Based on current trends, the global population is expected to exceed nine billion people by the year 2050. To adequately feed a population this size, crop production must increase by 60-110% (8–10, 12). Unfortunately, the amount of arable land is insufficient to meet these requirements without significant increases in grain yield per acre of land used for production. There has previously been significant investment in genetic improvement in the grain yield of many cereal crops such as corn. However, there has been comparatively minimal investment in sorghum breeding for grain yield. There are many genetic and environmental factors that influence the variation of grain yield, making it challenging to identify pathways that influence grain yield. One way to simplify the approach is to analyze subcomponent traits that affect grain yield in order to identify alleles that modify overall grain output.

It has been hypothesized that grain number has a larger genetic contribution than grain weight to grain yield. This hypothesis argues that grain weight is more dependent upon nutrient resources (i.e., sugars from photosynthesis, amino acids from nitrogen assimilation) available during grain filling, than grain number; this hypothesis has been supported by several studies

(31–33, 47, 112, 157, 158). However, an alternative hypothesis postulates that that nutrient limitation also impacts panicle architecture and overall grain number (33). This hypothesis argues that if nutrient limitation occurs during the phase of reproductive development when grain number is established then grain number will be affected whereas nutrient limitation that occurs after fertilization has occurred will be more likely to affect the weight of each grain.

Grain weight is largely determined during the grain filling stage assuming similar grain number per plant. The duration and rate of grain filling greatly affects the weight of the grain. Nutrient availability can affect multiple subcomponents of grain weight. In this chapter, the genetic basis of variation in grain composition, stem biomass, and stem hollowing were analyzed to determine if variation in these traits affects grain weight or grain yield.

Nutrient allocation depends on source activity and sink strength. During seed development, the plant is the source of nutrients (sugars, amino acids, etc.) whereas the developing seed is a sink for these compounds. Once the grain filling stage starts, the “sink strength” of the panicle is given priority over the other sink-related activities of the plant (49–51). This phenomenon results in the remobilization of carbon and nitrogen from vegetative tissues to the grain sometimes resulting in leaf senescence and stem hollowing (2, 159). Environmental factors also play a role in the development of the seed and adverse environments that affect carbon or nitrogen availability can reduce grain weight.

The remobilization of nutrients from the stem to the seed during grain filling can result in a hollowing of the stem. Stem hollowness varies dramatically amongst different sorghum varieties. Sorghum lines that have been selected for sucrose accumulation (sweet sorghum) and biomass usually have solid stems that are tall and large. (2, 3, 5, 160). As stem hollowing would reduce the available biomass, it has been selected against in these lines. In the case of high biomass sorghum the stem is a strong sink for biomass accumulation during the crop's long vegetative phase. The timing of the start of stem hollowing is under genetic control. In some parental lines, stem hollowing begins to occur at the grain filling stage once the plant has reached anthesis. However, in some grain sorghum genotypes, stem hollowing has been found to start around the time of floral induction with a dramatic increase in hollowing once the plant reaches anthesis. This increase in stem hollowing has been found to continue through grain maturity (161, 162). In this chapter, stem hollowing is reported in the BTx642 x Tx7000 RIL population grown in College Station in 2013.

Grain composition has an important impact on human and animal nutrition. Some areas of the world where sorghum is typically grown for human consumption are facing the challenge of malnutrition (163). Identifying QTL that create more nutritious sorghum would potentially be helpful in reducing malnutrition especially if this trait can be improved without reducing yield.

Grain weight varies among cultivars grown in similar environments with thousand grain weight ranging from 30g to 80g per panicle (164). Grain composition also varies significantly among genotypes. In the current study near infrared spectroscopy (NIR) was used to estimate the grain composition (165). The composition traits that were analyzed were protein, moisture, fat, fiber, ash, starch, phenol, 3-deoxyanthocyanidins, and tannins.

Proteins are composed of amino acids that are essential in human and animal diets. Proteins range from 7-15% of the total grain weight of sorghum(166). An increase in grain number typically results in a lower percentage of protein in each individual grain due to resource limitations (167). Identifying QTL for overall grain yield that also increase in overall protein/seed would be beneficial to improving sorghum nutrition.

The fat or lipid content of sorghum grain is typically very low compared to other cereals, making up just 2-4% of total weight (168). In developed countries, low fat foods are an ideal nutrient source (169). However, in sub-Saharan Africa where sorghum is grown as a primary food source, higher fat content could be beneficial to these populations that lack sufficient calorie intake (163, 170).

Fiber, classified as any non-starch polysaccharide, is an important part of dietary intake and balance. Sorghum grains typically contain 1-4% fiber (171). Due to sorghum’s relatively high dietary fiber content, it has been identified as an excellent food source for diabetics because high fiber diets that have been found to increase control of blood sugar levels and decrease insulin levels (172).

Ash is between 1-3% of sorghum grain composition and correlated with protein content (171). Whether this correlation is due to linked genes or one gene is unknown. Starch is utilized for energy storage, and makes up the largest portion of the grain by weight. Starch is a polysaccharide composed of a large number of glucose molecules that may be broken down into individual carbohydrates to provide a long lasting supply of energy (170, 173). Starch comprises 70-80% of the sorghum grain and is primarily located within the endosperm. Starch content is negatively correlated with protein and fat content (171). These are the three key nutrients to the human diet, therefore it would be ideal if there were a way to overcome this tradeoff. There are several biological reasons for this tradeoff, including carbon limitations; however, it is also possible that genetic linkages play a role in the tradeoff between these nutrients. It would be beneficial to break these linkages to improve the overall grain quality.

Phenols play an important role in grain quality that can be a huge asset to overall health of the consumer due to their antioxidant properties and other health benefits (174). 3-Deoxyanthocyanins, common sorghum phenols, are flavonoids that are not often found in cereal crops and are often identified by their black pigmentation phenotype in the grain and surrounding area (175).

Tannins are considered the most important group of flavonoids and have a wide range of concentrations in sorghum lines. Despite the potential health benefits of flavonoids, their bitter taste dictates a balance between potential health benefits and edibility (176).

Results

Three populations were utilized to analyze the genetic basis of variation in grain composition and other traits that could impact grain weight: BTx642 x Tx7000, BTx623 x IS3620c, and SC170 x M35. Basic information about these populations is described in Chapter II.

BTx642 x Tx7000

In Chapter II, QTL that modulate grain yield, grain weight and grain number were identified through analysis of the BTx642 x Tx7000 RIL population grown in College Station (2011-2013) and Lubbock (2012). In this study, QTL that modulate grain weight in this population was analyzed using MQM and data from College Station (2012). This analysis identified 5 additional QTL, shown in Figure 17, generating a total of 7 QTL identified for grain weight. While the MQM analysis identified additional QTL, the QTL on LG03 and LG01 were comprised of two adjacent peaks. Further analysis showed that the ‘split’ peaks were caused by a single RIL with a breakpoint within the region spanning the two adjacent QTL on LG01 and LG03. Assuming the phenotypes and genotypes are accurate, this indicates that there are two QTL present in these regions of the genome that are affecting the trait of interest. However, with only a single breakpoint, there is not enough statistical power to confirm whether the peak splitting results are caused by allelic variation in two different genes in each locus or if the data derived from the single RIL is inaccurate via an error in the phenotype or genotype data. To determine which hypothesis is correct, further

analysis of this RIL population and fine mapping of the region will be required. Therefore, for the purposes of this study, the regions on LG01 and LG03 spanning the 'split' QTL peaks will be called one single QTL resulting in 5 QTL identified via MQM for grain weight.

CIM analysis carried out in Chapter II also identified the QTL for grain weight on LG01 at 6.6 cM and LG03. MQM analysis identified additional QTL on LG01, LG06, and LG08. The best-fit Anova model includes seven QTL and explains 82.5% of the total variance found in the grain weight phenotype.

Table 5. MQM model results for grain weight. Seven QTL were identified explaining 82.5% of the variance found in the BTx642 x Tx7000 RIL population grown in College Station in 2012.

Model Formula: y~Q1+Q2+Q3+Q4+Q5+Q6+Q7							
	df	**SS**	**MS**	**LOD**	**% var.**	**P-value(Chi^2)**	**P-value(F)**
Model	7	2603.525	371.9322	34.51942	82.56847	0	0
Error	83	549.646	6.622241				
Total	90	3153.171					

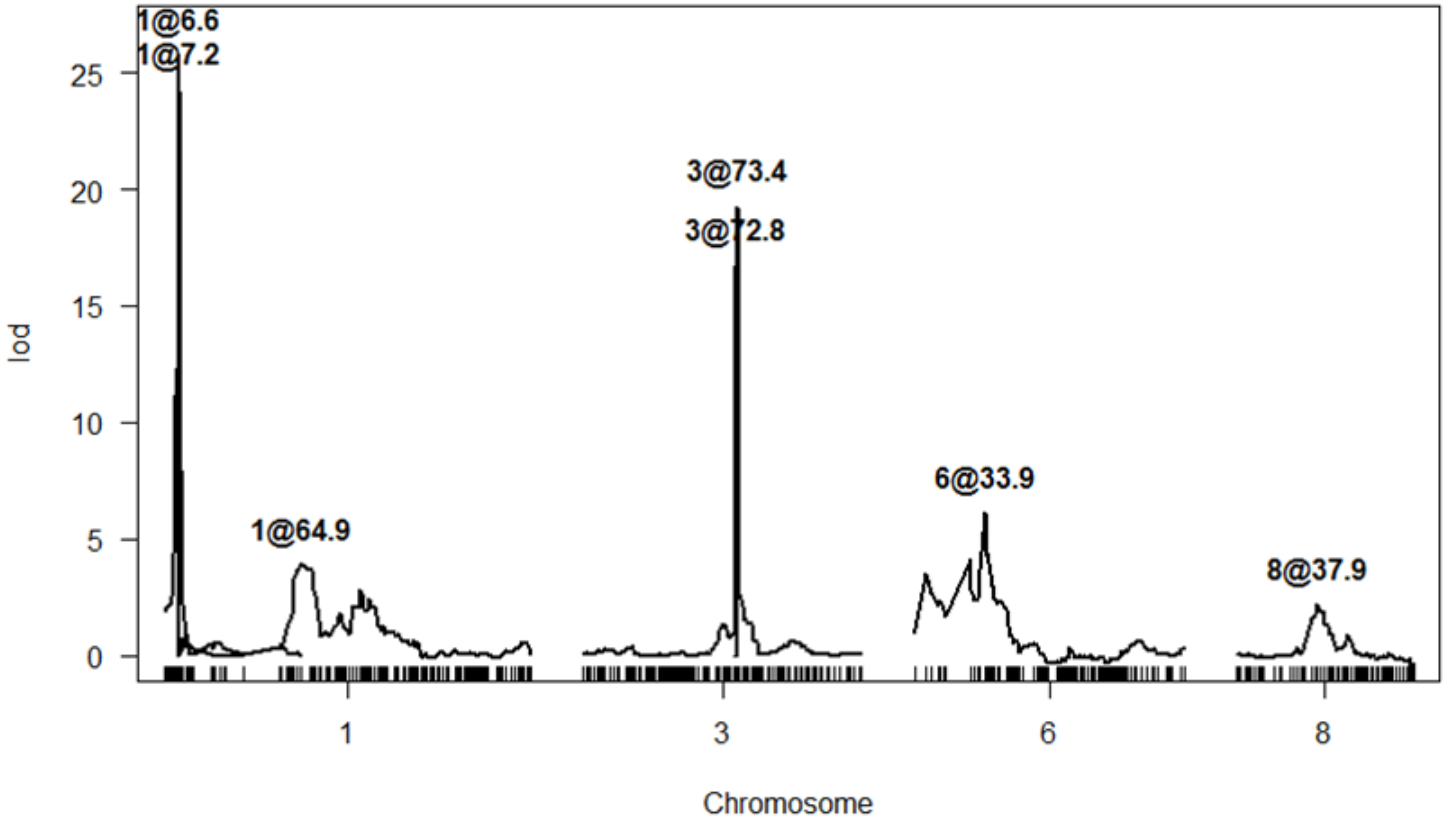

Figure 17. MQM analysis of QTL for grain weight using data derived from BTx642 x Tx7000 population grown in College Station (2012). QTL were identified on LG01 at 6.6cM and 64.9cM, LG03, LG06, and LG08.

Grain Composition

A QTL on LG03 (LOD = 5.2) for grain weight and grain yield was identified in the BTx642 x Tx7000 RIL population based on field data collected in College Station in the 2011 and 2012 (Figure 11). No corresponding QTL for grain number was identified in this region of the genome. The increase in weight per grain could be due to greater availability of nutrients (i.e., sugars, amino acids) during the grain filling phase, longer duration of grain filling, or a difference in the rate of grain filling. Differences in the rate of grain filling could be due to higher rates of starch accumulation that would result in variation in seed composition. Nutrient supply can be affected by differences in photosynthesis, nitrogen assimilation and remobilization of nutrients from leaves and stems during the grain-filling phase. Variation in remobilization could cause differences in stem composition and the extent of stem aerenchyma formation or hollowing at grain maturity.

Grain composition, stem hollowing, and stem composition were quantified to determine if these traits could potentially be influencing grain weight. Figure 18 depicts QTL that modify grain composition estimated via NIR. The QTL for protein, moisture, fat, fiber, starch, phenol, and tannins align with the grain weight and grain yield QTL identified on LG03.

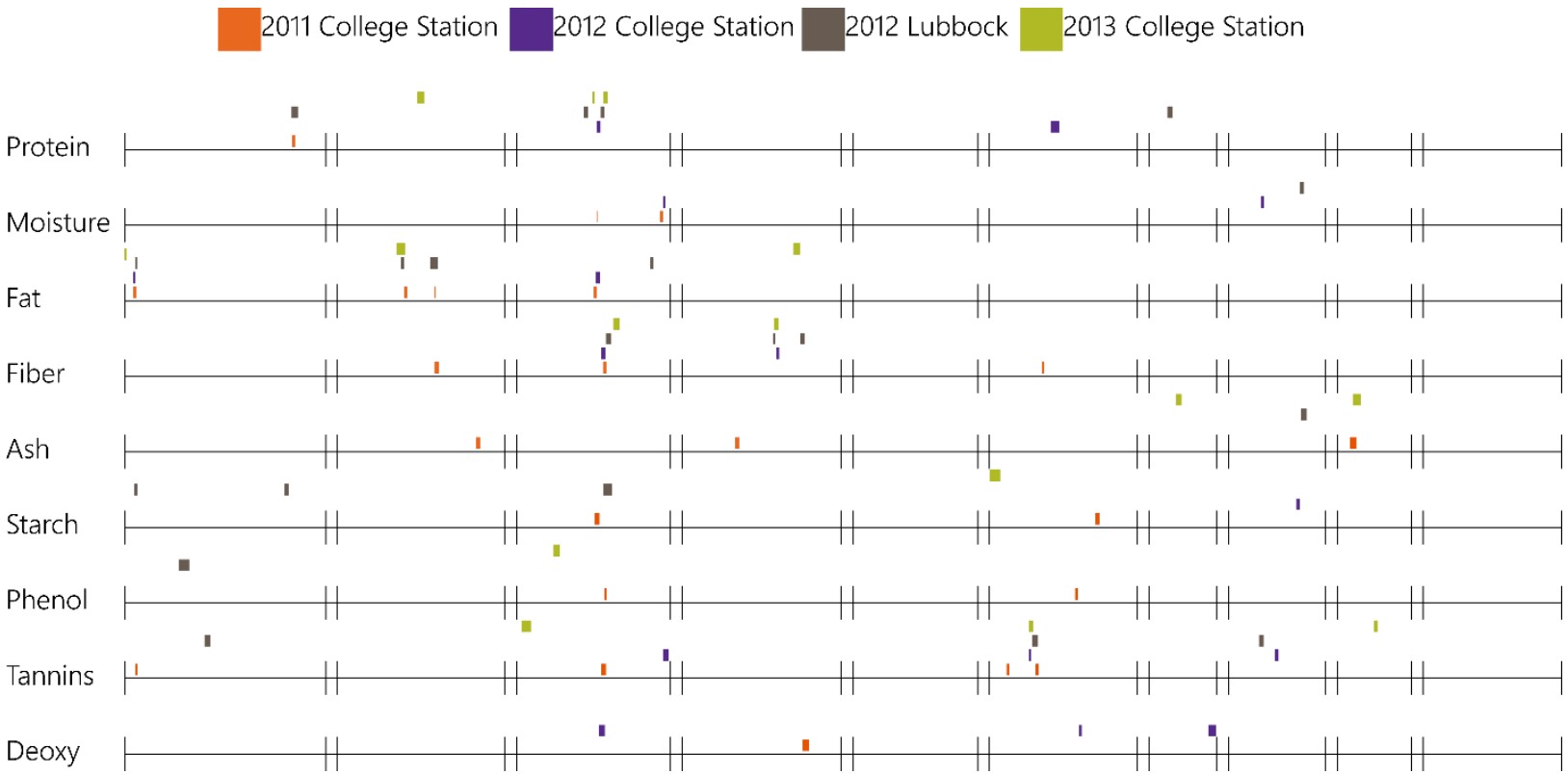

Figure 18. All QTL identified in grain composition analysis via NIR for the BTx642 x Tx7000 population. Results are color coded by the year and environment the population was grown. The height of the box representing a QTL corresponds to its relative LOD score while the width of each box indicates the 1 LOD interval of the identified QTL. The QTL on LG03 that is reported consistently across traits co-aligns with the previously identified grain weight QTL. Other QTL of note are on LG01 in fat that co-aligns with another grain weight QTL that was previously identified and QTL on LG06 and LG08 that co-align with MQM grain weight identified QTL.

Table 6. Information describing QTL identified for grain composition through analysis of the 2011 College Station BTx642 x Tx7000 RIL population. Chr. indicates chromosome, R2 is the amount of the additive variance that is explained and Add. is the additive variance. Phenotyping was performed via NIR spectroscopy.

Population	Trait	Chr.	Add.	R2	Peak (cM)	Peak (bp)	Left (cM)	Right (cM)	Left (bp)	Right (bp)
2011 CS	Protein	1	0.2404	0.1197	145.91	72591532	145	147.8	72331973	73098980
	Moisture	3	0.1541	0.1806	69.81	54663551	69.8	70.4	54404000	54819702
	Moisture	3	-0.1176	0.1007	125.71	72030091	124.1	127.1	71320639	72383129
	Fat	1	-0.1473	0.1544	7.81	3420389	7.2	10	3362675	4302055
	Fat	2	-0.1162	0.0814	58.61	54481016	57.8	60.2	53941313	55135152
	Fat	2	0.1712	0.1654	84.51	61204394	83.9	85.3	60659770	61504658
	Fat	3	-0.1353	0.1218	68.11	53636312	66.9	69.1	53308040	53917180
	Fiber	2	0.0433	0.1135	84.51	61204394	83.9	87.9	60659770	61924719
	Fiber	3	0.0661	0.2728	76.11	57147750	74.8	77.4	57040676	57244330
	Fiber	6	-0.0361	0.0818	46.41	43347513	45.8	47.4	43277496	43430387
	Ash	2	-0.0119	0.1177	121.91	71054674	119.7	123.6	70568207	71758736
	Ash	4	0.0094	0.0801	47.11	10061141	46	49.4	9697028	10305734
	Ash	5	-0.0113	0.1136	43.31	6919360	41.3	45.1	6919360	9432519
	Ash	9	-0.0132	0.1537	13.61	2228497	11	16.1	2128527	2449105
	Starch	3	-0.3196	0.2315	69.81	54663551	67.5	71.7	53428461	55904211
	Starch	6	0.2906	0.1777	93.71	53863027	91.5	95.6	53386932	53977646
	Phenol	3	-0.4532	0.1647	76.11	57147750	75.8	78.1	57040676	57298200
	Phenol	6	-0.4027	0.1371	75.91	49691842	74.3	76.5	48682099	50118603
	Tannins	1	1.1286	0.1015	10.61	5666314	9.1	11.2	4033307	5949209
	Tannins	3	-1.003	0.0847	76.11	57147750	73.4	77.3	56411014	57213858
	Tannins	6	-1.5611	0.1977	40.51	42116585	40	42.3	42116399	42690597

Table 6. Continued.

Population	Trait	Chr.	Add.	R2	Peak (cM)	Peak (bp)	Left (cM)	Right (cM)	Left (bp)	Right (bp)
2011 CS	Deoxy	4	-3.8139	0.1322	113.11	63491837	109.5	115.1	62898338	63597714
2012 CS	Protein	3	-0.3329	0.2793	72.21	55904211	71	73.9	54819702	56480518
	Protein	6	0.2819	0.1144	56.41	44382825	53.4	60.5	44050445	45491397
	Moisture	3	-0.4486	0.2277	127.41	72507817	126.6	128.5	72096240	72655192
	Moisture	8	0.3703	0.1380	29.21	5320802	28.2	30.2	4727743	5512414
	Fat	1	-0.1683	0.1427	8.31	3539934	7.6	9.4	3362675	4302055
	Fat	3	-0.2039	0.2157	69.81	54663551	68.2	72.2	53636312	55904211
	Fiber	3	0.0491	0.1922	74.61	57040676	73.3	76.8	56129399	57210048
	Fiber	4	0.0487	0.1882	82.91	56632355	81.5	84	55264139	56726882
	Starch	8	-0.3636	0.1309	60.11	57275212	58.4	61.2	57058747	57450260
	Tannins	3	1.7653	0.1814	128.51	72655192	126.7	131.4	72096240	73580488
	Tannins	6	-1.8431	0.2114	35.11	37266064	34.2	35.7	32050494	38113611
	Tannins	8	-1.3245	0.1053	41.41	7708496	39.7	43.1	7558555	50345079
	Deoxy	3	4.0855	0.1488	73.41	56411014	71.3	76.3	55369997	57210048
	Deoxy	6	-3.5884	0.1144	78.81	50972610	77.6	79.8	50801461	51165047
	Deoxy	7	3.4397	0.1080	56.61	53056971	51.7	57.7	8909425	53854257
2012 Lubbock	Protein	1	0.3348	0.1540	147.71	73098988	144.2	149.9	72261564	74409109
	Protein	3	0.3166	0.0836	60.11	51157728	58.2	62.1	16063294	51817379
	Protein	3	-0.3082	0.1248	73.41	56411014	72.7	75.7	56120997	57109343
	Protein	7	-0.3092	0.1349	19.01	2570907	16.2	20.6	2433772	3116461
	Moisture	8	-1.7388	0.1817	62.41	57534912	61.4	64.9	57450260	58176658
	Fat	1	-0.1865	0.2204	10.01	4302055	9.5	10.6	4033307	5666314

Table 6. Continued.

Population	Trait	Chr.	Add.	R2	Peak (cM)	Peak (bp)	Left (cM)	Right (cM)	Left (bp)	Right (bp)
	Fat	2	-0.1381	0.1075	56.41	14796208	54.7	57.8	12575090	53941313
	Fat	2	0.1247	0.0860	84.51	61204394	80.2	86.9	59800156	61924719
	Fat	3	-0.1628	0.1725	117.11	70018532	115.9	118.5	70010936	70191059
	Fiber	3	0.0578	0.2398	79.51	57708488	77.4	81.9	57244330	57915403
	Fiber	4	0.0501	0.1578	80.01	54492089	78.8	80.9	54315892	54956859
	Fiber	4	-0.0399	0.1028	104.91	61812417	102.5	106.5	61463113	62556780
	Ash	8	-0.0649	0.1628	63.51	58036393	62.2	67.5	57453140	58355788
	Starch	1	-0.2246	0.1082	10.01	4302055	8.3	11	3539934	5867148
	Starch	1	-0.2753	0.1607	139.21	70531158	138.3	141.9	70039294	71999152
	Starch	3	-0.2301	0.1111	77.41	57244330	75.2	82.4	57040676	58281325
	Phenols	1	1.9519	0.1349	54.41	15836709	46.8	55.8	13474003	15867215
	Tannins	1	7.7248	0.1287	71.01	22234301	69.1	74	21535461	24801931
	Tannins	6	-7.9304	0.1384	40.01	42116399	37.3	42	40146359	42690597
	Tannins	8	7.1601	0.1236	63.51	58036393	61.4	65.1	57450260	58176658
2013 CS	Protein	2	-0.3023	0.1071	73.01	58905718	68.8	75.1	57585201	59132746
	Protein	3	-0.4425	0.2570	66.91	53308040	65.9	67.2	52431637	53428461
	Protein	3	-0.5545	0.4127	77.41	57244330	75.3	78.4	57040676	57298200
	Fat	1	-0.1838	0.2983	0.01	479688	0	1.4	0	1828060
	Fat	2	-0.116	0.1180	56.41	14796208	51.4	58.6	8450847	54481016
	Fat	4	0.1459	0.1773	97.01	59579024	96.4	102	59519205	60959029
	Fiber	3	0.0573	0.2288	85.81	59601798	83.9	89.1	58846871	60686952
	Fiber	4	0.0469	0.1494	82.21	55592484	79.6	83.4	54466094	56726882
	Ash	7	-0.0117	0.1454	57.71	53854257	53.6	58.3	9732489	55673335

Table 6. Continued.

Population	Trait	Chr.	Add.	R2	Peak (cM)	Peak (bp)	Left (cM)	Right (cM)	Left (bp)	Right (bp)
	Ash	9	-0.0146	0.2209	16.41	2519763	13.4	20.1	2228497	2635934
	Starch	6	0.2858	0.1748	5.51	847299	0.4	9.2	334083	1067340
	Phenols	3	0.4088	0.1510	33.11	7415173	32	37.5	7257188	8167691
	Tannins	3	1.2702	0.1444	10.01	2055161	4.3	12.4	1070703	2185251
	Tannins	6	-1.3132	0.1736	36.31	38113611	34.5	37.9	32050494	40220770
	Tannins	9	-1.1492	0.1226	32.41	5326907	31.5	34.8	5129932	5744810

The QTL for grain weight and grain number on LG01 co-localize with a previously identified QTL for panicle shape (154). However, there are many different potential causative alleles underneath these QTL. Interestingly, a QTL for seed fat content was consistently identified across all years and environments on LG01 also co-aligns with QTL for grain weight and grain number. Additionally, in the 2012 Lubbock population, a starch QTL on LG01 co-aligns with the grain weight and grain number QTL on LG01. Due to its consistency across all years and environments, the QTL for grain fat content on LG01 is of particular interest as it co-aligns with a grain weight QTL that also appeared in all environmental conditions. The starch QTL on LG01 only appeared under water deficit conditions.

The grain weight QTL on LG08 in the 2012 CS environment co-localizes with a tannin content QTL found via MQM analysis. The QTL on LG08 also co-localizes with a QTL for grain ash content found in 2012 Lubbock. The grain weight QTL on LG06 aligned with QTL for phenol and tannins across multiple years.

Stem Composition

The genetic basis of differences in the change in stem biomass composition between the onset of anthesis and grain maturity were analyzed to determine if genetic variation in stem remobilization could be detected. Stem biomass composition at the onset of anthesis and at grain maturity was analyzed. The difference in stem biomass at grain maturity and anthesis

provided a measure of the extent of biomass remobilization from stems to the grain. The results of composite interval mapping are shown in Figure 19. The positions of these QTL are described in Table 6. One consistent QTL on LG02 was identified for variation in stem ash, sucrose, lignins, galactan, arabinan, soluble carbohydrates, ethanol and starch. Two other QTL consistently appeared on LG06 that aligned with QTL for several composition traits. These three QTL aligned with QTL for stem biomass at grain maturity and the change in biomass traits between anthesis and grain maturity. Two additional QTL for variation in stem biomass were consistently identified at anthesis. One of these QTL was located at the beginning of LG03 and aligned with QTL for ash, sucrose, lignin, xylan, soluble carbohydrates, and starch. The second QTL was located on LG09 and aligned with QTL for protein, glucan, galactan, and structural carbohydrates.

The second QTL on LG01 co-aligns with a QTL for stem moisture content was reported at the change in stem biomass composition.

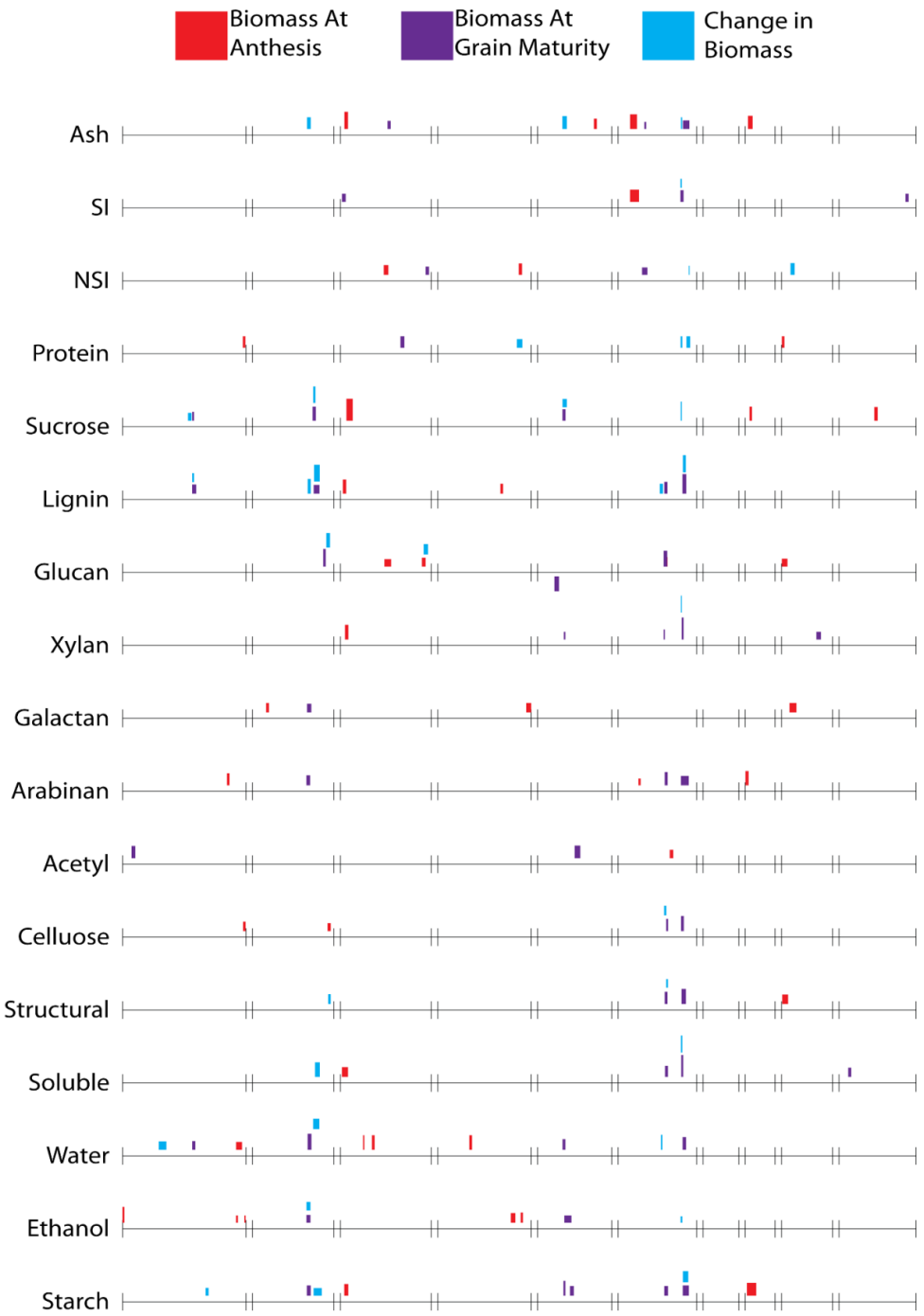

Figure 19. QTL analysis results from NIR analysis of stem composition in the BTx642 x Tx7000 RIL population. The height of the boxes corresponds to the relative LOD score while the width of the boxes corresponds to a 1 LOD interval in cM. Three QTL are identified consistently in stem biomass composition at grain maturity and the change in stem biomass composition from anthesis to grain maturity; one of these QTL is located on LG02 while two QTL are located on LG06. QTL identified on LG03 and LG09 were identified for stem biomass at anthesis.

Table 7. QTL information from the 2013 College Station BTx642 x Tx7000 RIL population for stem composition at grain maturity. Chr. indicate chromosome, R2 is the amount of the additive variance that is explained and Add. Var. indicates the additive variance. Phenotyping was performed via NIR spectroscopy.

Dev. Stage	Trait	Chr.	R^2	Add. Var.	Peak (cM)	Peak (bp)	Left (cM)	Right (cM)	Left (bp)	Right (bp)
Anthesis	Ash	3	0.2353	0.226	9.21	1,991,940	6.2	11.6	1,638,809	2,235,920
	Ash	5	0.131	0.1701	90.11	55,598,516	86.9	91.6	54,592,959	56,607,774
	Ash	6	0.2013	-0.2046	26.51	3,449,360	18.7	29.1	1,032,887	4,711,341
	Ash	8	0.1701	0.1886	9.21	2,133,644	4.1	11.1	679,896	3,300,851
	SI	6	0.2126	-0.185	26.51	3,449,360	18.5	32.2	1,032,887	6,428,223
	NSI	3	0.161	0.1055	69.21	53,322,260	67.3	74.4	51,454,564	55,441,525
	NSI	4	0.1789	0.1096	128.51	64,611,864	125	129.9	64,268,215	65,021,164
	Protein	1	0.1142	-0.2136	188.61	72,435,660	185.3	189.6	71,837,886	72,938,507
	Protein	8	0.1621	0.2542	57.71	49,740,362	56.4	60.3	49,032,189	50,116,222
	Sucrose	3	0.3784	-0.9867	11.21	2,087,154	9.4	19.1	1,991,940	4,103,048
	Sucrose	8	0.1947	-0.6863	9.21	2,133,644	6.5	10.2	679,896	2,449,617
	Sucrose	10	0.1773	0.7941	56.31	10,075,847	54.4	59.8	8,876,673	12,145,020
	Lignin	3	0.1852	0.2413	4.31	1,447,138	3.7	8.8	1,177,165	1,991,940
	Lignin	4	0.1281	-0.2104	98.21	59,064,346	96.5	100.7	58,935,877	60,153,138
	Glucan	3	0.1002	0.323	75.01	55,441,525	67.9	78.2	51,454,564	56,782,235
	Glucan	3	0.1252	-0.3822	127.21	70,597,525	125.8	131.6	70,230,977	71,832,221
	Glucan	6	0.1283	-0.3628	75.01	51,139,358	71.6	76.3	48,740,934	52,094,273
	Glucan	8	0.1073	0.3324	57.71	49,740,362	56.4	65	49,032,189	50,817,616
	Xylan	3	0.2361	0.2457	9.21	1,991,940	7.1	12.3	1,638,809	2,235,920
	Galactan	2	0.1682	-0.0179	22.91	4,899,014	19.4	26.9	4,720,368	5,470,637
	Galactan	4	0.155	-0.0168	139.21	66,691,746	138.5	142	66,646,029	67,116,462
	Galactan	8	0.1591	-0.0169	73.01	51,620,632	68.3	78.8	50,817,616	53,057,572

Table 7. Continued.

Dev. Stage	Trait	Chr.	R^2	Add. Var.	Peak (cM)	Peak (bp)	Left (cM)	Right (cM)	Left (bp)	Right (bp)
Anthesis	Arabinan	1	0.1737	-0.1196	162.61	65,987,823	160.9	165.2	65,752,402	67,379,971
	Arabinan	6	0.2012	-0.13	33.81	6,428,223	31.3	34.8	4,711,341	32,354,931
	Arabinan	8	0.1963	0.1258	0.01	679,896	0	5	679,896	2,133,644
	Acetyl	6	0.1563	-0.0828	82.81	53,328,895	79.8	85.3	52,338,107	53,664,189
	Cellulose	1	0.1922	0.4955	188.61	72,435,660	185.9	189.6	71,837,886	72,938,507
	Cellulose	2	0.1358	-0.4215	120.11	71,422,224	116.1	120.7	69,910,974	71,789,496
	Structural Carb.	8	0.1781	0.8971	63.61	50,116,222	57	65.7	49,587,492	50,817,616
	Soluble Carb.	3	0.1555	-0.8232	4.31	1,447,138	2.4	11.5	421,680	2,235,920
	Water	1	0.1101	-0.1707	183.21	71,191,866	175.1	184.2	67,785,160	71,837,886
	Water	3	0.2113	-0.3003	35.61	7,940,280	35	37	7,543,264	8,150,254
	Water	3	0.112	0.2137	51.01	10,683,496	49.1	52.3	10,391,393	11,160,047
	Water	4	0.2132	-0.2379	49.91	10,094,021	48.6	52.9	9,429,083	10,372,107
	Ethanol	1	0.2513	0.0329	1.11	1,415,784	0	2.8	1,132,065	1,964,168
	Ethanol	1	0.1352	0.0214	177.91	71,099,630	167.9	185	67,785,160	71,837,886
	Ethanol	1	0.1312	0.021	188.61	72,435,660	188	189.6	72,288,774	72,938,507
	Ethanol	4	0.1362	-0.023	113.01	62,670,633	112.8	119.7	62,339,651	63,335,696
	Ethanol	4	0.1499	-0.0234	129.11	64,794,148	128	131.1	64,400,937	65,463,145
	Starch	3	0.193	-0.7818	8.31	1,967,752	5.9	12	1,452,327	2,235,920
	Starch	8	0.2297	-0.8649	8.01	2,133,644	2.6	16.7	679,896	3,300,851
Grain Maturity	Ash	3	0.1205	0.3064	48.31	10,391,393	45.8	50.3	9,985,015	10,683,496
	Ash	6	0.1056	-0.2926	41.51	42,728,446	40.6	43.1	41,856,240	42,795,561

Table 7. Continued.

Dev. Stage	Trait	Chr.	R^2	Add. Var.	Peak (cM)	Peak (bp)	Left (cM)	Right (cM)	Left (bp)	Right (bp)
Grain Maturity	Ash	6	0.1207	-0.305	109.01	58,326,028	100.2	110.1	56,303,525	58,558,664
	SI	3	0.1186	0.1688	4.31	1,447,138	2.6	8.3	421,680	1,967,752
	SI	6	0.1483	-0.1968	97.31	55,318,076	96.4	100.9	55,157,298	56,393,047
	SI	10	0.1837	0.2201	76.91	52,525,422	74.2	79.1	51,120,566	53,232,079
	SI	10	0.1214	-0.1819	104.61	56,728,379	102.6	107.2	56,511,327	56,922,188
	NSI	3	0.1282	0.197	133.01	71,832,221	131.6	136.6	70,597,525	72,711,436
	NSI	6	0.1019	-0.1422	41.51	42,728,446	37	45.3	40,201,125	44,708,620
	Protein	3	0.1594	-0.2433	95.11	60,677,970	92.7	98.6	59,753,639	61,863,809
	Sucrose	1	0.1013	-1.2227	108.31	52,217,159	107.5	110.1	51,826,648	53,516,081
	Sucrose	2	0.1774	-1.6011	94.31	62,441,950	92.7	97.6	61,759,693	63,514,945
	Sucrose	5	0.1337	1.3999	41.91	6,523,469	38.6	43.1	5,325,838	6,857,282
	Sucrose	6	0.1371	1.4461	101.41	56,393,047	99.8	106.6	56,303,525	57,729,673
	Lignin	1	0.1006	0.5603	108.31	52,217,159	107.3	113.5	51,826,648	53,679,180
	Lignin	2	0.0994	0.5531	99.31	63,514,945	94.5	103.2	62,441,950	65,244,625
	Lignin	6	0.1398	-0.6633	74.41	50,347,055	71	76.2	48,499,624	52,094,273
	Lignin	6	0.2429	-0.864	101.41	56,393,047	99.5	105.3	56,230,045	57,222,080
	Glucan	2	0.2347	0.738	112.01	69,695,964	109.1	113.1	68,327,380	69,910,974
	Glucan	6	0.1895	-0.5299	72.01	49,176,678	70.3	75.8	48,499,624	52,094,273
	Xylan	5	0.0881	-0.4052	41.91	6,523,469	40.8	43	5,325,838	6,857,282
	Xylan	6	0.1405	-0.6482	72.01	49,176,678	70.5	72.3	48,499,624	49,772,604
	Xylan	6	0.3358	-0.8978	99.01	56,230,045	98.3	101.1	55,318,076	56,393,047
	Xylan	9	0.0921	0.4065	16.51	2,421,341	14.4	21.4	2,250,695	2,774,240
	Galactan	2	0.1321	0.0253	89.11	60,696,892	84.1	90.5	59,579,955	61,543,010
	Arabinan	2	0.1504	0.217	84.71	59,949,245	83.2	88.8	59,579,955	60,696,862
	Arabinan	6	0.1975	-0.2489	75.01	51,499,148	71.9	76.4	48,740,934	52,094,273

Table 7. Continued.

Dev. Stage	Trait	Chr.	R^2	Add. Var.	Peak (cM)	Peak (bp)	Left (cM)	Right (cM)	Left (bp)	Right (bp)
Grain Maturity	Arabinan	6	0.1239	-0.1927	105.21	57,222,080	96.8	108.8	55,157,298	58,326,028.00
	Acetyl	1	0.1921	0.116	14.61	7,315,100	13.6	19.6	6,457,030	11,843,676.00
	Acetyl	5	0.201	-0.1238	58.91	16,063,396	57.2	66	12,504,766	50,009,642.00
	Cellulose	6	0.2017	-1.0735	75.61	51,499,148	74.4	77.2	50,347,055	52,094,273.00
	Cellulose	6	0.211	-1.2877	98.41	55,997,638	97	101.4	55,157,298	56,393,047.00
	Structural Carb.	6	0.1538	-1.264	75.01	51,139,358	71.9	76.2	48,740,934	52,094,273.00
	Structural Carb.	6	0.1939	-1.419	99.01	56,230,045	97.8	104.7	55,318,076	57,222,080.00
	Soluble Carb.	6	0.1538	1.6212	75.01	51,139,358	72.6	76.9	49,772,604	52,094,273.00
	Soluble Carb.	6	0.285	2.5978	98.41	55,997,638	97.6	100.5	55,318,076	56,393,047.00
	Water	10	0.1042	1.3334	15.31	2,238,012	14.3	18.6	1,901,431	2,856,016.00
	Water	1	0.1018	-0.3964	108.31	52,217,159	107.5	112.2	51,826,648	53,516,081.00
	Water	2	0.1821	-0.5454	89.11	60,696,862	85	91	59,949,245	61,543,010.00
	Water	5	0.1316	0.459	41.91	6,857,282	38.8	42.9	5,325,838	6,857,282.00
	Water	6	0.147	0.4876	101.41	56,393,047	99.6	104.9	56,303,525	57,222,080.00
	Ethanol	2	0.1357	0.0954	84.71	59,949,245	83.3	89.3	59,579,955	61,374,549.00
	Ethanol	5	0.0747	-0.0707	47.11	7,352,477	43.1	50.5	6,857,282	11,125,059.00
	Ethanol	6	0.0956	-0.0804	101.41	56,393,047	99	109.9	56,230,045	58,558,664.00
	Starch	2	0.1356	-1.2812	88.71	60,696,862	83.8	89.7	59,579,955	61,374,549.00
	Starch	5	0.2098	1.6507	41.91	6,523,469	40.1	43	5,325,838	6,857,282.00
	Starch	5	0.142	1.3281	52.21	11,125,059	49.6	56.5	8,856,612	12,504,766.00
	Starch	6	0.1135	1.4996	72.01	49,176,678	71.4	77.1	48,740,934	52,094,273.00
	Starch	6	0.1435	1.3781	104.21	56,586,356	99.7	109.3	56,303,525	58,461,757.00

Table 7. Continued.

Dev. Stage	Trait	Chr.	R^2	Add. Var.	Peak (cM)	Peak (bp)	Left (cM)	Right (cM)	Left (bp)	Right (bp)
Net Change	Ash	2	0.1905	-0.4059	87.71	60,696,862	84.2	89.7	59,579,955	61,374,549
	Ash	5	0.1999	0.4463	41.51	6,523,469	38.5	45.3	5,325,838	7,352,477
	Ash	6	0.1755	0.3741	97.31	55,318,076	96.8	98.7	55,157,298	56,230,045
	SI	6	0.1569	0.2305	97.31	55,318,076	96.5	98.4	55,095,011	55,997,638
	NSI	6	0.1564	0.202	109.71	58,461,757	109	110.4	58,326,028	58,810,612
	NSI	8	0.2144	0.1902	73.01	51,620,632	69.7	76.2	50,817,616	53,057,572
	Protein	4	0.1501	0.3586	128.51	64,611,864	122	130.3	63,335,696	65,021,164
	Protein	6	0.1887	-0.3761	97.31	55,318,076	96.6	99	55,157,298	56,230,045
	Protein	6	0.1753	-0.3612	108.31	58,326,028	105.6	111.4	57,222,080	60,847,137
	Sucrose	1	0.0914	1.2148	103.51	51,522,284	100.8	106.3	50,944,662	51,826,648
	Sucrose	2	0.2216	1.8442	94.31	62,441,950	93.7	97	61,759,693	63,514,945
	Sucrose	5	0.1051	-1.285	41.51	6,523,469	38.7	45.3	5,325,838	7,352,477
	Sucrose	6	0.2302	-1.9132	97.31	55,318,076	96.7	98.4	55,157,298	56,230,045
	Lignin	1	0.0821	-0.4953	108.31	52,217,159	107.5	110.3	51,826,648	53,516,081
	Lignin	2	0.1713	-0.7405	88.71	60,696,862	85.3	89.7	59,949,245	61,374,549
	Lignin	2	0.2016	-0.775	100.31	63,514,945	95.1	103.5	62,441,950	65,244,625
	Lignin	6	0.0912	0.5403	66.71	47,513,584	64.3	69	46,942,206	48,297,388
	Lignin	6	0.1786	0.7477	101.41	56,393,047	100	104.5	56,303,525	57,222,080
	Glucan	2	0.2471	-0.65	116.51	70,665,664	113.7	119.3	69,910,974	71,422,224
	Glucan	3	0.1665	-0.537	132.01	71,832,221	128.6	135	70,597,525	72,173,211
	Xylan	6	0.2737	0.7381	97.31	555,318,076	96.8	98.4	55,157,298	55,997,638
	Cellulose	6	0.1285	1.0316	72.01	49,176,678	70.9	74.4	48,499,624	50,347,055
	Structural Carb.	2	0.1269	-1.1083	117.11	70,816,798	116.8	120.6	70,665,664	71,789,496

Table 7. Continued.

Dev. Stage	Trait	Chr.	R^2	Add. Var.	Peak (cM)	Peak (bp)	Left (cM)	Right (cM)	Left (bp)	Right (bp)
Net Change	Structural Carb.	6	0.109	1.079	75.01	51,139,358	74.4	76.9	50,347,055	52,094,273
	Soluble Carb.	2	0.2122	1.7517	100.31	63,514,945	96.5	103.8	62,441,950	65,244,625
	Soluble Carb.	6	0.2477	-2.3213	97.31	55,318,076	96.9	99	55,157,298	56,230,045
	Water	1	0.1007	0.4186	65.71	16,465,677	55.9	67.6	13,616,615	18,272,587
	Water	2	0.1422	0.4872	97.31	63,514,945	93.6	103.1	61,759,693	65,244,625
	Water	6	0.0876	-0.399	66.71	47,513,904	64.7	69.7	46,942,206	48,297,388
	Water	6	0.1875	-0.6339	97.31	55,318,076	96.9	99	55,157,298	56,230,045
	Ethanol	2	0.1591	-0.1093	84.71	59,949,245	83.4	89.3	59,579,955	61,374,549
	Ethanol	6	0.0881	0.0863	97.31	55,318,076	96.6	99	55,157,298	56,230,045
	Starch	1	0.0967	1.1041	131.01	56,707,521	128.1	132.6	56,245,572	57,158,862
	Starch	2	0.0992	1.1154	101.31	63,514,945	94.3	106.8	62,441,950	68,327,380
	Starch	6	0.1501	-1.3612	104.21	56,586,356	99.9	108.4	56,303,525	58,326,028

Stem Hollowing

The extent of stem hollowing was phenotyped because this is correlated with stem resource remobilization to the panicle. Stem hollowness was phenotyped at two different time points: anthesis and grain maturity and at three different locations along the stem; the peduncle, the middle of the stem, and the lower stem.

Several QTL for stem hollowing were identified (Figure 20). Information about the positions and other information about these QTL is summarized in Table 8. The most consistent QTL for stem hollowing was identified on LG07. This QTL was consistently identified in analyses of mid-stem hollowing. The QTL for the change in stem hollowness often co-aligned with QTL for stem hollowness at the anthesis indicating that the genes involved continue to be active through grain maturity. QTL for peduncle hollowness did not align with vegetative stem hollowness indicating these traits are under separate genetic regulation.

Figure 20. QTL analysis results for stem hollowing data in the 2013 College Station BTx642 x Tx7000 population. The height of the QTL indicates the relative LOD score and the length represents the 1 LOD interval in cM. The main QTL identified is on LG07 in all phenotypes at the middle internode.

Table 8. QTL information from the 2013 College Station BTx642 x Tx7000 RIL population for stem hollowness. Chr. indicate chromosome, R2 is the amount of the additive variance that is explained and AddVar indicates the additive variance.

Dev. Stage	Trait	Chr	R^2	AddVar	Peak cM	Peak bp	cM Left	cM Right	bp Left	bp Right
Anthesis	Hollowness at Peduncle	1	0.3337	-0.2159	132.11	57,141,926	131.5	133.5	56,707,521	57,158,862
	Hollowness at Middle	2	0.1599	0.0373	106.81	68,327,380	105.5	109.4	65,731,619	69,615,903
	Hollowness at Middle	2	0.0946	-0.0363	120.71	71,789,496	120.7	125.2	71,422,224	73,082,593
	Hollowness at Middle	7	0.2611	-0.0544	29.81	4,204,880	29.1	33.2	3,952,824	5,101,119
	Hollowness at Bottom	1	0.2531	-0.0516	179.91	70,500,000	171.3	183	68,000,000	71,191,866
	Hollowness at Bottom	6	0.1399	-0.0443	34.91	32,354,931	33.7	36.2	6,400,000	40,120,199
Grain Maturity	Hollowness at Peduncle	6	0.1906	-0.1071	79.81	52,338,107	79.5	82.8	52,094,273	53,328,895
	Hollowness at Middle	7	0.1673	0.0884	39.11	5,180,259	34.2	40.5	4,252,105	5,186,838
	Hollowness at Bottom	6	0.1578	-0.0703	101.41	56,393,047	100.2	105.2	56,303,525	57,222,080

BTx623 x IS3620c

BTx623 x IS3620c RILs grown in the field in College Station (2012) were analyzed for grain composition traits and the information was used to identify QTL that modulate these traits. The resulting QTL from this study are shown in Figure 21. A QTL on LG02 was identified for both tannins and phenols. Also, a QTL on LG04 was identified for variation in ash, tannins, and phenols. Tannins and phenols consistently appear to have overlapping QTL. The QTL on LG02 has a high R^2 value explaining over 20% of the total genetic variance of both tannins and phenols. Detailed information for these QTL is shown in Table 9.

The BTx623 x IS3620c population showed no QTL for grain weight. This is surprising as these two parents have a large difference in the weight per grain; however, these findings suggest that the variation in grain weight in this population is driven by differences in grain number. The grain yield QTL from this population overlapped with QTL for grain number and its subcomponents.

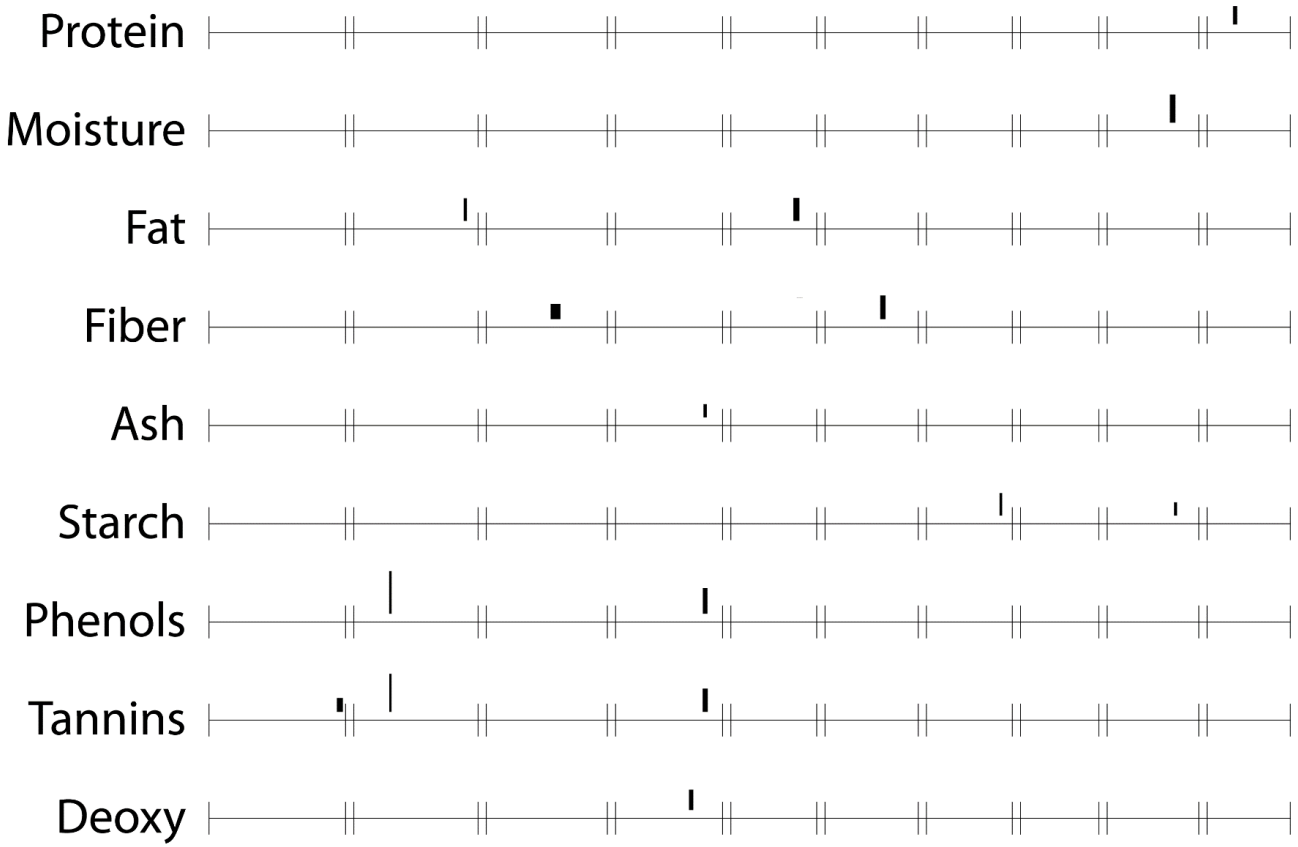

Figure 21. All QTL identified in grain composition analysis via NIR for the BTx623 x IS3620c 2012 College Station population. The height of the box representing a QTL corresponds to its relative LOD score while the width of each box indicates the 1 LOD interval of the identified QTL in cM. The QTL identified on LG02 and LG04 shown across multiple composition phenotypes, specifically phenols and tannins.

Table 9**.** QTL information for the BTx623 x IS3620c 2012 College Station population grain composition results. Chr. indicate chromosome, R2 is the amount of the additive variance that is explained and AddVar indicates the additive variance. Phenotyping was performed via NIR spectroscopy.

Trait	Chr.	R2	AddVar	cM peak	bp peak	cM Left	cM Right	bp Left	bp Right
Protein	10	0.102111	-0.3157	33.91	5,620,238	33.1	35.6	5,375,742	6,206,947
Moisture	9	0.166453	-0.1853	80.91	52,770,876	78.4	81.8	52,099,663	52,882,707
Fat	2	0.124722	-0.1284	135.81	72,307,103	135.3	137.2	71,874,825	72,638,591
Fat	5	0.133377	0.1369	79.01	57,621,392	78.4	82	57,481,755	58,216,468
Fiber	3	0.079537	0.0361	84.81	58,258,957	81.7	87.7	57,505,491	58,725,408
Fiber	6	0.14141	0.0489	69.81	52,481,031	69.1	72.3	52,247,000	53,155,569
Ash	4	0.079432	0.0128	109.51	61,803,781	108.5	110.4	61,537,356	62,040,437
Starch	7	0.141853	0.3851	90.91	60,776,023	90.2	91.8	60,604,810	60,886,344
Starch	9	0.078157	0.2922	83.11	52,939,589	82.7	84.5	52,900,944	53,050,536
Phenols	2	0.245821	1.2453	44.91	8,956,343	43.9	45.4	8,752,460	9,348,427
Phenols	4	0.14342	-0.9705	109.51	61,803,781	108.2	110.9	61,537,356	62,047,608
Tannins	1	0.06678	-1.3162	160.21	72,333,449	158.3	161.9	71407886	72,210,372
Tannins	2	0.214333	2.3914	44.91	8,956,343	44.2	45.4	8,878,778	9,111,526
Tannins	4	0.120447	-1.816	109.51	1,803,781	108.1	111.2	61,474,109	62,086,779
Deoxy	4	0.130091	-3.6938	92.71	58,008,566	91.1	93.7	57,281,304	58,281,320

SC170 x M35-1

QTL results from the SC170 x M35-1 grain NIR data are shown in Figure 22. For this population, QTL that contributed to grain weight were identified on LG01, LG06, and LG09. QTL on LG06 that aligned with grain weight QTL were also identified for starch, protein, and fat conents of grain composition. This QTL on LG06 aligns with a QTL that was identified in the BTx642 x Tx7000 population for grain composition. This QTL was consistently identified in both populations, and it is responsible for a large portion of the genetic variance in this population, contributing ~25% of the genetic variance for the phenotypes modulated by this QTL. A QTL on LG09 for grain weight was aligned with a QTL for grain starch content in the same population. Additional QTL were identified in the grain composition analysis and are reported in Figure 22. Detailed information for these QTL are shown in Table 10.

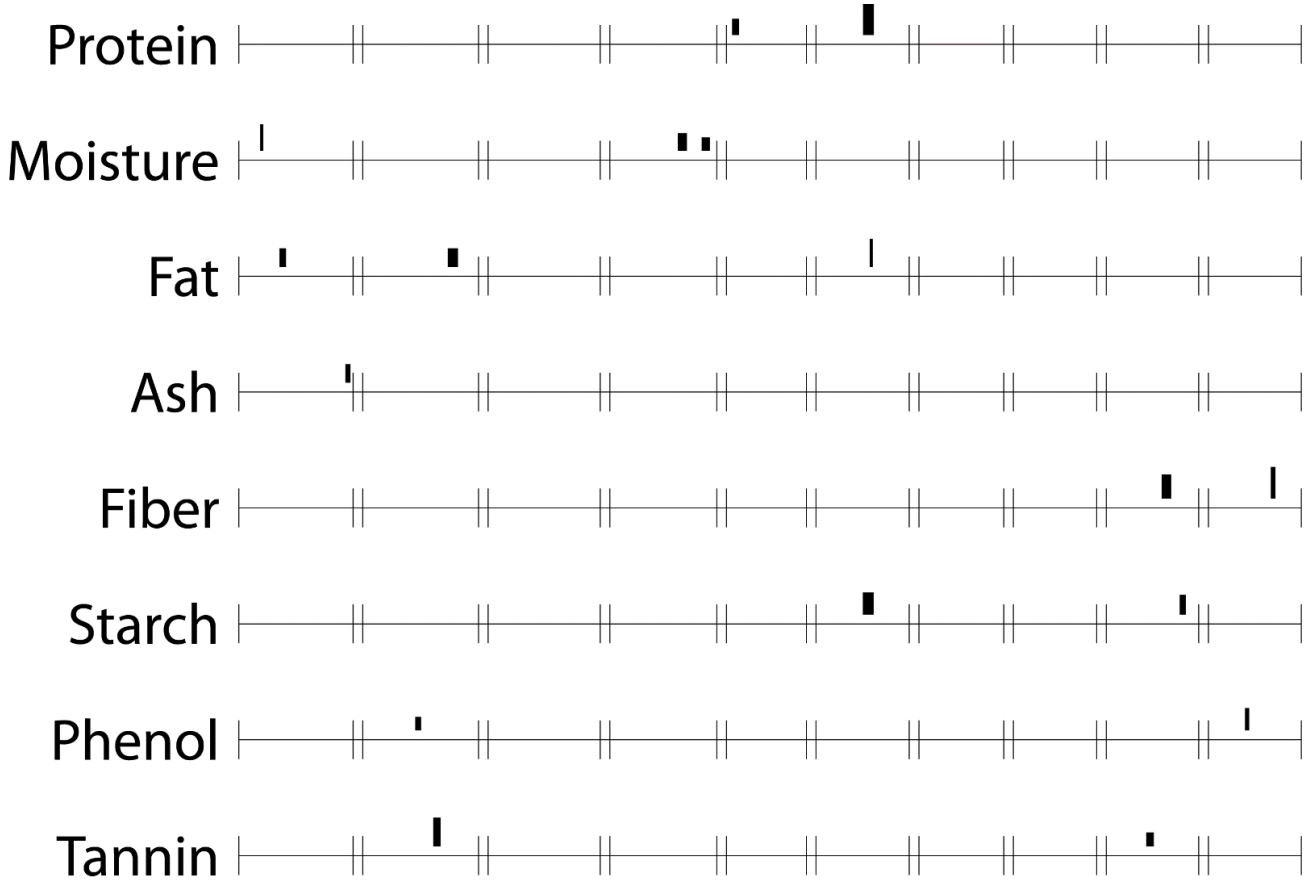

Figure 22. All QTL identified in grain composition analysis via NIR for the SC170 x M35-1 population. The height of the box representing a QTL corresponds to its relative LOD score while the width of each box indicates the 1 LOD interval of the identified QTL in cM. Several QTL that were identified in grain weight for this population in Chapter II were identified here on LG06 and LG09.

Table 10. SC170 x M35-1 QTL information for grain composition. Chr. indicate chromosome, R^2 is the amount of the additive variance that is explained and AddVar indicates the additive Phenotyping was performed via NIR spectroscopy.variance.

Trait	Chr.	R^2	Add. Var.	Peak (cM)	Peak (bp)	Left (cM)	Right (cM)	Left (bp)	Right (bp)
Protein	5	0.10726	0.4515	10.11	2,100,000	6.3	12.7	1,700,000	2,700,000
Protein	6	0.26112	0.7073	52.61	43,000,000	49.2	59.5	41,500,000	47,700,000
Moisture	1	0.272356	0.2127	23.71	9,200,000	22.5	24.7	8,172,966	10,800,000
Moisture	4	0.176553	0.1885	74.21	56,300,000	70.6	79.4	55,700,000	58,000,000
Moisture	4	0.126458	-0.1606	99.21	64,800,000	95.5	103.2	63,100,000	65,557,214
Fat	1	0.143964	-0.1632	45.01	19,300,000	42.5	48.3	18,200,000	20,200,000
Fat	2	0.153835	0.1696	93.81	70,646,966	88.7	98.3	69,950,000	71,837,128
Fat	6	0.24618	0.2132	58.31	44,755,457	54.8	59.9	44,800,000	47,800,000
Ash	1	0.188522	-0.0186	113.11	71,520,000	110.9	115.2	71,000,000	72,000,000
Fiber	9	0.165934	0.045	63.61	50,500,000	57.8	66.9	49,800,000	51,000,000
Fiber	10	0.249899	-0.0534	67.91	55,100,000	65.1	69	53,905,115	55,250,000
Phenol	2	0.119243	-0.3466	57.01	58,022,189	54.9	60.1	57,226,739	58,800,000
Tannin	2	0.214527	-1.119	78.81	63,420,498	73.7	80.1	61,500.000	63,621,131
Tannin	4	0.209355	1.1151	99.21	65,000,000	94.8	102.2	63,100,000	65,196,742
Tannin	9	0.092929	-0.7399	45.81	5,700,000	41.9	48.8	4,281,719	7,175,946

Discussion

Grain yield is a complex trait with many subcomponents that affect the overall phenotype. The two main subcomponents of grain yield are grain weight and grain number. Both of these phenotypes have their own subcomponent traits that may be analyzed to better understand the genetic control of variation of grain yield. In this chapter, several grain weight subcomponents, including grain composition, stem hollowing, and stem composition, were analyzed. Quantifying stem hollowing and stem composition at anthesis, the start of the grain filling process, and at grain maturity, was done to quantify the reallocation of resources from the stem to the grain. Changes in stem composition may also provide insight into the mechanisms by which grain size is determined.

In the BTx642 x Tx7000 population, five QTL were identified for grain weight using MQM analysis. Four of these five QTL aligned with QTL that modify grain composition. The fifth grain weight QTL aligned with a stem biomass QTL. The lack of overlap between QTL for stem composition/hollowing and grain composition was a surprising result as it was hypothesized that the reallocation of stem resources would have a direct correlation to the allocation of new resources to the grain. It is possible that these QTL could not be identified in grain composition but could still be altering the grain weight. However, these QTL did not align consistently with QTL from the overall trait of grain weight or yield suggesting that is not the case. It is likely that the main resources for determining grain weight do not come from the stem biomass, but instead

through ongoing photosynthesis and nitrogen assimilation that occurs during grain filling. This hypothesis could be tested by analyzing rates of photosynthesis and nitrogen assimilation under good conditions and by studying this population under nutrient stress from anthesis to grain maturity and determining of if the stem nutrient QTL align with grain composition QTL.

Grain Composition for BTx642 x Tx7000 RIL population

Several QTL were identified for grain composition in the BTx642 x Tx7000 population (Figure 18). While the identified QTL were fairly stable across years, there were also several QTL for composition traits that aligned with the QTL for overall grain weight and overall grain yield. Most notably, the QTL for grain weight and grain yield on LG03 aligned with QTL for grain protein, moisture, fat, fiber, starch, phenol, and tannins content on LG03. These results suggest that the grain yield QTL on LG03 may correspond to a QTL that modulates grain composition.

It was initially hypothesized that the QTL on LG03 could align with QTL for changes in stem biomass that occur from anthesis to grain maturity. However, this QTL was aligned only with a QTL for stem protein content. The mechanistic basis of this QTL correlation is unknown, but it could be important if it modulates grain weight and grain yield phenotypes.

It has been argued that there is a direct link between grain number and grain yield; therefore making grain number a potentially more important and genetically controllable factor for grain yield (47). However, due to the

consistency of the QTL identified in the NIR grain composition data between years and environments, this study suggests that, for the years and environments tested, grain weight and grain composition are more important to the overall grain yield phenotype. Therefore, the results of this study indicate that grain weight has a high level of genetic control. The appearance of the grain weight QTL that align with grain yield QTL suggests that for the populations and environments studied, variation in grain weight contributes significantly to variation in grain yield. MQM analysis of grain weight supports this as the generated model explains significantly more of the variance in grain weight than the MQM model of grain number as shown in Figure 13.

Stem Biomass and Hollowing for BTx642 x Tx7000 RIL population

It was hypothesized that a portion of the resources used for grain filling were remobilized from stems between anthesis and grain maturity. This hypothesis was supported by the fact that QTL resulting from the change in stem biomass from anthesis to grain maturity overlapped with the QTL reported for grain maturity. Additionally, QTL for stem biomass at anthesis had no overlap with any other biomass category. QTL that appeared consistently in the change in stem biomass did not overlap with QTL identified for grain composition. This suggests that the genes contributing to the change in biomass of the stem are not the same genes that are impacting nutrient content in the grain. It should be noted that, despite different QTL being identified, this does not indicate that there is not a correlation between change in biomass and grain composition or

stem hollowing isn't a reallocation of nutrients from the stem to grain. Instead, the lack of overlap in QTL shows that the reallocation of resources that occurs during stem hollowing is a more complex pathway than hypothesized and all of these factors cannot be captured via QTL mapping alone.

QTL Across Populations

The BTx623 x IS3620c population shows a higher genetic control of grain yield through grain number rather than grain weight. This finding is the opposite of what was observed in the BTx642 x Tx7000 population. Given the IS3620c parent's small seed phenotype and wild aspects of this parental line, this result is not surprising. No QTL were identified for grain weight in the BTx623 x IS3620c population; however, there were a few QTL of note that appeared in multiple grain composition traits on LG02 and LG04. These QTL appear to have an effect on tannin and phenol content. Not only do these QTL show a consistent impact but a large percentage (~40%) of the genetic variance in tannins and phenols is explained by these two QTL. This result indicates that these two QTL have an important role in controlling the antioxidant levels of sorghum grain.

The SC170 x M35-1 population supports the importance of the tannin and phenol content QTL found on LG04 as the same QTL in this population as well. Once again, it was observed that this QTL explains a significant amount of the variance identified (~21%). Any QTL identified in more than one population, is noteworthy as it may play an important role in the sorghum-breeding program.

Analysis of SC170 x M35-1 identified a QTL on LG06 for grain protein, moisture, and starch composition that was identified in the BTx642 x Tx7000 population. This QTL also aligned with QTL for grain weight and multiple phenotypes for grain composition. These QTL are important to note as their appearance across populations indicates that multiple alleles are prevalent in the sorghum germplasm. This prevalence makes identification of the best allele even more important to find as these loci play a large role in variance explanation in many different populations.

Future Directions

Further analysis of grain weight traits is necessary to develop a better understanding of the grain-filling pathway. Additionally further studies will identify interactions between traits that determine the grain weight produced by an individual plant. As there are many environmental factors, this can be a challenging study. However, by optimizing the grain weight under various conditions, more QTL may be identified that are affecting the overall phenotype. Eventually, a pathway for the determination of grain weight in sorghum could be elucidated.

Materials and Methods

For this study, grain yield was quantified on a per plant basis where plants had one main culm. By observing grain yield on a per plant basis in fields that were thinned to a standard plant density, factors such as plant density and tillering, which changes plant density, were minimized as modifiers of grain yield

per plant. All populations that are reported here were field trials from 2011 to 2013 in College Station and Lubbock. All populations were irrigated except 2013 College Station and 2012 Lubbock of the BTx642 x Tx7000. Grain sorghum genotypes normally produce one panicle per plant when grown at standard planting densities. Grain trait data was collected on 5 panicles per plot (genotype) using a randomized replicated planting design. As a result, 2 sets of 5 panicles were analyzed for each genotype. Panicles were not collected from plants at the end of plots because these plants intercept more light and have greater access to water compared to plants within rows.

The standard for quantifying the individual weight per grain produced by a plant is to measure the weight of 1000 grains. A seed counter was utilized to count out 1000 grains that were subsequently weighed. Grain composition data was collected via NIR. This is a method used to estimate composition once a calibration curve has been established (165). Equations were derived by the Rooney lab (165). The grain samples were loaded into the NIR machine in batches of ~100 grain that were randomly selected from the entire sample. Each sample was run twice to account for technical errors. The stem hollowness data was collected at both the anthesis and grain maturity harvest. The leaf sheaths were removed from the stem and then the stem was cut at 3 different locations: peduncle, a middle internode, and the 2nd to lowest internode. All of these sample sites were cut in the middle of the internode or peduncle to eliminate any affects from the nodes themselves. Hollowness was measured on a 1 to 5 rating

scale to denote the degree to which the stem had hollowed. After these samples were collected, the stems were chopped and dried. The stems were then ground and used for stem biomass composition data via NIR as well. These samples were also taken at both anthesis and grain maturity.

BTx642 x Tx7000 and BTx623 x IS3620c are both RIL populations meaning that these populations have been selfed to the point that the heterozygosity in these populations has been minimized. SC170 x M35-1 were in the F4 indicating that heterozygosity levels should be at approximately 12.5%. All QTL analyses were run under the assumptions of the populations being RILs as this best matched level of heterozygosity that was present.

Genotyping data for QTL mapping was prepared by Digital Genotyping (DG) and then run on the Illumina HiSeq for all populations For BTx642 x Tx7000, over 10,000 unique markers were identified. Approximately 2,750 informative markers were identified. Informative markers are markers that provide new information to the refinement of QTL due to breakpoints found in a given population. This was a necessary step for this population to perform the MQM analysis in a reasonable amount of time as the 10,000 permutation calculations take a significant amount of time for each additional marker so eliminating markers that were not informative sped up the analysis. The SC170 x M35-1 genetic map contained ~600 markers and BTx623 x IS3620c had ~850 markers in the maps used for QTL mapping.

All QTL mapping was performed through QTL Cartographer (58). MQM was performed via the Rqtl package (61, 156).). All populations were run as RILs with the Kosambi mapping function. BTx642 x Tx7000 has ~90 RILs in its population. BTx623 x IS3620c has ~400 RILs in its population and SC170 x M35-1 F4s has ~200 lines. All lines were phenotyped for each trait using 10 biological replicates as described above.

CHAPTER IV

IDENTIFICATION OF *MATURITY LOCUS (MA_2)* IN *SORGHUM BICOLOR*

Introduction

Sorghum [*Sorghum bicolor* (L.) Moench] is a cereal crop native to Africa that is commonly used as both food and forage stock (1, 6, 177, 178). Sorghum is a photoperiod sensitive short day plant with critical photoperiods ranging from 11-14 hours (179, 180). When grain sorghum was introduced into more temperate climates, breeders selected for earlier flowering times to optimize grain yield in the new environment (38, 39, 62). In the U.S., this led to the identification of six maturity loci, Ma_1-Ma_6, which when recessive, resulted in an earlier flowering phenotype in long days. In optimal conditions, overall grain yield is positively correlated with longer flowering times (39). In temperate climates, grain maturation needs to occur before frost or onset of poor growing seasons and as a consequence, earlier flowering times were often better suited for agricultural purposes (38).

Earlier flowering times were introduced into sorghum lines following the identification of mutations in six different loci that control flowering time (Ma_1 – Ma_6). These maturity genes enabled breeders to select for the optimal days to flowering (DTF) in a given environment (39). By crossing different varieties of Milo, Quinby and Karper created 4 different phenotype classes of flowering: early, intermediate, late, and ultra-late. The DTF of these phenotype classes ranged from ~40 to 100 days. Genetic analysis showed that 3 maturity loci

contributed to the variation in flowering time, *Ma*$_1$, *Ma*$_2$, and *Ma*$_3$ (38). The phenotypic impacts of *Ma*$_1$ – *Ma*$_3$ were observed in long, 14-hour photoperiod day conditions and no phenotypic impacts were shown under short, 10-hour photoperiod day conditions. *Ma*$_1$ – *Ma*$_3$ were named in order of their phenotypic effect; with *Ma*$_1$ having the largest impact. Notably, the phenotypic effects of *Ma*$_2$ and *Ma*$_3$ were attenuated in *ma*$_1$ genetic backgrounds. Due to this epistatic interaction, it has been hypothesized that *Ma*$_2$ and *Ma*$_3$ were downstream of *Ma*$_1$ in the flowering time pathway.

Quinby later identified a 4th maturity locus (*Ma*$_4$) in Hegari *Ma*$_1$*Ma*$_2$*Ma*$_3$*ma*$_4$ (62). *Ma*$_4$ was found by crossing Hegari with different Milo lines, including 60M which is *ma*$_1$*ma*$_2$*ma*$_3$*Ma*$_4$. The flowering time distribution in this cross showed segregation of 4 distinct maturity genes. *Ma*$_1$ – *Ma*$_4$ are widely considered the four classic maturity genes in sorghum; however breeders continued to see variation in DTF that could not be explained by only *Ma*$_1$ – *Ma*$_4$ (39). For example, *Ma*$_5$ and *Ma*$_6$ were identified in populations that showed even higher photoperiod sensitivity and delayed flowering in long days (>12.4 hours) (39).

There has been considerable effort to identify the causative genes/alleles that correspond to the 6 maturity loci. *Ma*$_3$ was identified to be *PHYTOCHROME B* (*PHYB*) (63). The recessive *ma*$_3$ was reported as the result of a frameshift mutation resulting in a stop codon in 58M. The phenotypes displayed by the *PHYB* mutant sorghum included very early flowering, a reduction of chlorophyll

content, and lack of sensitivity to red light (63). This result is similar to what is observed in *PHYB* mutants in Arabidopsis (64–66).

Ma_1 was identified as the *PSEUDO RESPONSE REGULATOR PROTEIN 37 (PRR37)* (67). Under long days, *SbPRR37* represses *CONSTANS* and EHD1 which encode activators of flowering time. The overall effect on flowering time caused by *SbPRR37* is influenced by both the circadian clock and photoperiod (67). *Ma_6* has been identified as *GRAIN NUMBER, PLANT HEIGHT AND HEADING DATE (Ghd7)* (67).*Ghd7* works with *PRR37* to repress floral initiation by repressing the expression of EHD1. *Ma_5* has been identified as *PHYTOCHROME C (PHYC)* through genetic analysis and sequence variants of PHYC found in R07007 (*ma_5*), 90M (*$ma_{5\text{-}1}$*), 100M (*$ma_{5\text{-}1}$*), and BTx623 (*Ma_5*) (68). The identification of 4 of the 6 sorghum maturity genes has enabled great progress towards an understanding the flowering time pathway of sorghum.

Results

Map-based Cloning of the Ma_2 Gene

To identify maturity gene *Ma_2*, a population was derived by crossing Hegari with 80M. Hegari is *$Ma_1Ma_2Ma_3ma_4Ma_5Ma_6$* while 80M is *$Ma_1ma_2Ma_3Ma_4Ma5Ma_6$;* therefore, *Ma_2* and *Ma_4* were expected to segregate in this population. The F2 population derived from this cross was grown in long days in a greenhouse and plants were phenotyped for days to flowering and genotyped using Digital Genotyping (59). QTL for flowering time were identified on LG02 and LG10 (Figure 23). Prior analysis showed that *Ma_4* was located on

LG10 and the allele causing early flowering on LG10 was derived from Hegari (*ma*$_4$). Therefore *Ma*$_2$ was identified as the QTL located on LG02 spanning from 67.3 Mbp to 69.1 Mbp (Figure 23). The allele causing early flowering corresponding to the QTL on LG02 was derived from 80M (*ma*$_2$).

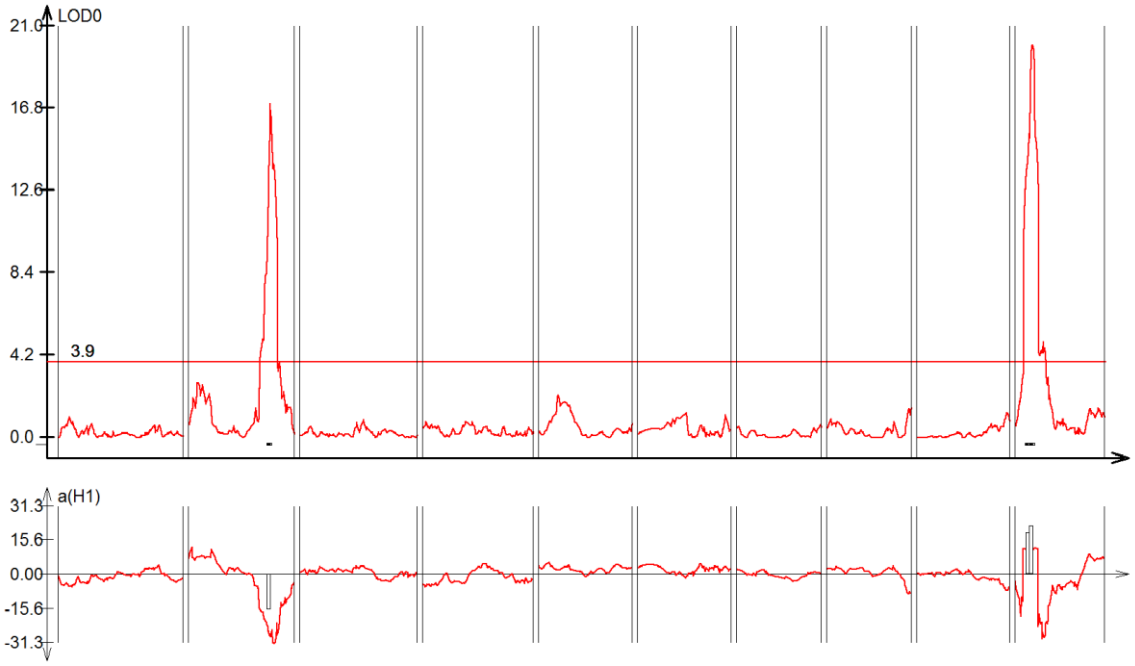

Figure 23. QTL map of Hegari*80M F2s with both *Ma*$_2$ and *Ma*$_4$ identified on LG02 and LG10 respectively.

A fine mapping population was created to aid with the identification of *Ma*$_2$. Lines from the Hegari x 80M populations which were heterozygous across the *Ma*$_2$ QTL region, but not segregating for *Ma*$_4$, were planted in long days in August 2013. Approximately 1000 plants from this population were phenotyped for flowering time and genotyped via DG analysis in order to fine map the region encoding *Ma*$_2$. After calculating the recombination frequency that was expected

within the region of interest, it was determined using Eq. 2 that 1000 plants would be needed to get the desired number of recombinants.

$$recombination\ frequency = \frac{num.of\ recombinants}{total\ num.of\ progeny} \quad 2$$

The genetic distance spanning the *Ma₂* locus is 2 cM corresponding to a physical distance of ~1.8 Mbp. Based on Eq. 2 above, it was expected that 20 recombinants in the 2 cM *Ma₂* locus would be found by screening the 1000 plants derived from the *Ma₂* heterogeneous inbred family (HIF) population. DG was performed for all plants. This study narrowed the region down to ~600kb (67.72Mb-68.33Mb) as shown in Table 11. This process was repeated, growing out approximately 100 plants from 5 of the HIF lines that were found to have heterozygosity across the region of interest and 25 plants from 2 additional HIFs that had a breakpoint in the region of interest with heterozygosity still present on one side of the breakpoint. This fine mapping population narrowed down the region to ~500kb (67.72Mb-68.22Mb) as shown in Table 12.

Early attempts to map *Ma₂* used a population derived from a cross of 80M and 100M. The population segregated 3:1 as expected, however, no QTL were identified because one arm of chromosome 2 in 80M and 100M was nearly identical in sequence and no DG markers were identified. The *Ma₂* locus lies in this region suggesting that a mutation in *Ma₂* occurred in this region of 100M resulting in a recessive *ma₂* allele. In this case, we would expect that a comparison of 100M and 80M genomic sequences across the fine mapped *Ma₂* locus would identify a mutation in the gene corresponding to *Ma₂*. Genome

sequences of 80M and 100M were obtained and aligned to the reference genome sequence. The genomic region spanning *Ma*$_2$ was then scanned to identify SNPs or INDELS within the region of interest between 67.72 Mb and 68.22 Mb on LG02. Due to common ancestry, there were very few sequence differences within this arm of chromosome 2 for these two parental lines. Only 1 SNP located in Sobic.002G302700 was identified in the 80M whole genome sequence within the entire ~500kb region on LG02 (Figure 24). The mutation resulted in a stop codon in the gene sequence from 80M consistent with a gene corresponding to *Ma*$_2$.

Table 11. First fine mapping population for identification of *Ma*$_2$ breakpoints of interest. Lines that show breakpoints that narrow down the region are shown.

Position (bp)	**HIF-1**	**HIF-2**	**HIF-3**	**HIF-4**	**HIF-5**	**HIF-6**
Days to Flowering:	**65**	**90**	**66**	**101**	**63**	**94**
67330461	bb	ab	bb	-	bb	bb
67330473	bb	ab	bb	-	bb	bb
67340729	ab	ab	bb	ab	ab	bb
67469853	ab	ab	bb	-	ab	bb
67698634	bb	ab	bb	ab	bb	bb
67720272	bb	ab	bb	ab	bb	-
68327569	bb	bb	ab	aa	bb	ab

Table 12. Second fine mapping population for identification of *Ma₂* breakpoints of interest. Lines that show breakpoints that narrow down the region are shown.

Position (bp)	**Hegari**	**80M**	**HIF-1**	**HIF-2**	**HIF-3**	**HIF-4**	**HIF-5**
Days to Flowering:	**99**	**74**	**70**	**86**	**86**	**87**	**63**
67176490	aa	bb	aa	ab	aa	ab	aa
67429275	aa	bb	ab	ab	ab	ab	ab
67626147	aa	bb	ab	bb	ab	bb	bb
67626162	aa	bb	ab	bb	ab	bb	bb
67626174	aa	bb	ab	bb	ab	bb	bb
67629889	aa	bb	ab	ab	ab	ab	bb
68229806	aa	bb	bb	ab	bb	ab	bb
68871492	aa	bb	bb	ab	bb	bb	ab

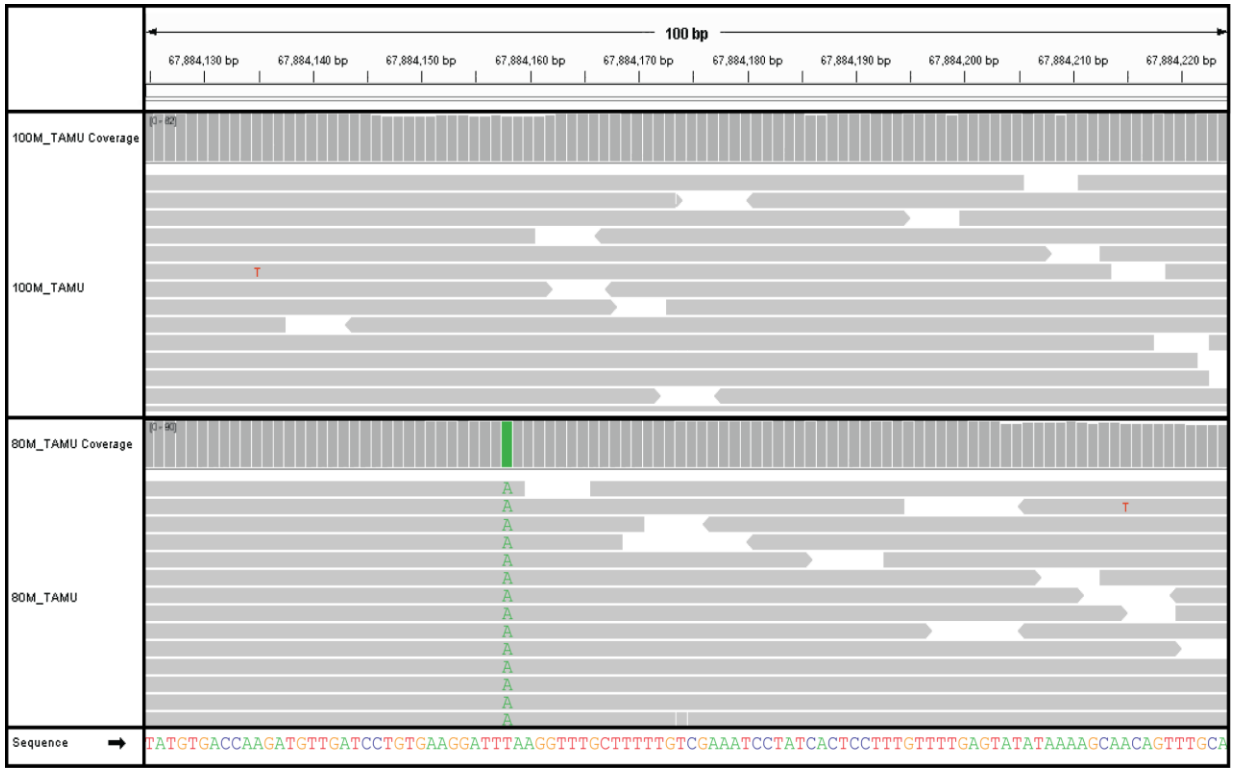

Figure 24. Whole genome sequence of 100M and 80M at the *ma*$_2$ mutation location. The top is showing the 100M whole genome sequence compared to the reference genome and the bottom is showing 80M. The T to A mutation is shown in green.

Sequencing Ma_2 Gene

80M was sequenced to confirm the SNP identified in the whole genome sequence (Figure 25). The same T to A point mutation was identified in 80M while both 100M and BTx623, the reference genome, have the T for dominant *Ma_2*. This gene was identified using Phytozome as Sobic.002G302700.1 (Figure 26) (96). The T to A point mutation in this gene creates the stop codon UAA at the end of the third of eleven exons. Sobic.002G302700 is a zinc ion binding protein suggesting it is a transcription factor.

Table 13 shows additional variation in the *Ma_2* gene that was found in the different lines that have currently been whole genome sequenced. The first SNP identified was in 80M and is the T to A point mutation that was found previously. This mutation was not found in any of the other sorghum lines that have been whole genome sequenced. Three lines stand out in the table shown: PI 586430, S. propinquum 369-1 and S. proprinquum 369-2. These lines have several mutations identified in the *Ma_2* gene; however, variation does not imply an effect on expression.

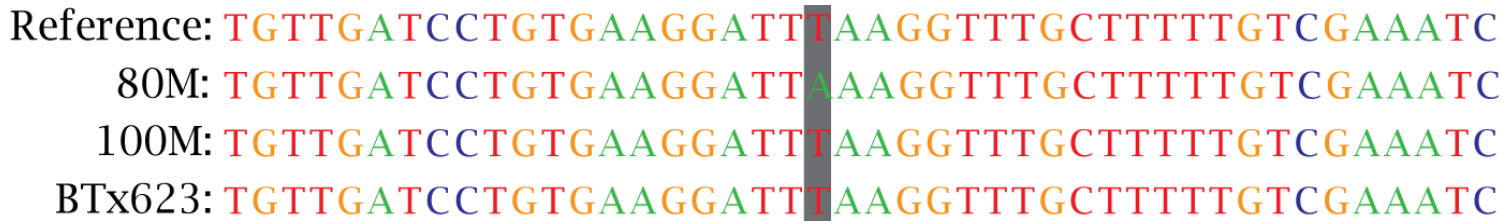

Figure 25. Confirmation of *ma₂* mutation via sequencing. The T to A point mutation seen in line 3 is the 80M sequence at this location. This matches the findings of the whole genome sequence comparisons.

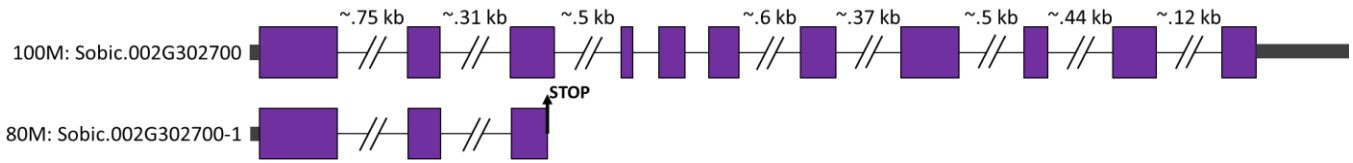

Figure 26. *Ma₂* gene. Exons are shown by boxes and introns are shown by line. UTR regions are grey boxes.

Table 13. Variation identified in Ma2 gene from the different lines that have whole genome sequencing data. Variation is highlighted in yellow. SNPs and indels were identified in this region. Three lines stand out in the table shown: PI 586430, S. propinquum 369-1 and S. proprinquum 369-2 as having a high number of mutations. Any of these mutations could lead to a ma2 but these potential loss of functions have not been investigated.

POS	REF	80M	IBC-E-3843	IS3614-2	IS3620C	Kilo-IBC-E-382	PI300119	PI586430	PI226096
67884158	T	A/A	T/T	T/T	T/T	T/T	T/T	T/T	T/T
67882687	G	G/G	G/C	G/G	G/G	./.	G/G	G/G	G/G
67882698	G	G/G	G/G	G/G	G/G	./.	G/G	T/T	G/G
67882737	G	G/G	./.	./.	G/G	./.	./.	./.	G/G
67882786	C	C/C	./.	C/C	C/C	./.	C/C	./.	C/C
67882804	G	G/G	G/G	G/G	G/G	G/G	G/G	./.	G/G
67883724	T	T/T	T/T	T/T	T/T	T/T	T/T	C/C	T/T
67883732	G	G/G	G/G	G/G	G/G	G/G	G/G	A/A	G/G
67884143	A	A/A	A/A	A/A	A/A	A/A	A/A	C/C	A/A
67884778	A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A
67887383	G	G/G	G/G	G/G	G/G	G/G	G/G	A/A	G/G
67884645	A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A
67885815	C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
67887574	C	C/C	C/C	T/T	C/C	C/C	C/C	C/C	C/C
67882645	AC	AC/AC	AC/AC	AC/AC	AC/AC	AC/AC	AC/AC	A/A	AC/AC
67887675	G	G/G	G/G	G/G	G/G	T/T	G/G	G/G	G/G
67887699	A	A/A	A/A	A/A	A/A	A/A	A/A	C/C	A/A
67887732	C	C/C	C/C	C/C	C/C	C/C	C/C	T/T	C/C

Table 13. Continued.

POS	REF	80M	IBC-E-3843	IS3614-2	IS3620C	Kilo-IBC-E-382	PI300119	PI586430	PI226096
67887745	A	A/A	A/A	A/A	A/A	A/A	A/A	G/G	A/A
67887794	A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A
67887852	C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
67887886	A	A/A	A/A	A/A	A/A	A/A	A/A	T/T	A/A
67887985	AAT	AAT/AAT	AAT/AAT	AAT/AAT	AAT/AAT	AAT/AAT	AAT/AAT	A/A	AAT/AAT
67887993	T	./.	./.	T/T	T/T	T/T	TA/TA	T/T	TA/TA
67887994	T	T/T	./.	T/T	A/A	T/T	./.	T/T	./.
67888013	G	G/G	G/G	G/G	G/G	G/G	G/G	A/A	G/G
67888094	G	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A
67888121	G	G/G	G/G	G/G	G/G	G/G	G/G	C/C	G/G
67882611	T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T
67882612	T	T/T	T/T	T/T	T/T	T/T	T/T	C/C	T/T

Table 13. Continued

POS	REF	PI226096	PI330272	Sprop-369-2	Sprop-369-1	Zengada-IBC-E-308
67884158	T	T/T	T/T	T/T	T/T	T/T
67882687	G	G/G	G/G	G/G	G/G	G/G
67882698	G	G/G	G/G	G/G	G/G	G/G
67882737	G	G/G	./.	C/C	G/C	./.
67882786	C	C/C	C/C	T/T	T/T	C/C
67882804	G	G/G	G/G	A/A	G/A	G/G
67883724	T	T/T	T/T	T/T	T/T	T/T
67883732	G	G/G	G/G	G/G	A/A	G/G
67884143	A	A/A	A/A	C/C	C/C	A/A
67884778	A	A/A	A/A	A/A	A/C	A/A
67887383	G	G/G	G/G	G/G	G/G	G/G
67884645	A	A/A	A/A	A/A	A/G	A/A
67885815	C	C/C	C/C	C/C	C/T	C/C
67887574	C	C/C	C/C	C/C	C/C	C/C
67882645	AC	AC/AC	AC/AC	A/A	A/A	AC/AC
67887675	G	G/G	T/T	G/G	G/G	T/T
67887699	A	A/A	A/A	C/C	C/C	A/A
67887732	C	C/C	C/C	T/T	T/T	C/C
67887745	A	A/A	A/A	A/A	A/G	A/A
67887794	A	A/A	A/A	G/G	A/G	A/A
67887852	C	C/C	C/C	C/C	C/T	C/C
67887886	A	A/A	A/A	A/A	T/T	A/A
67887985	AAT	AAT/AAT	AAT/AAT	./.	./.	AAT/AAT
67887993	T	TA/TA	T/T	./.	./.	T/T
67887994	T	./.	T/T	./.	./.	T/T
67888013	G	G/G	G/G	A/A	A/A	G/G
67888094	G	A/A	A/A	A/A	A/A	A/A
67888121	G	G/G	G/G	G/G	G/G	G/G
67882611	T	T/T	T/T	T/T	T/C	T/T
67882612	T	T/T	T/T	C/C	C/C	T/T

MQM Analysis

MQM analysis of data from the Hegari x 80M F2 population was employed to identify additional flowering time QTL and interactions amongst the QTL. The best model was identified to be a 3 gene model with 1 interaction. This model had the highest pLOD of 42.11 and is depicted in Figure 27. An additional QTL was identified on LG09 and an interaction was found between Ma_2 and Ma_4. Figure 28 shows the genotype and phenotype relationship at each of the identified QTL. 80M is parent A and Hegari is parent B. Genotypes containing recessive ma_2 and ma_4 have average DTF of ~50 days compared to ~150 days for genotypes that are dominant for both genes. The QTL on LG09 has significantly less effect on the overall phenotype causing a modeled decrease in DTF to ~100 days.

The interaction between Ma_2 and Ma_4 is shown in Figure 29 and Figure 30. The previous observation of Ma_2's affect being negated by ma_4 is confirmed in the model shown in Figure 29. The ma_4 line shows no significant variation in flowering time in any Ma_2 backgrounds.

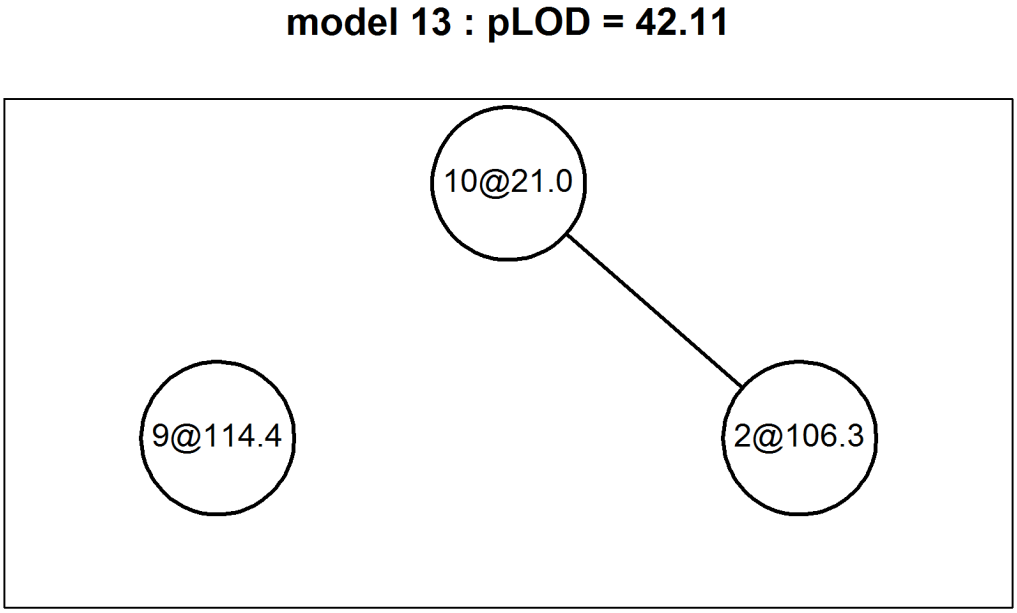

Figure 27. Best model for flowering time in the Hegari x 80M population identified via MQM. Each circle denotes an identified QTL with LG number and cM position. An interaction is denoted by connecting the circles via a line. The pLOD for this model was 42.11 which was the highest amongst the potential models scanned. This model shows three QTL impacting flowering time with one interation term between the QTL on LG02 (*Ma*$_2$) and the QTL on LG10 (*Ma*$_4$).

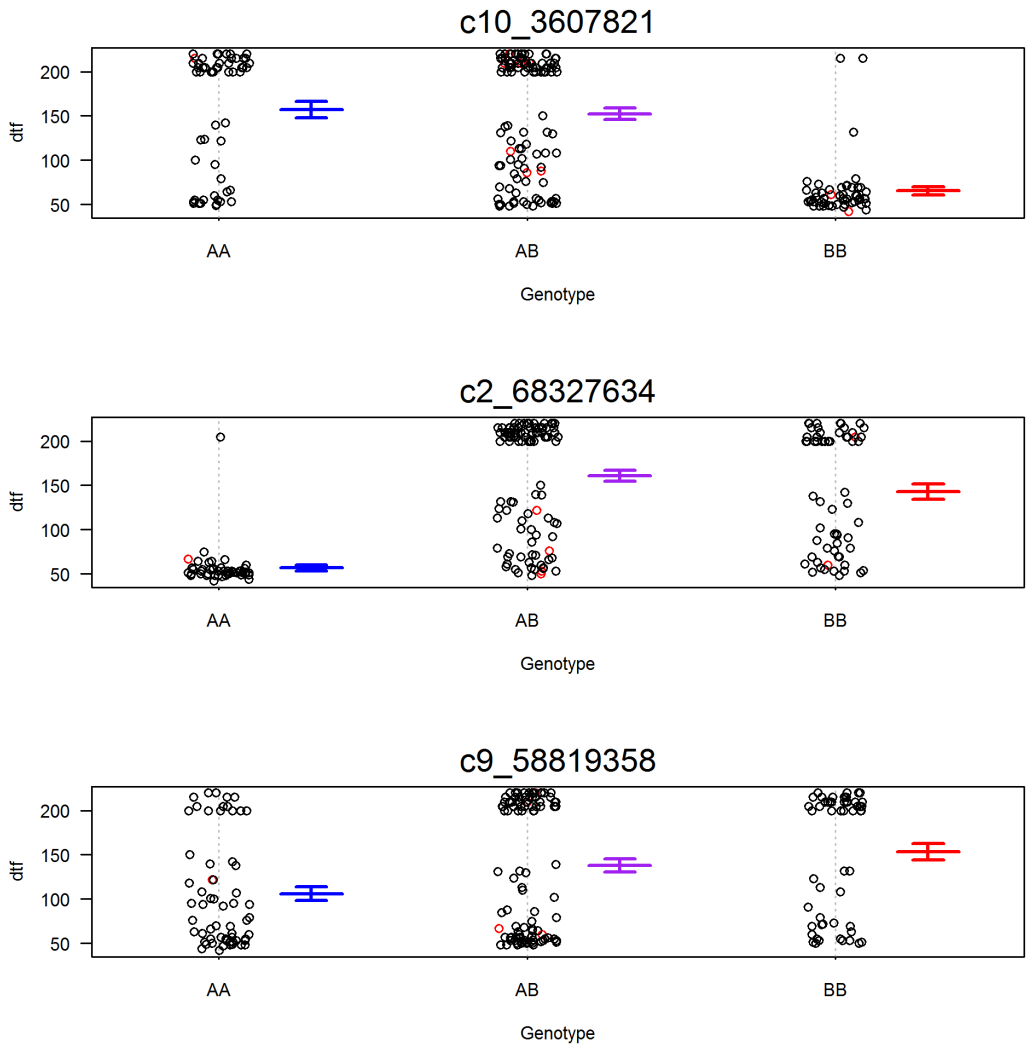

Figure 28. Genotype and Phenotype effects at each QTL identified.

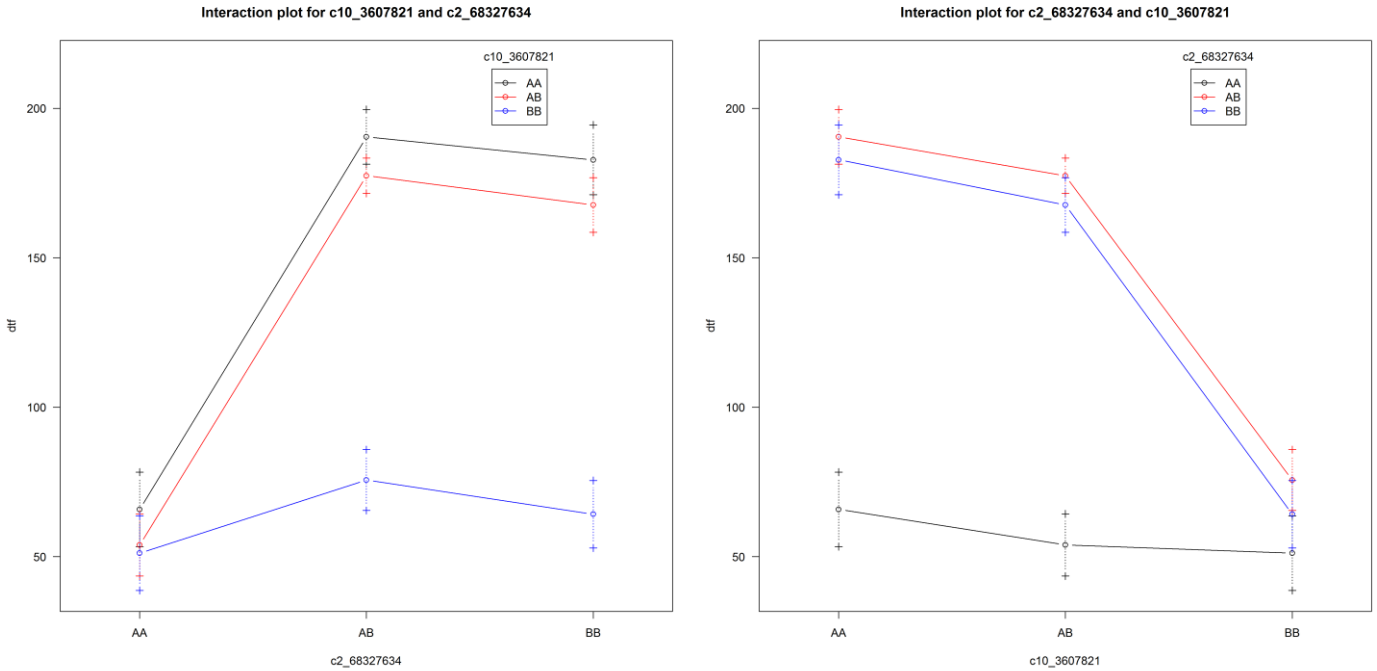

Figure 29. Interaction plots for *Ma*$_2$ and *Ma*$_4$. The graph on the left shows *Ma*$_2$ on the x-axis and flowering time on the y-axis. The different lines are the different genotypes of *Ma*$_4$. The graph on the right shows *Ma*$_4$ on the x-axis and *Ma*$_2$ genotypes as the variable. Both graphs show that if one of these maturity genes is recessive, both do not show a phenotype. When both are dominant or heterozygous, there is a large delay in flowering.

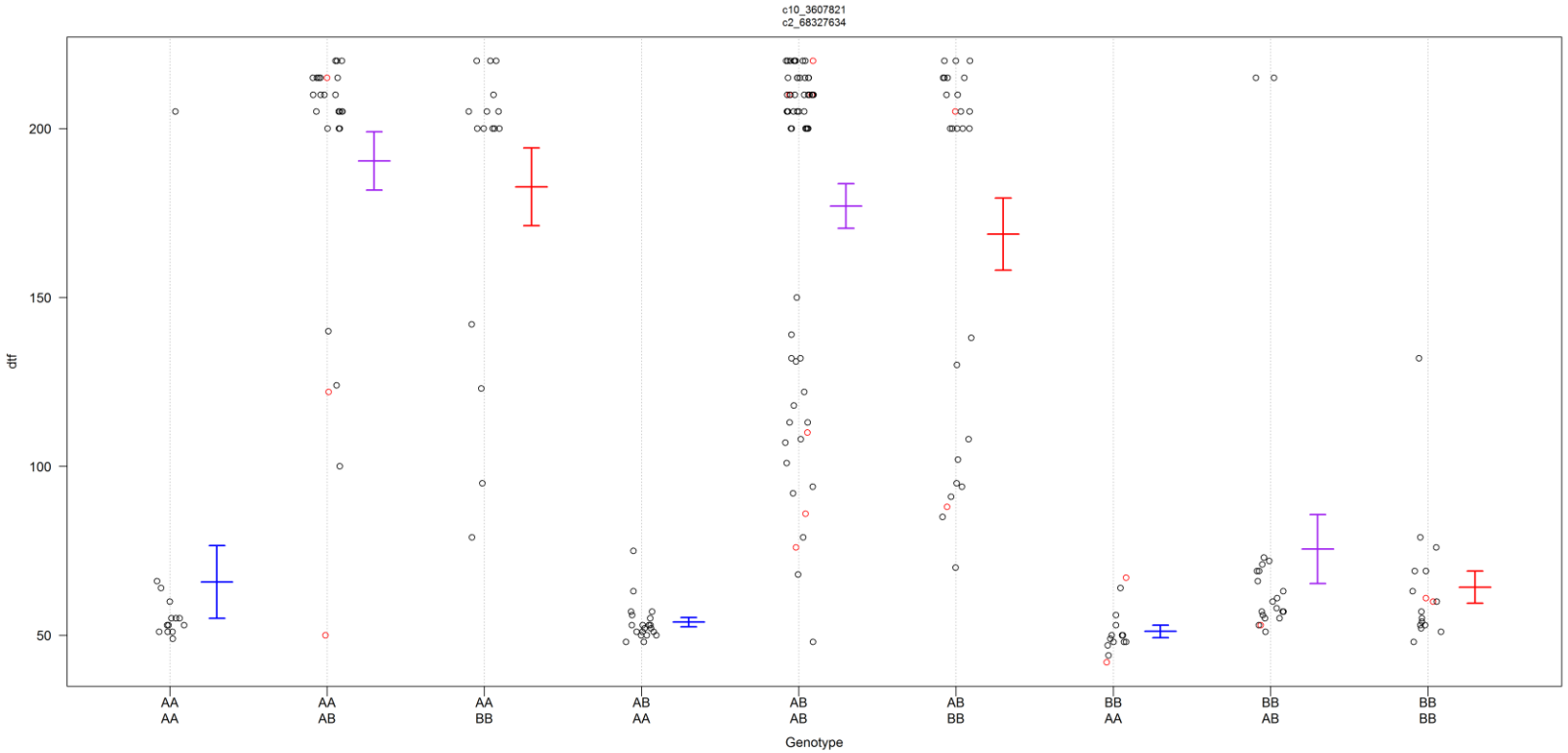

Figure 30. The genotypic combinations of *Ma*$_2$ and *Ma*$_4$ graphed showing their corresponding phenotype. This graph shows that if one of these maturity genes is recessive, both do not show a change in flowering time. When both are dominant or heterozygous, there is a large delay in flowering.

Discussion

A fine mapping population of 100M x 80M was created to identify the *Ma₂* gene; however, due to the lack of SNPs found on LG02 in this population, this proved to be an ineffective population to identify the causative gene. No SNPs were identified on the arm of LG02 encoding *Ma₂* explaining why no QTL was detected in the prior study even though the population segregated 3:1 for flowering time. In this study, a population constructed from Hegari x 80M segregating for *Ma₂* and *Ma₄* was analyzed. This population had numerous SNPs across LG02 useful for mapping *Ma₂* QTL. During the fine mapping process, the region was narrowed down to ~500kb, a region spanning ~75 genes. Fine mapping was unable to further delimit the region of interest for *Ma₂* because there was a low recombination frequency in this region of the sorghum genome in the populations used for fine mapping.

Ma₂'s identification as Sobic.002G302700.1 was initially done through analysis of whole genome sequences of 100M and 80M. No SNPs were identified in the initial 100M x 80M map across the arm of chromosome 2 where *Ma₂* is located which led to the hypothesis that there were few if any differences between 100M and 80M on this chromosome arm with the exception of *Ma₂*. This hypothesis was confirmed when only 2 SNPs were identified within the fine mapped *Ma₂* region. Only one of these SNPs was in the *ma₂* parent, 80M. This SNP leads to a stop codon causing a loss of function of *Ma₂*. Sobic.002G302700.1 is a zinc finger transcription factor with unknown function.

The results of this study show that the ability of *Ma₁* and *Ma₆* to repress flowering is significantly reduced in *ma₂ma₂* genetic backgrounds. In prior studies, it was shown that recessive *ma₂* caused early flowering in *Ma₁* backgrounds (181) and the current study extends this observation to both *Ma₁* and *Ma₆*. *Ma₂* causes late flowering at 30C when *Ma₁* is dominant but earliness when *ma₁* is recessive, especially when plants are grown at 20°C (62, 181). Also, when *ma₂* is recessive, heterozygosity at *Ma₁* causes late flowering relative to homozygosity. Since *Ma₂* modifies *Ma₁* (and *Ma₆*) action but also can act in *ma₁ma₆* null backgrounds, this factor could act down-stream of *Ma₁*. Analysis of the expression of various flowering time genes in 100M and 80M via qRT-PCR and RNAseq could provide further insight into the interaction of *Ma₂* with the other genes in the flowering time pathway.

The MQM analysis identified an additional QTL on LG09 in addition to the interaction between *Ma₂* and *Ma₄*. The interaction between *Ma₂* and *Ma₄* had been previously hypothesized due to the effect of *Ma₂*'s phenotype only being identified in a dominant *Ma₄* background, as noted during the fine mapping process. The additional QTL on LG09 shows the utility of MQM analysis since the effect of this QTL is only ~4% of the variation of flowering time identified in the Hegari x 80M F2 population. Both *Ma₂* and *Ma₄* account for >20% of the variance.

In Figure 29 and Figure 30, the interaction of *Ma₂* and *Ma₄* are explored in more detail. The *ma₄* line shows no significant variation in flowering time in any

Ma₂ backgrounds. Similarly, when *ma₂* is recessive, allelic variation in *Ma₄* does not alter flowering time indicating that the action of *Ma₄* is dependent on *Ma₂*. Additionally, the heterozygous genotype and dominant genotype for both of these QTL do not have a significant difference in phenotype; suggesting no heterozygous advantage for either of these QTL. These figures are an important visualization as it shows the trend in phenotypes, emphasizing the dramatic differences the genotypes cause in phenotype as well as showing the minimal effect of outliers on this dataset.

It has been previously determined that photoperiod alters the action of *Ma₂* on flowering time, as dominant alleles delay flowering in long days but not short days. *Ma₂* expression or activity may also be modulated by the circadian clock. It is hypothesized that *Ma₂* is downstream of *Ma₁* and *Ma₆*. It is well documented that *Ma₁* and *Ma₆* repress flowering time. *Ma₁* and *Ma₆* repress expression of the *SbCN* genes which activate the transition from vegetative to floral meristem.

Future Directions

Sequencing of Allelic Variants

Multiple mutations may have occurred in the *Ma₂* gene during the selection period of earlier flowering. Other sorghum lines that do not share a family history with 80M that have been noted to be *ma₂* include Early Kalo, Kalo, and Fargo. These lines in addition to SM80, which should have the same

mutation as identified in 80M, will be grown out and *Ma$_2$* sequenced to determine if an allelic series for *Ma$_2$* can be identified.

Analyze expression of Ma$_2$ gene to identify expression alleles and their affects in the flowering time pathway

The recessive allele of *ma$_2$* in 80M corresponds to a mutation that creates a premature stop codon. The gene identified, Sobic.002G302700.1 is a zinc finger protein which led to the hypothesis that *Ma$_2$* is a transcription factor. Previous observations and MQM analysis have noted that *Ma$_2$* only has a phenotypic effect on flowering time if *Ma$_1$* and *Ma$_4$* are in their dominant forms. Analysis of gene expression would further the understanding of *Ma$_2$* regulation and function as well as its dependency on the circadian clock and light (104). Leaf tissue was collected at 3 hour intervals for 48 hours in long days in varying temperature of either 20 or 30°C. The first 24 hours were under regular long day light conditions while the second 24 hours were under constant light. These varying conditions will provide insight into the light, circadian clock, and temperature pathways as all are major factors that affect flowering time. Expression of known genes involved in the flowering time pathway such as *PRR37, GHD7, EHD1, FT* and *ZCN8* in *Ma$_2$* and *ma$_2$* genotypes will also help determine where *Ma$_2$* acts in the flowering time regulatory network.

Materials and Methods

Fine Mapping Through Genotyping and Phenotyping

All HIFs were grown in a greenhouse in long day conditions, 14 hours light and 10 hours dark. Plants were fertilized with Peters Professional Allrounder fertilizer every 2 weeks. Flowering time phenotypes were recorded at anthesis. Plants were bagged at anthesis to ensure self-fertilization. Tissue was collected for DNA extraction (zymo) and then DG was performed at Fse1 depth (59). Individuals that could provide additional information via breakpoints were subsequently analyzed using DG at Ngo Depth. Breakpoints were analyzed to delimit the region and narrow down the search for candidate genes.

MQM Analysis

MQM was run using the Rqtl package (61). Penalty scores were calculated for heavy, moderate, and light penalties using 10,000 iterations. These penalties, along with the phenotypes, genotype map, and previously identified QTL were used in the analysis.

Allelic Variants Sequencing

100M, 80M, Fargo, Kalo, R07007, and SM80 were planted in flats in greenhouse 204A. Tissue was taken for DNA extraction (zymo). Primers were designed using 80M.

Circadian Cycling Experiment

100M and 80M were planted in greenhouse 204A under long day conditions. They were thinned to a density of 3 plants per pot at 7 days. Osmocote was mixed into the soil prior to planting for fertilization. At 30 days, half of the plants were put into a growth chamber under 14 hours of light and 10 hours of dark at 60% humidity and at 20°C while the other half is put into a growth chamber under the same conditions except for at 30°C. Samples were taken starting at the top of the 14 hour light period on day 4 of being in the growth chamber and continued every 3 hours for 48 hours. On the second day of collections, plants were exposed to continuous illumination. At each time point, the top 3 leaves from 3 different plants were pooled into 1 collection for both the 20°C and 30°C temperatures. This experiment was performed in 3 replications for each genotype. Total RNA was extracted from each collection (zymo), used for synthesis of cDNA (thermo-scientific), and analyzed using qRT-PCR.

CHAPTER V

CONCLUSIONS

Numerous QTL identified during the conduct of this research modulate grain yield or traits that have the potential to impact grain yield. QTL in one region of LG01 affected grain weight and grain number, but not overall grain yield. It was hypothesized that this QTL could be caused by alleles of a single gene that affects grain number and that changes in grain number affect grain weight. This hypothesis is consistent with the observation that the parental allele from BTx642 that decreases grain number increases grain weight. The lack of a QTL for grain yield at this location on LG01 is also consistent with this hypothesis. Moreover, the grain number/weight QTL on LG01 affected grain number more consistently than grain weight across years and environments suggesting stronger genetic control for grain number by this locus. Therefore, it may be beneficial to fix the allele for increased grain number in grain crop genotypes. However, in the BTx623 x IS3620c population an aligned QTL on LG01 affected grain number and yield, but not grain weight. If the genes/alleles that modulate grain number, weight and yield in this region of LG01 in the two populations are the same; this suggests that two loci may be involved.

A QTL for grain weight and grain yield was identified on LG03 in the populations studied across multiple years and environments. There was no corresponding QTL for grain number in this region of the genome. QTL for grain composition aligned with the grain weight QTL on LG03 including those for

protein, moisture, fat, fiber, starch, phenol, and tannins phenotypes. These findings illustrate the importance of grain weight to overall grain yield and suggest that differences in grain composition could affect or be correlated with grain weight. I hypothesized that the grain yield QTL and the grain composition QTL on LG03 are the result of the same causative allele; therefore a change in grain composition has the potential to impact the overall grain yield. This QTL on LG03 is an example of how phenotyping subcomponents of grain yield can lead to a better understanding of the underlying cause of QTL identified in grain yield. Fine mapping of this QTL on LG03 would be the next step. Fine mapping should show quick results as for each fine mapping round, phenotypes could be done of overall grain yield, grain weight, and grain composition to delimit the region.

QTL for changes in stem biomass that provide insight into remobilization did not overlap with QTL identified for grain composition. However, a QTL on LG03 was aligned with a QTL that modulates stem protein content. This could be the result of protein remobilization from the stem to the grain. Further study into this pathway will provide a better understanding of how the plant uses its resources to optimum efficiency enabling grain weight to be maximized. Looking into photosynthetic rate of different sorghum lines that vary in stem hollowing formation would also be informative into this process. I would hypothesize that ongoing photosynthesis plays a larger role in grain filling than remobilization based off the data seen in this dissertation; however, I would hypothesize that

the photosynthesis rate is higher in plants with no stem hollowing to compensate for these nutrients.

Many QTL that contribute to the variation of overall grain yield were identified in this study of grain sorghum populations. A logical next step in this research is to identify the genes underlying the QTL. This will enable better understanding of the biochemical pathways that affect grain yield in sorghum. A QTL on LG01 was identified in multiple populations that affected grain number, weight, and subcomponents of grain yield. Further analysis of QTL that map to this region of LG01 could provide insight into the tradeoff between grain weight and grain number. This region is already relatively small (600kb); however, one limitation to consider in the positions in this data is that the BTx642 x Tx7000 population is only 91 RILS. This causes some of these cut offs to be based off of one RIL's breakpoints. Typically in fine mapping, repetition of breakpoint analysis is needed in order to confirm the cut off and feel confident in the delimited region still containing the gene of interest. HIFs need to be created in order to have a larger number of lines to consider with more breakpoints in the region of interest. Preferably, these lines will be isogenic at many of the other known grain yield QTL identified in these studies so the observed affects can be attributed to the gene of interest.

Flowering time is an important trait that affects grain yield. Overall grain yield is positively correlated with longer flowering times (39); however, in temperate climates, grain maturation needs to occur before frost or onset of

adverse environmental conditions during the growing season (i.e., drought, heat, insect pressure). As a consequence, earlier flowering times are often associated with improved grain yield (39). In the U.S., this led to the identification of six maturity loci, Ma_1-Ma_6, which when recessive, resulted in an earlier flowering phenotype in long days (38, 39, 62). Many of these maturity genes have been identified and characterized.

Ma2 was originally identified through the study of populations derived from Early White Milo ($ma_1Ma_2Ma_3Ma_4Ma_5ma_6$) and Dwarf Yellow Milo ($Ma_1ma_2ma_3Ma_4Ma_5ma_6$) (181). 80M and 100M lines derived from this population differed in Ma_2 alleles. A population derived from crossing 80M and 100M segregated 3:1 for flowering time as expected. Genetic maps based on this population lacked most of LG02 and did not contain a QTL corresponding to Ma_2 suggesting that the mutation causing ma_2 occurred in a gene within the nearly isogenic region of LG02. In this study, a population was constructed from Hegari x 80M that segregated for Ma_2 and Ma_4. Analysis of HIFs derived from this population narrowed the interval encoding Ma_2 to ~500kb, a region spanning ~75 genes. Comparison of whole genome sequences derived from 100M and 80M across the Ma_2 locus identified a SNP mutation in the gene Sobic.002G302700 in 80M resulting in a stop codon and loss of gene function. Based on these results, Sobic.002G302700, a gene encoding a zinc finger transcription factor, is tentatively identified as Ma_2.

In prior studies, it was shown that recessive *ma₂* caused early flowering in *Ma₁* backgrounds (181) and the current study showed that recessive *ma₂* causes early flowering in *Ma₁*, *Ma₆* and *Ma₁Ma₆* genetic backgrounds. *Ma₂* causes late flowering at 30°C when *Ma₁* is dominant but earliness when *ma₁* is recessive, especially when plants are grown at 20°C (62, 181). Also, when *ma₂* is recessive, heterozygosity at *Ma₁* causes late flowering relative to homozygosity. Since *Ma₂* modifies the action of *Ma₁* (and *Ma₆*) but also can act in *ma₁ma₆* null backgrounds, it is hypothesized that *Ma₂* acts down-stream of *Ma₁* and *Ma₆*. Analysis of the expression of genes in the sorghum flowering pathway in 100M and 80M via qRT-PCR and RNAseq could provide further insight into the interaction of *Ma₂* with the other genes in the flowering time pathway.

In summary, the identification of numerous QTL that modify grain yield, and in depth understanding of the biochemical pathways modulated by these QTL, could provide information useful for increasing the yield of grain sorghum.

REFERENCES

1. Dahlberg J, Berenji J, Sikora V, Latković D (2011) Assessing sorghum [Sorghum bicolor (L) Moench] germplasm for new traits: Food, fuels & unique uses. *Maydica* 56(2):85–92.
2. Murray SC, et al. (2008) Genetic improvement of sorghum as a biofuel feedstock: I. QTL for stem sugar and grain nonstructural carbohydrates. *Crop Sci* 48(6):2165–2179.
3. Han KJ, et al. (2012) Agronomic Considerations for Sweet Sorghum Biofuel Production in the South-Central USA. *Bioenergy Res* 5(3):748–758.
4. Mann JA, Kimber CT, Miller FR (1983) *The Origin and Early Cultivation of Sorghums in Africa* (Texas Agricultural Experiment Station, College Station).
5. Zhao YL, et al. (2009) Biomass yield and changes in chemical composition of sweet sorghum cultivars grown for biofuel. *F Crop Res* 111(1):55–64.
6. Taylor JRN, Schober TJ, Bean SR (2006) Novel food and non-food uses for sorghum and millets. *J Cereal Sci* 44(3):252–271.
7. Doggett H (1988) *Sorghum* (Longman Scientific & Technical).
8. Alexandratos N, Bruinsma J (2012) World agriculture towards 2030/2050. *Land use policy* 20(4):375.
9. Rosegrant M, van Boekel M a. JS, Hubert B, Ortiz R (2010) The Future of Food: Scenarios for 2050. *Crop Sci* 50(Supplement 1):S--33--S--50.

10. Melorose J, Perroy R, Careas S (2015) World Population Prospects: The 2015 Revision, Key Findings and Advance Tables. Working Paper No. ESA/P/WP.241. *United Nations, Dep Econ Soc Aff Popul Div* 1:1–59.
11. Floros JD, et al. (2010) Feeding the world today and tomorrow: The importance of food science and technology. *Compr Rev Food Sci Food Saf* 9(5):572–599.
12. Summary E (2009) FAO's Director-general on how to feed the world in 2050. *Popul Dev Rev* 35(4):837–839.
13. Thornton PK (2010) Livestock production: recent trends, future prospects. *Philos Trans R Soc B Biol Sci* 365(1554):2853–2867.
14. Ray DK, Mueller ND, West PC, Foley JA (2013) Yield Trends Are Insufficient to Double Global Crop Production by 2050. *PLoS One* 8(6).
15. Paterson AH, et al. (2009) The Sorghum bicolor genome and the diversification of grasses. *Nature* 457(7229):551–556.
16. Paterson AH, H. A (2008) Genomics of sorghum. *Int J Plant Genomics* 2008:362451.
17. Chaw S-M, Chang C-C, Chen H-L, Li W-H (2004) Dating the monocot-dicot divergence and the origin of core eudicots using whole chloroplast genomes. *J Mol Evol* 58(4):424–441.
18. Kellogg EA (2001) Update on Evolution Evolutionary History of the Grasses 1. 125(March):1198–1205.
19. Sage RF (2004) The evolution of C4 photosynthesis. 341–370.

20. Ueno O (1998) Induction of kranz anatomy and C4-like biochemical characteristics in a submerged amphibious plant by abscisic acid. *Plant Cell* 10(4):571–584.

21. Hibberd JM, Sheehy JE, Langdale J a (2008) Using C4 photosynthesis to increase the yield of rice-rationale and feasibility. *Curr Opin Plant Biol* 11(2):228–231.

22. Parry M a J, et al. (2011) Raising yield potential of wheat. II. Increasing photosynthetic capacity and efficiency. *J Exp Bot* 62(2):453–467.

23. Shi J, et al. (2009) Unraveling the complex trait of crop yield with quantitative trait loci mapping in Brassica napus. *Genetics* 182(3):851–861.

24. Stephens JC, Holland RF (1954) Cytoplasmic Male-Sterility For Hybrid Sorghum Seed Production1. *Agron J* 46(1):20.

25. Fehr WR, et al. (1981) Genetic Contributions to Yield Gains of Five Major Crop Plants. (7).

26. Perez-Prat E, van Lookeren Campagne MM (2002) Hybrid seed production and the challenge of propagating male-sterile plants. *Trends Plant Sci* 7(5):199–203.

27. Wright H (1980) Commercial hybrid seed production. *Hybrid Crop Plants*:161–176.

28. Xiao J, Li J, Yuan L, Tanksley SD (1995) Dominance is the major genetic basis of heterosis in rice as revealed by QTL analysis using molecular

markers. *Genetics* 140(2):745–754.

29. Hochholdinger F, Hoecker N (2007) Towards the molecular basis of heterosis. *Trends Plant Sci* 12(9):427–432.
30. Robinson AHF, Comstock RE, Khalil A, Harvey PH (2013) Dominance Versus Over-Dominance in Heterosis : Evidence from Crosses between Open- Pollinated Varieties of Maize. 90(851):127–131.
31. Gupta PK, Rustgi S, Kumar N (2006) Genetic and molecular basis of grain size and grain number and its Relevance to Grain Productivity in Higher Plants. *Genome* 49(6):565–571.
32. Sinclair TR, Jamieson PD (2006) Grain number, wheat yield, and bottling beer: An analysis. *F Crop Res* 98(1):60–67.
33. Fischer R a. (2008) The importance of grain or kernel number in wheat: A reply to Sinclair and Jamieson. *F Crop Res* 105(1–2):15–21.
34. Gerik T, Bean B, Vanderlip R (1993) *Sorghum Growth and Development*.
35. Li H, Freeling M, Lisch D (2010) Epigenetic reprogramming during vegetative phase change in maize. 107(51).
36. Sinclair TR, Horie T (1989) Leaf Nitrogen, Photosynthesis, and Crop Radiation Use Efficiency: A Review. *Crop Sci* 29(1):90.
37. Quinby JR (1966) Fourth Maturity Gene Locus in Sorghum. 516–518.
38. Quinby JR, Karper R (1945) The inheritance of three genes that influence time of floral initiation and maturity date in Milo. *J Am Soc Agron* 37(901):916–936.

39. Rooney WL, Aydin S (1999) Genetic Control of a Photoperiod-Sensitive Response in Sorghum bicolor (L.) Moench. (Table 1):3–6.

40. Wu Y, et al. (2016) The QTL GNP1 Encodes GA20ox1, Which Increases Grain Number and Yield by Increasing Cytokinin Activity in Rice Panicle Meristems. *PLoS Genet* 12(10):e1006386.

41. Vollbrecht E, Springer PS, Goh L, Buckler ES, Martienssen R (2005) Architecture of floral branch systems in maize and related grasses. *Nature* 436(7054):1119–1126.

42. Ogunlela VB, Eastin JD (1984) Effect of Elevated Night Temperature During Panicle Development on Sorghum (Sorghum bicolor L.) Yield Components. *Cereal Res Commun* 12(3):245–251.

43. Yang Z, van Oosterom EJ, Jordan DR, Doherty A, Hammer GL (2010) Genetic Variation in Potential Kernel Size Affects Kernel Growth and Yield of Sorghum. *Crop Sci* 50(2):685–695.

44. Yang Z, van Oosterom EJ van, Jordan DR, Hammer GL (2009) Pre-anthesis ovary development determines genotypic differences in potential kernel weight in sorghum. *J Exp Bot* 60(4):1399–1408.

45. Burow G, Xin Z, Hayes C, Burke J (2014) Characterization of a Multiseeded () Mutant of Sorghum for Increasing Grain Yield. *Crop Sci* 54(5):2030.

46. Tolk JA, Schwartz RC (2017) Do More Seeds per Panicle Improve Grain Sorghum Yield? *Crop Sci* 57:490–496.

47. Peltonen-Sainio P, Kangas A, Salo Y, Jauhiainen L (2007) Grain number dominates grain weight in temperate cereal yield determination: Evidence based on 30 years of multi-location trials. *F Crop Res* 100(2):179–188.

48. Xing Y, Zhang Q (2010) Genetic and molecular bases of rice yield. *Annu Rev Plant Biol* 61:421–442.

49. Serrago RA, Alzueta I, Savin R, Slafer GA (2013) Understanding grain yield responses to source–sink ratios during grain filling in wheat and barley under contrasting environments. *F Crop Res* 150:42–51.

50. Rogers GS, Milham PJ, Gillings M, Conroy JP (1996) Sink Strength May Be the Key to Growth and Nitrogen Responses in N-Deficient Wheat at Elevated CO2. *Aust J Plant Physiol* 23(3):253–264.

51. Ho LC (1988) Metabolism and Compartmentation of Imported Sugars in Sink Organs in Relation to Sink Strength. *Annu Rev Plant Physiol Plant Mol Biol* 39:355–378.

52. Yoshida S (1972) Physiological Aspects of Grain Yield. *Annu Rev Plant Physiol* 23(1):437–464.

53. Girin T, et al. (2014) Brachypodium: A promising hub between model species and cereals. *J Exp Bot* 65(19):5683–5686.

54. Guillon F, et al. (2012) A comprehensive overview of grain development in Brachypodium distachyon variety Bd21. *J Exp Bot* 63(2):739–755.

55. Yang J, Zhang J (2005) Grain Filling of Cereals Under Soil Drying. *New Phytol* 169(2):223–236.

56. Zhang Z, et al. (2013) Genetic Analysis of Grain Filling Rate Using Conditional QTL Mapping in Maize. *PLoS One* 8(2):e56344.

57. Lander ES, Botsteins 'b D Mapping Mendelian Factors Underlying Quantitative Traits Using RFLP Linkage Maps.

58. Wang S, Basten CJ, Zeng ZB (2005) Windows QTL cartographer version 2.5. *Stat Genet*.

59. Morishige DT, et al. (2013) Digital genotyping of sorghum -- a diverse plant species with a large repeat-rich genome. *BMC Genomics* 14(1):448.

60. Lander ES, et al. (1987) MAPMAKER: An interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics* 1(2):174–181.

61. Broman KW, Wu H, Sen S, Churchill GA (2003) R/qtl: QTL mapping in experimental crosses. *Bioinforma Appl NOTE* 19(7):889–890.

62. Quinby JR (1966) Fourth maturity gene locus in sorghum. *Crop Sci* 6(6):516–518.

63. Childs KL, et al. (1997) The Sorghum Photoperiod Sensitivity Gene, Ma3, Encodes a Phytochrome B. *Plant Physiol* 113(2):611–619.

64. Somers DE, Sharrock RA, Tepperman JM, Quail PH (1991) The hy3 Long Hypocotyl Mutant of Arabidopsis Is Deficient in Phytochrome B. *Plant Cell* 3(12):1263–1274.

65. Reed JW, Nagatani A, Elich TD, Fagan M, Chory J (1994) Phytochrome A and Phytochrome B Have Overlapping but Distinct Functions in

Arabidopsis Development. *Plant Physiol* 104(4):1139–1149.

66. Reed JW, Nagpal P, Poole DS, Furuya M, Chory J (1993) Mutations in the gene for the red/far-red light receptor phytochrome B alter cell elongation and physiological responses throughout Arabidopsis development. *Plant Cell* 5(2):147–157.

67. Murphy RL, et al. (2014) Ghd7 (Ma6) Represses Sorghum Flowering in Long Days: Alleles Enhance Biomass Accumulation and Grain Production. *Plant Genome* 7(2):0.

68. Yang S, et al. (2014) Sorghum phytochrome B inhibits flowering in long days by activating expression of SbPRR37 and SbGHD7, repressors of SbEHD1, SbCN8 and SbCN12. *PLoS One* 9(8).

69. Li L, Li X, Liu Y, Liu H (2016) Flowering responses to light and temperature. *Sci China Life Sci* 59(4):403–408.

70. Möller-Steinbach Y, Alexandre C, Hennig L (2010) Flowering Time Control. *Methods in Molecular Biology (Clifton, N.J.)*, pp 229–237.

71. Srikanth A, Schmid M (2011) Regulation of flowering time: all roads lead to Rome. *Cell Mol Life Sci* 68(12):2013–2037.

72. Fowler S, et al. (1999) GIGANTEA: a circadian clock- controlled gene that regulates photoperiodic flowering in Arabidopsis and encodes a protein with several possible membrane- spanning domains. *EMBO J* 18(17):4679–4688.

73. Hamner KC, Bonner J (1938) Photoperiodism in Relation to Hormones as

Factors in Floral Initiation and Development. *Bot Gaz* 100(2):338–431.

74. Hamner KC (1940) Interrelation of light and darkness in photoperiodic induction. *Bot Gazette1* 101(3):658–687.

75. Song YH, Ito S, Imaizumi T (2010) Similarities in the circadian clock and photoperiodism in plants. *Curr Opin Plant Biol* 13(5):594–603.

76. Jeong S, Clark SE (2005) Photoperiod regulates flower meristem development in Arabidopsis thaliana. *Genetics* 169(2):907–15.

77. Hill CB, Li C (2016) Genetic Architecture of Flowering Phenology in Cereals and Opportunities for Crop Improvement. *Front Plant Sci* 7(December):1–23.

78. Cober E, Curtis D, Stewart D, Morrison M (2014) Quantifying the Effects of Photoperiod, Temperature and Daily Irradiance on Flowering Time of Soybean Isolines. *Plants* 3(4):476–497.

79. Bünning E (1960) Circadian Rhythms and the Time Measurement in Photoperiodism. *Cold Spring Harb Symp Quant Biol* 25:249–256.

80. Nozue K, et al. (2007) Rhythmic growth explained by coincidence between internal and external cues. *Nature* 448(7151):358–361.

81. Panda S, et al. (2002) Coordinated Transcription of Key Pathways in the Mouse by the Circadian Clock. *Cell* 109(3):307–320.

82. Gekakis N, et al. (1998) Role of the CLOCK Protein in the Mammalian Circadian Mechanism. *Science* 280(5369).

83. Berson DM, Dunn FA, Takao M (2002) Phototransduction by Retinal

Ganglion Cells That Set the Circadian Clock. *Science* 295(5557).

84. Harmer SL, et al. (2000) Orchestrated Transcription of Key Pathways in Arabidopsis by the Circadian Clock. *Science* 290(5499).

85. Snaith PJ, Mansfield TA (1986) The Circadian Rhythm of Stomatal Opening: Evidence for the Involvement of Potassium and Chloride Fluxes. *J Exp Bot* 37(2):188–199.

86. Atkins KA, Dodd AN (2014) Circadian regulation of chloroplasts. *Curr Opin Plant Biol* 21:43–50.

87. Ruby NF, et al. (2002) Role of Melanopsin in Circadian Responses to Light. *Science* 298(5601):2211–2213.

88. Bretzl H (1903) Botanische forschungen des Alexanderzuges. *Königliche Gesellschaft der Wissenschaften zu Göttingen*.

89. Daan S (2010) A History of Chronobiological Concepts. *The Circadian Clock*, ed Albrecht U (Springer New York, New York, NY), pp 1–35.

90. McClung CR (2006) Plant circadian rhythms. *Plant Cell* 18(4):792–803.

91. De Marian JJ. (1729) Observation Botanique. *Hist l'Academie R des Sci Paris*:35.

92. Darwin C, Darwin F (1880) *The Power of Movement in Plants* (London). 1st Ed.

93. Millar A, Carre I, Strayer C, Chua N, Kay S (1995) Circadian clock mutants in Arabidopsis identified by luciferase imaging. *Science* 267(5201).

94. Y N, et al. (2007) Genetic linkages of the circadian clock-associated

genes, TOC1, CCA1 and LHY, in the photoperiodic control of flowering time in Arabidopsis thaliana. *Plant Cell Physiol* 48(7):925–937.

95. Somers DE, Webb AA, Pearson M, Kay SA (1998) The short-period mutant, toc1-1, alters circadian clock regulation of multiple outputs throughout development in Arabidopsis thaliana. *Development* 125(3):485–94.

96. Harmon FG, Imaizumi T, Kay SA (2007) The Plant Circadian Clock: Review of a Clockwork Arabidopsis. *Annual Plant Reviews Volume 21: Endogenous Plant Rhythms* (Blackwell Publishing Ltd).

97. Imaizumi T, Schultz TF, Harmon FG, Ho LA, Kay SA (2005) FKF1 F-Box Protein Mediates Cyclic Degradation of a Repressor of CONSTANS in Arabidopsis. *Science* 309(5732):293–297.

98. Fornara F, et al. (2009) Arabidopsis DOF Transcription Factors Act Redundantly to Reduce CONSTANS Expression and Are Essential for a Photoperiodic Flowering Response. *Dev Cell* 17(1):75–86.

99. Turck F, Fornara F, Coupland G (2008) Regulation and Identity of Florigen: FLOWERING LOCUS T Moves Center Stage. *Annu Rev Plant Biol* 59:579–594.

100. Imaizumi T, Kay SA (2006) Photoperiodic control of flowering: not only by coincidence. *Trends Plant Sci* 11(11):550–558.

101. Liu B, Zuo Z, Liu H, Liu X, Lin C (2011) Arabidopsis cryptochrome 1 interacts with SPA1 to suppress COP1 activity in response to blue light.

Genes Dev 25(10):1029–1034.

102. Ishikawa M, Kiba T, Chua N-H (2006) The Arabidopsis *SPA1* gene is required for circadian clock function and photoperiodic flowering. *Plant J* 46(5):736–746.

103. Nuñez F, Yamada T (2017) Molecular Regulation of Flowering Time in Grasses. *Agronomy* 7(1):17.

104. Murphy RL, et al. (2011) Coincident light and clock regulation of pseudoresponse regulator protein 37 (PRR37) controls photoperiodic flowering in sorghum. *Proc Natl Acad Sci U S A* 108(39):16469–16474.

105. Boden SA, et al. (2013) Thermal stress effects on grain yield in Brachypodium distachyon occur via H2A.Z-nucleosomes. *Genome Biol* 14(6):R65.

106. Muchow RC, Wilson GL (1976) Photosynthetic and storage limitations to yield in Sorghum bicolor (L. Moench). *Aust J Agric Res* 27(4):489–500.

107. Harsant J, Pavlovic L, Chiu G, Sultmanis S, Sage TL (2013) High temperature stress and its effect on pollen development and morphological components of harvest index in the C3 model grass Brachypodium distachyon. *J Exp Bot* 64(10):2971–2983.

108. Gallagher JN, Biscoe PV, Scott RK (1975) Barley and its Environment. V. Stability of Grain Weight. *J Appl Ecol* 12(1):319–336.

109. Fischer RA, et al. (1997) Wheat Yield Progress Associated with Higher Stomatal Conductance and Photosynthetic Rate, and Cooler Canopies.

Crop Sci 38(6):1467–1475.

110. Lawlor DW (2002) Carbon and nitrogen assimilation in relation to yield: mechanisms are the key to understanding production systems. *J Exp Bot* 53(370):773–787.
111. Araus J., Slafer GA, Royo C, Serret MD (2008) Breeding for Yield Potential and Stress Adaptation in Cereals. *CRC Crit Rev Plant Sci* 27(6):277–412.
112. Dolferus R, Ji X, Richards RA (2011) Abiotic stress and control of grain number in cereals. *Plant Sci* 181(4):331–341.
113. Khanday I, Yadav SR, Vijayraghavan U (2013) Rice LHS1/OsMADS1 controls floret meristem specification by coordinated regulation of transcription factors and hormone signaling pathways. *Plant Physiol* 161(4):1970–83.
114. Morris GP, et al. (2013) Population genomic and genome-wide association studies of agroclimatic traits in sorghum. *Proc Natl Acad Sci U S A* 110(2):453–458.
115. Elgersma A, Śnieżko R (1988) Cytology of seed development related to floret position in perennial ryegrass (Lolium perenne L.). *Euphytica* Volume 39:59–68.
116. Chen FQ, Hayes PM (1989) A comparison of Hordeum bulbosum-mediated haploid production efficiency in barley using in vitro floret and tiller culture. *Theor Appl Genet* 77(5):701–704.

117. Müntzing A, Tometorp G, Munt-Petersen K (1937) Tetraploid Barley Produced by Heat Treatment. *Hereditas* 22(3):401–406.

118. Miralles DJ, Richards RA, Slafer GA (2000) Duration of the stem elongation period influences the number of fertile florets in wheat and barley. *Funct Plant Biol* 27(10):931–940.

119. Borrill P, Fahy B, Smith AM, Uauy C (2015) Wheat grain filling is limited by grain filling capacity rather than the duration of flag leaf photosynthesis: A case study using NAM RNAi plants. *PLoS One* 10(8):1–14.

120. Wang G, Kang MS, Moreno O (1999) Genetic analyses of grain-filling rate and duration in maize. *F Crop Res* 61(3):211–222.

121. Hammer G, I.J. B (2002) Genotype and Environment Effects on Dynamics of Harvest Index during Grain Filling in Sorghum. *Agron Journal2* 95(1):199–206.

122. Yadav SK, Lakshmi G (2005) Influence of water defict at vegetative, anthesis and grain filling stages on water relation and grain yield in sorghum. *Indian J Plant Physiol* 10(1):20–24.

123. Vanderlip RL, et al. (1998) Grain sorghum production handbook. *Grain sorghum Prod Handb*:1–18.

124. Henry RJ, Kettlewell RS (1996) *Cereal Grain Quality* (Chapman & Hall, London, UK). 1st Ed.

125. Short HL, Epps EAJ (1976) Nutrient Quality and Digestibility of Seeds and Fruits from Southern Forests. *J Wildl Manage* 40(2):283–289.

126. He P, et al. (1999) Genetic analysis of rice grain quality. *Theor Appl Genet* 98(3):502–508.

127. Mazur B, Krebbers E, Tingey S (1999) Gene Discovery and Product Development for Grain Quality Traits. *Science* 285(5426).

128. Guimarães IG, Pezzato LE, Barros MM, Tachibana L (2008) Nutrient digestibility of cereal grain products and by-products in extruded diets for nile tilapia. *J World Aquac Soc* 39(6):781–789.

129. Black JL (2001) Selection and storage of cereal grains for livestock. *Proceedings 51st Australian Cereal Chemistry Conference*, pp 62–65.

130. Mangal V, Ghorade RB, Thakare DP, Kalpande V V (2017) Identification of Promising Parents for Grain Yield and Early Maturity in Kharif Sorghum (Sorghum bicolor (L .) Moench). *Int J Curr Microbiol Appl Sci* 6(3):1131–1136.

131. Benech Arnold R., Fenner M, Edwards PJ (1991) Changes in germinability, ABA content and ABA embryonic sensitivity in developing seeds of Sorghum bicolor (L.) Moench. induced by water stress during grain filling. *New Phytol* 118(2):339–347.

132. Schnable PS, et al. (2009) The B73 Maize Genome: Complexity, Diversity, and Dynamics. *Science* 326(5956).

133. Sanchez AC, Subudhi PK, Rosenow DT, Nguyen HT (2002) Mapping QTLs associated with drought resistance in sorghum (Sorghum bicolor L. Moench). *Plant Mol Biol* 48(5–6):713–726.

134. Khan S, Gabriel HF, Rana T (2008) Standard precipitation index to track drought and assess impact of rainfall on watertables in irrigation areas. *Irrig Drain Syst* 22:159–177.

135. Foyer CH, Zhang H (2011) *Nitrogen metabolism in plants in the post-genomic era* (Wiley-Blackwell).

136. Ghannoum O, et al. (2005) Faster Rubisco Is the Key to Superior Nitrogen-Use Efficiency in NADP-Malic Enzyme Relative to NAD-Malic Enzyme C4 Grasses. *PLANT Physiol* 137(2):638–650.

137. Cui Z, et al. (2009) Interaction between genotypic difference and nitrogen management strategy in determining nitrogen use efficiency of summer maize. *Plant Soil* 317(1):267–276.

138. Debaeke P, et al. (1996) Grain nitrogen content of winter bread wheat (Triticum aestivum L.) as related to crop management and to the previous crop. *Eur J Agron* 5(3–4):273–286.

139. Myers RJK, Foale MA, Smith FW, Ratcliff D (1987) Tissue concentration of nitrogen and phosphorus in grain sorghum. *F Crop Res* 17(3–4):289–303.

140. Borrell A, Hammer G, van Oosterom EJ (2001) Stay-green: A consequence of the balance between supply and demand for nitrogen during grain filling? *Ann Appl Biol* 138(1):91–95.

141. Brkljacic J, et al. (2011) Brachypodium as a Model for the Grasses: Today and the Future. *Plant Physiol* 157(1):3–13.

142. Downes R (1972) Effect of temperature on the phenology and grain yield of *Sorghum bicolor*. *Aust J Agric Res* 23(4):585.

143. Gupta, Pushpendra Kumar, Varshney R ed. (2004) *Cereal Genomics II* (Springer Netherlands). 2nd Ed.

144. Kärkkäinen K, Savolainen O, Koski V (1999) Why do plants abort so many developing seeds: bad offspring or bad maternal genotypes? *Evol Ecol* 13(3):305–317.

145. Jorge Castro, Gómez JM, García D, Zamora R, Hódar JA (1999) Seed predation and dispersal in relict Scots pine forests in southern Spain. *Plant Ecol* 145(1):115–123.

146. Sarvas R (1962) Investigations on the flowering and seed crop of Pinus silvestris. *Metsat tutkimuslaitoksen Julk* 53(4).

147. Klein A-M, et al. (2007) Importance of pollinators in changing landscapes for world crops. *Proc R Soc London B Biol Sci* 274(1608).

148. Paterson AH, et al. (2009) The Sorghum bicolor genome and the diversification of grasses. *Nature* 457(7229):551–556.

149. Tang H, et al. (2013) Seed shattering in a wild sorghum is conferred by a locus unrelated to domestication. *Proc Natl Acad Sci U S A* 110(39):15824–15829.

150. Konishi S, et al. (2006) An SNP Caused Loss of Seed Shattering During Rice Domestication. *Science* 312(5778).

151. Cai H-W, Morishima H (2000) Genomic regions affecting seed shattering

and seed dormancy in rice. *Theor Appl Genet* 100(6):840–846.

152. Rakshit S, Wang Y-H (2016) *The Sorghum Genome* (Springer International Publishing).
153. Berenji J, Dahlberg J, Sikora V, Latkovi D (2011) Origin, History, Morphology, Production, Improvement, and Utilization of Broomcorn [Sorghum bicolor (L.) Moench] in Serbia. *Econ Bot* 65(2):190–208.
154. Li S, et al. (2009) *Short panicle1* encodes a putative PTR family transporter and determines rice panicle size. *Plant J* 58(4):592–605.
155. Singh VK, Singh AK, Chand R, Singh BD (2012) Genome wide analysis of disease resistance mlo gene family in sorghum [sorghum bicolor (l.) Moench].
156. Arends D, Prins P, Jansen RC, Broman KW (2010) R/qtl: high-throughput multiple QTL mapping: Fig. 1. *Bioinformatics* 26(23):2990–2992.
157. Can ND, Yoshida T (1999) Grain Yield of Sorghum Cultivars in a Double Cropping System. *Plant Prod Sci* 2(2):121–124.
158. Richards RA (2000) Selectable traits to increase crop photosynthesis and yield of grain crops. *J Exp Bot* 51(90001):447–458.
159. Borrell AK, Hammer GL, Douglas ACL (2000) Does maintaining green leaf area in sorghum improve yield under drought? I. Leaf growth and senescence. *Crop Sci* 40(4):1026–1037.
160. Antonopoulou G, Gavala HN, Skiadas I V., Angelopoulos K, Lyberatos G (2008) Biofuels generation from sweet sorghum: Fermentative hydrogen

production and anaerobic digestion of the remaining biomass. *Bioresour Technol* 99(1):110–119.

161. Acciaresi HA, Guiamet JJ (2010) Below- and above-ground growth and biomass allocation in maize and Sorghum halepense in response to soil water competition. *Weed Res* 50(5):481–492.
162. Piper JK, Kulakow PA (1994) Seed yield and biomass allocation in Sorghum bicolor and F1 and backcross generations of S. bicolor × S. halepense hybrids. *Can J Bot* 72(4):468–474.
163. Pelletier DL, Frongillo EA, Schroeder DG, Habicht JP, Habicht JP (1995) The effects of malnutrition on child mortality in developing countries. *Bull World Health Organ* 73(4):443–8.
164. Chiremba C, Rooney LW, Beta T (2012) Microwave-Assisted Extraction of Bound Phenolic Acids in Bran and Flour Fractions from Sorghum and Maize Cultivars Varying in Hardness. *J Agric Food Chem* 60(18):4735–4742.
165. Dykes L, Hoffmann L, Portillo-Rodriguez O, Rooney WL, Rooney LW (2014) Prediction of total phenols, condensed tannins, and 3-deoxyanthocyanidins in sorghum grain using near-infrared (NIR) spectroscopy. *J Cereal Sci* 60(1):138–142.
166. Taylor JRN, Schüssler L (1986) The protein compositions of the different anatomical parts of sorghum grain. *J Cereal Sci* 4(4):361–369.
167. Wyss CS, Czyzewicz JR, Below FE (1991) Source-Sink Control of Grain

Composition in Maize Strains Divergently Selected for Protein Concentration. *Crop Sci* 31(3):761.

168. Rooney LW (1978) Sorghum and pearl millet lipids.
169. Popkin BM, Adair LS, Ng SW (2012) Global nutrition transition and the pandemic of obesity in developing countries. *Nutr Rev* 70(1):3–21.
170. Dicko MH, Gruppen H, Traoré AS, Voragen AGJ, Van Berkel WJH (2006) Sorghum grain as human food in Africa: relevance of content of starch and amylase activities. *African J Biotechnol* 5(5):384–395.
171. Douglas JH, Sullivan TW, Bond PL, Struwe FJ (1990) Nutrient Composition and Metabolizable Energy Values of Selected Grain Sorghum Varieties and Yellow Corn 1 , 2 by genetic , agronomic , and managerial factors industry . Nutritionists usually rely on tabular feed composition values to formulate least-cost. 908.
172. Anderson JW, et al. (2009) Health benefits of dietary fiber. *Nutr Rev* 67(4):188–205.
173. Tan H-Z, Li Z-G, Tan B (2009) Starch noodles: History, classification, materials, processing, structure, nutrition, quality evaluating and improving. *Food Res Int* 42(5):551–576.
174. Dykes L, Rooney LW (2006) Sorghum and millet phenols and antioxidants. *J Cereal Sci* 44(3):236–251.
175. Awika JM, Rooney LW (2004) Sorghum phytochemicals and their potential impact on human health. *Phytochemistry* 65(9):1199–1221.

176. Awika JM, Rooney LW, Waniska RD (2005) Anthocyanins from black sorghum and their antioxidant properties. *Food Chem* 90(1):293–301.

177. Gerik T, Bean B, Vanderlip R (1993) Sorghum: State of the Art and Future Perspectives. *Sorghum: State of the Art and Future Perspectives*, ed Prasad IC and V. (American Society of Agronomy and Crop Science Society of America, Inc.).

178. Blum A, Golan G, Mayer J, Sinmena B (1997) The effect of dwarfing genes on sorghum grain filling from remobilized stem reserves, under stress. *F Crop Res* 52(1–2):43–54.

179. Craufurd P., Qi A (2001) Photothermal adaptation of sorghum (Sorghum bicolour) in Nigeria. *Agric For Meteorol* 108(3):199–211.

180. Sanon M, et al. (2014) Photoperiod sensitivity of local millet and sorghum varieties in West Africa. *NJAS - Wageningen J Life Sci* 68:29–39.

181. Quinby JR (1974) *Sorghum improvement and the genetics of growth* (Texas Agricultural Experiment Station, College Station, TX).