THE IDENTIFICATION OF GENES AFFECTING MATURITY AND BIOMASS YIELD IN *SORGHUM BICOLOR*

A Dissertation

by

ANNA LYNN CASTO

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 Mullet
Committee Members,	Scott Finlayson
	Elizabeth Pierson
	Hisashi Koiwa
Head of Department,	Dirk Hays

May 2019

Major Subject: Molecular and Environmental Plant Sciences

ABSTRACT

Lignocellulosic biofuels are a promising source of renewable liquid fuel. Sorghum is a drought-resilient C4 grass that is grown for grain, forage, sugar, and biomass. It is a promising lignocellulosic biofuel feedstock because of its ability to accumulate large amounts of biomass with minimal inputs. Isolating and characterizing genes contributing to traits relevant to biomass yield is an important step in optimizing sorghum as a dedicated bioenergy crop. To that end, the three studies presented here use quantitative trait locus (QTL) mapping to explore the genetics of sorghum development as it relates to biomass accumulation. In the first study, six unique QTL are mapped across two environments for flowering time, stem length, and biomass characteristics in the expanded BTx623/IS3620C population with an improved genetic map. Two flowering time loci are identified that were not found in earlier studies of this population. Biomass QTL in this population colocalize with known flowering QTL and the *Dwarf* loci, highlighting the importance of lengthened vegetative growth and stem length to biomass accumulation. In the second study, a single gene encoding a NAC family transcription factor is isolated from a QTL for variation in stem aerenchyma formation. *SbNAC_D* expression is elevated in internodes soon after they reach full length where expression is correlated with aerenchyma formation. Expression of *SbNAC_D* is also correlated with the expression of the programmed cell death gene, *SbXCP1*. Lastly, the *Maturity2* QTL from the population Hegari/80M is fine-mapped to a gene encoding a lysine methyl transferase. Analysis of the circadian expression of sorghum flowering time genes reveal that *Ma*$_2$ contributes to

delayed flowering by enhancing the expression of *SbPPR37* and *SbCO*, which are co-repressors of flowering under long days.

DEDICATION

To my parents

who gave me everything I needed to succeed

To Grander and Big Bill

who supported every step of my education

ACKNOWLEDGEMENTS

I would like to thank my committee chair, Dr. John Mullet, for his guidance and patience throughout my PhD. I would also like to thank my committee members for their help refining and improving my research process. Thanks especially to Dr. Scott Finlayson who allowed me to use his lab's resources during my research.

Many thanks to my many labmates, many of whom are mentors and friends. My days in lab were brightened by your company. Thanks especially to Dr. Daryl Morishige for his advice and patient help over the years. Also, to Dr. Dirk Hays and LeAnn Hague for making me feel welcome in the MEPS program.

Thank you to my family who were a source of unwavering support throughout my PhD. A special thanks to my partner, Jorge, who provided daily support and encouragement as well as much needed perspective. I also thank my many friends in the Brazos Runners Club without whom I would not have been able to maintain my health and happiness during my doctoral studies.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by a dissertation committee consisting of Professor John Mullet of the Biochemistry and Biophysics Department, Professor Elizabeth Pierson of the Horticulture Department, Professor Scott Finlayson of the Soil and Crop Sciences Department, and Professor Hisashi Koiwa of the Horticulture Department.

The analyses in Chapter IV were completed in part by Dr. Ashley Mattison and Dr. Sara Olson. All other work conducted for this dissertation was completed by the student independently.

Funding sources

This work was funded in part by the Perry Atkisson Chair and the Great Lakes Bioenergy Research Council.

TABLE OF CONTENTS

Page

ABSTRACT ii

DEDICATION iv

ACKNOWLEDGEMENTS v

CONTRIBUTORS AND FUNDING SOURCES vi

TABLE OF CONTENTS vii

LIST OF FIGURES x

LIST OF TABLES xii

CHAPTER I INTRODUCTION 1

Future challenges for crop productivity 1
The mandate for biofuels and types of biofuel feedstocks 2
Sorghum bicolor and its potential as a bioenergy crop 6

CHAPTER II IDENTIFICATION OF QUANTITATIVE TRAIT LOCI ASSOCIATED WITH PLANT ARCHITECTURE AND BIOMASS 10

Introduction 10
Results and Discussion 15
Descriptive statistics 15
Quantitative trait loci 19
Flowering time, internode number, and leaf number 21
Stem characteristics 22
Biomass 25
Conclusion 27
Methods 28
Plant materials and growing conditions 28
Genotyping 28
Phenotyping 29
QTL mapping 29
Statistics 30

CHAPTER III IDENTIFICATION AND CHARACTERIZATION OF *SBNAC_D*, A MAJOR GENE INFLUENCING AERENCHYMA FORMATION IN THE STEM 31

Introduction 31
Materials and methods 34
Plant materials and growing conditions 34
Plant phenotyping and visualization of aerenchyma by microscopy and CT scanning 34
Quantitative trait locus (QTL) analysis and fine mapping 35
DNA sequencing and sequence analysis 36
Protein sequence analysis 36
cDNA sequencing and qRT-PCR 37
Statistical analyses 38
Results 38
Sorghum genotypes vary in stem aerenchyma accumulation 38
QTL-assisted identification of the D-gene 41
D-alleles in sorghum germplasm 54
Phylogenetic analysis of SbNAC_D 54
Expression of D-locus genes in RILs with BTx623 or IS3620c haplotypes 59
Developmental timing of stem aerenchyma accumulation and SbNAC_D expression 60
SbXCP1 expression is correlated with aerenchyma formation and regulated by SbNAC_D 65
Spatial analysis of aerenchyma accumulation in sorghum internodes 70
Discussion 71

CHAPTER IV *MATURITY2*, A NOVEL REGULATOR OF FLOWERING TIME IN *SORGHUM BICOLOR*, INCREASES EXPRESSION OF *SBPRR37* AND *SBCO* IN LONG DAYS DELAYING FLOWERING 79

Introduction 79
Methods 82
Plant growing conditions and populations 82
QTL mapping and multiple-QTL analysis 83
Fine mapping of the Ma_2 QTL 84
Circadian gene expression analysis 85
Ma_2 phylogenetic analysis 86
Ma_2 DNA sequencing and whole genome sequence analysis 86
Results 87
Effects of Ma_2 alleles on flowering pathway gene expression 87
Genetic analysis of Ma2 and Ma4 91
Epistatic interactions between Ma_2 and Ma_4 93
Ma2 candidate gene identification 94
Distribution of Ma_2 alleles in the sorghum germplasm 99

Discussion....101
Interactions between Ma_2 and Ma_4....104
Identification of a candidate gene for Ma_2....105

CHAPTER V CONCLUSIONS....108

REFERENCES....111

APPENDIX PRIMER SEQUENCES....129

LIST OF FIGURES

Page

Figure 1. Box and whisker plot of days to anthesis (DTA) data in the high-light (HL) and low-light (LL) environments. 16

Figure 2. Box and whisker plot of leaf number (LN) data in the high-light (HL) and low-light (LL) environments. 17

Figure 3. Box and whisker plot of total fresh weight (TFW) and stem fresh weight (SFW) data in the high-light (HL) and low-light (LL) environments. 18

Figure 4. Box and whisker plot of total dry weight (TDW) and stem dry weight (SDW) data in the high-light (HL) and low-light (LL) environments. 18

Figure 5. Stem aerenchyma rating scale, 1 (no aerenchyma) to 5 (high aerenchyma). ... 39

Figure 6. Variation in the extent of stem aerenchyma (AER) in diverse sorghum genotypes. 40

Figure 7. QTL associated with variation in aerenchyma (AER) formation were mapped in three biparental populations: BTx623*IS3620c (A, n = 380, RIL), BTx623*R07007 (B, n = 215, F5), and BTx623*Standard Broomcorn (C, n = 133, F2). 43

Figure 8. Phenotype-genotype plot of stem aerenchyma ratings in the F2 population BTx623*Standard Broomcorn at the marker with the highest LOD score in the QTL interval. 44

Figure 9. Positional cloning of SbNAC_D. 46

Figure 10. Transcript sequences of the candidate gene Sobic.006G147400 obtained from Sanger sequencing (top) and from Phytozome (bottom). 48

Figure 11. Protein sequence alignment of SbNAC_D homologs in select grass species. 49

Figure 12. Phylogenetic analysis of grass homologs of SbNAC_D. 56

Figure 13. Phylogenetic relationships of SbNAC_D to (A) SbNAC_D paralogs in the sorghum genome and (B) SbNAC_D orthologs in other plant species. 58

Figure 14. Phylogenetic relationships of SbNAC_D homologs in grasses to NAC-family VND genes involved in vascular differentiation and NAC-family genes involved in senescence. 59

Figure 15. SbNAC_D expression in three pairs of SbNAC_D/SbNAC_d1 RILs: (A) RIL226/RIL212, (B) RIL119/RIL474, (C) RIL392/RIL49. 62

Figure 16. Aerenchyma (AER) formation and expression of SbNAC_D in stem sections of R07020. 64

Figure 17. Expression of *SbXCP1* in (A) *SbNAC_d1/SbNAC_D* RILs and (B) R07020 internodes. 67

Figure 18. Expression of genes in the NAC_D QTL interval during the development of internodes of R07020 stems: (A) Sobic.006g147450, (B) Sobic.006g147500, (C) Sobic.006g147600, (D) Sobic.006g147800, (E) Sobic.006g147900, (F) Sobic.006g148000. 69

Figure 19. Microscopic (A) and CT scan (B) images of R07020 stem cross-sections at progressive stages of aerenchyma (AER) formation. 71

Figure 20. Circadian expression of genes regulating flowering in *S. bicolor* in 100M and 80M under long days. 88

Figure 21. Expression of the *S. bicolor FT*-like genes *SbCN8*, *SbCN12*, and *SbCN15* in long days at the expected peak of expression. 89

Figure 22. Circadian expression of *SbTOC1*, *SbLHY*, *SbGhd7*, and *SbEhd1* 90

Figure 23. Quantitative trail locus (QTL) map of flowering time in the Hegari/80M F_2 population. 92

Figure 24. Genotype x phenotype plots for the QTL on SBI02 and SBI10. 93

Figure 25. Interaction plots for the Ma2 QTL and the Ma4 QTL. 94

Figure 26. Fine-mapping of the Ma2 QTL. 95

Figure 27. Alignment of Sobic.002G302700 with its closest homologs in several plant species. 98

Figure 28. Circadian expression of Sobic.002G302700 in 100M and 80M. 99

Figure 29. A model of the flowering time regulatory pathway in *S. bicolor*. 103

LIST OF TABLES

Page

Table 1. Description of QTL for each trait measured in the high-light and low-light environments. 20

Table 2. QTL for DTA and TSL identified in this study and two previous studies of BTx623/IS3620C. 22

Table 3. Sub-population numbers for each internode. 25

Table 4. QTL for internode lengths in the subpopulation fixed for Dw3. 25

Table 5. 2 LOD QTL intervals from three biparental mapping populations. 42

Table 6. Genes in the fine-mapped QTL region 45

Table 7. *SbNAC_D* alleles identified in various genotypes of *Sorghum bicolor.* 52

Table 8. Coding region sequence variants in genes in the NAC_D QTL interval among the parental lines used in QTL mapping. 53

Table 9. Sequence variants of Sobic.002G203700 and their predicted effect on protein function 101

CHAPTER I

INTRODUCTION

Future challenges for crop productivity

Though crop yields have increased dramatically in the past 100 years through the efforts of plant breeders and production technology, crop productivity faces significant challenges in the years to come. The global human population is predicted to increase to nine to ten billion people by the year 2050. Alongside the increase in population, standards of living in many developing countries are improving. Diets higher in meat and dairy products are becoming the norm, so there is added pressure on crop supplies to provide sufficient feed for livestock as well. It is predicted that agricultural crops will have to be 60 – 110% more productive by 2050 to provide enough food and fiber for the global population [1]. Agricultural crops have experienced large gains in productivity over the past century. Maize (*Zea mays*) in particular has seen steady gains in yield since the 1930s due to intensive breeding efforts and improved crop management [2]. The Green Revolution of the 1960s increased yields of crops like wheat and rice by selecting for shorter plants that not only experienced less stalk lodging, but also responded to intensive fertilization by directing most of their photosynthate toward seed production [3,4]. The yield increases of the past are beginning to reach a plateau as many of the easily won yield gains have already been exploited [5]. Much of the available arable land in the world is

already being used for crop production, so further increases in yield cannot come from increasing the land area for growing crops [6].

The problems presented by population growth and crop productivity are inextricably linked to the additional modern challenge of climate change. Atmospheric CO_2 levels have risen from 284 ppm in 1832 to 407 ppm in 2018 due in large part to the increased use of fossil fuels [7–9]. Though at first glance higher CO_2 levels would seem to benefit plant growth and, by extension, crop yields, predicting yield in a changing climate has proven to be a complex problem. A 0.5% increase in average global temperature is correlated with more extreme changes in seasonal temperatures and weather events [10]. Changing temperatures and rising CO_2 levels have also changed ranges and life cycles of certain pests and plant diseases [10]. Overall, warming global temperatures have a net negative effect on crop productivity [11]. Indeed, under continued global warming, an average decline of 8% in crop yields has been predicted across some of the world's most food-insecure regions by 2050 [8]. Crops of the future will have to be more productive on the same amount of land and more resilient to a range of abiotic and biotic stresses.

The mandate for biofuels and types of biofuel feedstocks

A growing and developing human population requires more food and fuel, but continued widespread reliance on fossil fuels serves only to advance climate change. Furthermore, fossil fuels are a limited resource that will eventually run out. Renewable,

carbon-neutral sources of fuel and other forms of energy are needed to maintain a modern standard of living while reducing the rate of increase in atmospheric CO_2 levels. Biofuels that are derived plants and/or microalgae, are an attractive source of renewable energy since they have the potential to be completely or close to carbon neutral. Since plants capture CO_2 for their own growth, CO_2 released into the atmosphere through burning of biofuels will be re-captured by additional biofuel feedstock crops. In one study, sweet sorghums captured an estimated 12-51 Mg CO_2/ha [12]. Schmer et al [13] reported that use of corn- and switchgrass-based bioethanol could reduce greenhouse gas emissions by -29 – -396 g of CO_2 equivalent per MJ of ethanol/y. Recognizing the need to reduce greenhouse gas emissions and support development of renewable fuels, the United States 2005 Energy and Policy Act created the Renewable Fuel Standard (RFS) which required an increase in renewable fuel to 7.5 Bgal/yr by 2012 [14]. An update in 2007 increased this goal to 36 Bgal/y by 2022 and capped the volume of conventional biofuels derived from grain 15 Bgal/yr [14]. Currently, biomass-derived energy accounts for about 43% of U.S. renewable energy consumption with liquid biofuels accounting for about half of that, but this is still far short of the goals set out by the RFS [15].

Several types of biofuel feedstocks have emerged over the last few decades of biofuel research. Conventional biofuels use seed starch (mostly from corn) to ferment into ethanol. Almost all gasoline in the United States is currently at least 10% corn ethanol, so this type of biofuel production is already well-established in the U.S. However, there are several disadvantages of corn-based ethanol production. Corn is a cornerstone of the food

system as a major component of animal feed as well as multiple commercial products such as corn syrup. Hoekman et al [14] estimate that corn yield would have to increase an additional 40% to meet U.S. energy requirements under the RFS. Directing a large portion of corn yield to biofuel production could potentially increase corn prices and thus increase the cost of food. Further, dedicating conserved and/or non-crop land to growing more corn as a monoculture has potentially detrimental effects on the ecology of the areas where corn is grown [14]. Another option for directly converting plant carbohydrates to ethanol are sweet sorghum and sugarcane. Sweet sorghums and sugarcane both accumulate high concentrations of sugar in their stems [4]. Currently some sweet sorghum and sugarcane crops are being used as biofuel feedstocks. An advantage of high stem sugar plants such as these is that very little additional post-harvest processing is needed to obtain a product that can be fermented to ethanol. However, delaying the time between harvest and fermentation can alter the fermentable sugar quantity and composition in the stem, so a system for continuous harvest would be necessary for large-scale implementation [16].

Second generation biofuels use microalgae to produce carbohydrates and lipids, but for these systems to be productive there would have to be a large investment in the infrastructure needed to house algal biofuel production systems. Additional research is also needed to perfect the biofuel harvest and production from algae [17]. There is also some concern about the health risks to humans and aquatic life [17].

Another alternative biofuel feedstock is lignocellulosic biomass, which uses plant biomass materials (leaves, stems, etc) to convert plant structural and waste material to

fuel. In addition to crop waste such as corn silage and sugarcane/sorghum bagasse, a few plants are being developed as dedicated high-biomass cellulosic bioenergy crops. These include poplar (*Poplar spp.*), switchgrass (*Panicum virgatum*), *Miscanthus x giganteus*, and bioenergy sorghums. Some of the species being investigated for lignocellulosic feedstocks such as switchgrass and *Miscanthus* are perennials, which eliminate the need to remove plants every season and help with soil conservation [18,19]. Despite the advantages of low-input cultivation and year-round ground cover, farmer buy-in has been a problem for developing perennial crops as a viable biofuel feedstock. Perennial crops can take several years to be productive and requires that farmers dedicate land and time to growing a plant during their establishment phase [18]. This proves a significant risk factor for growers. A current disadvantage of all lignocellulosic feedstocks for biofuel is that the plant material requires chemical pretreatment before fermentation to access the cell wall polysaccharides [20]. Some plants form a secondary cell wall comprised of cellulose microfibrils in a complex polysaccharide matrix, but access to these polysaccharides is inhibited by lignin [21]. Pretreatment loosens the secondary cell wall so that enzymes can access and hydrolyze cell wall polysaccharides into fermentable sugars; however, pretreatment can be expensive and raises concerns about the ecological impact of waste material [20]. Some have tried circumventing the need for pretreatment by reducing the lignin content of cell walls in the plant through transgenics or mutagenesis of lignin biosynthetic genes [22–24]. Though there are a few hurdles for lignocellulosic biofuels to

overcome, the overall energy balance of these biofuels (>10:1) is more favorable than conventional biofuels [25].

Sorghum bicolor and its potential as a bioenergy crop

Sorghum bicolor is a multiuse crop that is grown globally for grain, sugar, animal feed, and biomass. According to the Food and Agriculture Organization, it is the fifth most important crop worldwide with 58 million tons of grain produced globally from 42.6 million ha [26]. The U.S. is the largest global producer of sorghum, where it is primarily grown for livestock feed and ethanol production. Sorghum's center of origin is thought to be in central Africa where it was domesticated 5000 - 8000 years ago and then spread throughout Africa and Asia before introduction to the US in the late 1800s [27]. In the U.S., sorghum has benefited from a long history of breeding for grain and forage yield, and recent breeding efforts have turned to developing high-biomass energy hybrids [25].

Sorghum has several advantages as a biofuel feedstock. Unlike many perennial biomass crops being investigated as biofuel feedstocks, sorghum has a relatively small genome (800 Mbp) and is diploid [25]. Many perennial crops have only recently been domesticated, while sorghum was domesticated thousands of years ago and has benefited from extensive breeding efforts in the last 100 years, including the development of cytoplasmically male sterile lines for hybrid seed production [25]. The sorghum research community has also developed many genetic resources, such as a fully sequenced genome and over 50 whole genome sequences of various sorghum genotypes [28]. Several

important sorghum genes have been cloned and characterized, and efforts at improving sorghum transformation are well underway [29–36]. Sorghum's drought-resilience and high nitrogen and radiation use efficiency make it possible to adapt sorghum for growth on marginal lands, thus reducing the impact of bioenergy crops may have on land area used to grow food crops. Multiple sorghum products can feed into a biofuel production system. Sorghum grain is high in starch and can be used for conventional starch-based biofuel [37]. Sweet sorghum stem exudate is up to 75% sucrose, so it can be extracted and directly fermented to ethanol as discussed above. High biomass sorghums can accumulate up 40-50 dry Mg/ha under ideal growing conditions, which was similar to perennial *Miscanthus* biomass yields [38]. As an annual crop, sorghum may be more attractive to growers who are already accustomed to annual crop production [25]. Additionally, sorghum has alternative markets as forage, sugar, or a specially grain that may reduce the risk for farmers that grow sorghum [25].

Optimizing sorghum for use as a biofuel feedstock will require the combined efforts of plant breeding, genomics, and biotechnology. The isolation and eventual manipulation of specific genes relevant to bioenergy sorghum varieties may help advance efforts in creating a biofuel feedstock to meet future energy demand. The sorghum stem accumulates up to 80% of the plant's biomass [39], so an understanding of the genetics of sorghum stem growth is critical to improving biomass yield of sorghum. Historically, four *Dwarf* (*Dw*) genes were described that controlled segregation of height in sorghum populations [40]. Three of the *Dw* loci (*Dw1*, *Dw2*, and *Dw3*) have been map-based cloned

[32–34]. Two of these genes are involved in auxin transport (*Dw3*) and response (*Dw2*), showing that auxin plays a major role in the regulation of sorghum stem height [32,34]. Numerous genetic studies of sorghum populations have identified other quantitative trait loci (QTL) associated with variation in height of sorghum plants. In addition to stem height, the composition of the stem is an important component of designing a biofuel feedstock. Numerous studies have explored the composition of sorghum stems in both sweet and bioenergy sorghums [41,42].

An extended vegetative phase, which is achieved through delayed flowering, allows bioenergy sorghums accumulate large amounts biomass. Sorghum is originally a tropical grass that evolved to delay flowering in photoperiods (or day-lengths) longer than 12 h. Years of breeding for adaptation to temperate climates typical of higher latitudes yielded sorghum varieties that are much less photoperiod sensitive and have shorter vegetative phases in long days [43]. Four flowering time or *Maturity* (*Ma*) loci were described based on segregation patterns for flowering time between the 1940s-60s [44]. Two additional *Ma* loci (Ma_5 and Ma_6) were described later [45]. Many of these genes associated with the maturity loci have been cloned and characterized, providing a greater understanding of how flowering time is modulated on the molecular level [29,30,46]. Understanding the molecular genetics of flowering time control in sorghum will allow sorghum crops to be adapted to a variety of environments and widen the possible range for sorghum production.

This dissertation aims to contribute to a greater understanding of the genetics influencing sorghum characteristics important for biomass accumulation with the further goal of advancing sorghum as a model C4 grass. To this end the following three studies are described: a survey of stem growth and biomass accumulation in a biparental population in multiple environments, the identification of a major gene contributing to aerenchyma formation in the stem, and the identification of the gene corresponding to *Maturity2*, a gene that delays flowering under long days.

CHAPTER II

IDENTIFICATION OF QUANTITATIVE TRAIT LOCI ASSOCIATED WITH PLANT ARCHITECTURE AND BIOMASS

Introduction

Plant architecture has direct consequences for crop yield. For example, grain yield is linked to panicle architecture traits such as the length and number of branches [47,48]. It is also linked to plant height as taller plants tend to lodge or break, making harvest of the grain impossible. Shorter plants are also able to direct more photosynthate to seed filling rather than toward maintenance of vegetative organs [37]. In terms of biomass yield, taller and thicker stems accumulate greater biomass, but plant height must still be balanced with the possibility of lodging. Another way plant architecture affects productivity of a plant is through leaf angle, which can change light interception and thus the radiation use efficiency of the plant. The genetics underlying plant architecture traits are points of potential manipulation to increase yield of the desired plant product.

Sorghum bicolor is a C4 grass that is grown globally for a wide range of plant products, such as forage, sugar, biomass, and grain. The *S. bicolor* germplasm shows a great deal of plasticity in plant architecture. For example, Broomcorns have large open panicles, while grain sorghums have smaller, densely-packed panicles. Bioenergy sorghums can grow up to 4 – 5 m tall, while grain sorghums barely reach 1 m [49]. Sorghum also shows variation for leaf angle, stem diameter, grain color, and tiller number

[48,50–52]. Some varieties will flower in as little as 60 days, while others will grow vegetatively for 180 days [40]. The diversity of the sorghum germplasm can be exploited to optimize varieties for various sorghum products and perhaps even create highly efficient multi-use varieties. A greater understanding of the genetic basis of these traits can improve the usage of the diversity of the sorghum germplasm.

Several genetic methods have been developed to identify regions of the genome associated with variation for a particular trait. One such method is quantitative trait locus (QTL) mapping. QTL mapping has three essential requirements: 1) a mapping population that segregates for one or more traits of interest, 2) molecular markers along the genome and 3) measurements of quantitative traits [53]. Mapping populations can be structured in many ways, but the most common are experimental crosses. Biparental mapping population are generated by crossing two parents that differ for a particular trait to generate the heterozygous F_1 plants. The F_1 plant is self-pollinated to generate an F_2 population that segregates for the trait of interest due to random assortment of chromosomes and recombination events during meiosis. It is possible to map QTL in F_2 populations, but the disadvantages of this approach are that large heterozygous regions still remain in the F_2 generation and the analysis cannot be replicated. When possible a population of F_2 plants is self-pollinated for several more generations to create a recombinant inbred line (RIL) population. RILs are fixed at most genetic loci with some residual heterozygosity remaining in each RIL. RIL populations can take years to generate, but RIL population

allows for replicated analysis and repeated use over time since the genetics a RIL are fairly stable [53].

In experimental mapping populations, molecular markers are genetic differences between the parental genomes that can be used to genotype the individuals of the population at a particular locus. Once generated, these markers are arranged in a genetic map based on linkage between markers (centiMorgans) or on physical distance if a reference genome is available. Many different methods have been developed to generate markers [54–57]. The earliest were restriction fragment length polymorphisms (RFLPs) in which genomic DNA is digested and the fragment length difference between the parents are compared [54]. If a mutation has occurred in a restriction site in one of the parents, then the restriction fragment lengths will differ between the parents (or INDELS within the fragment). Genomic DNA from each of the individuals in a biparental population can be digested and compared with the restriction sites of the parents to genotype the population. More recently, the affordability of next-generation sequencing has made it possible to produce high-density genetic maps using methods such as Digital Genotyping, which uses single nucleotide polymorphisms (SNPs) between the genomic DNA sequences of the parents as markers. This method is described in detail in Morishige et al [58].

The third requirement of QTL mapping is the measurement of a quantitative phenotype of interest in all the individuals in a population. A quantitative trait shows a continuous distribution of phenotypes across a population caused by contributions of

numerous loci. For example, plant height is a quantitative trait because height within germplasm does not occur in discrete intervals, but rather along a continuous curve. This continuous distribution of the phenotype is due to the presence of many genes that affect a single phenotype. However, in some crosses that contain one or a few main effect alleles, differences in height are inherited in a qualitative manner.

Statistical analysis of marker information and phenotype measurements can be used to generate a QTL map. The correlation between the genotypes at particular markers and the phenotype measurements indicate that genetic variation in that region is influencing the phenotype, so a QTL is called at that locus. Several statistical methods have been developed for QTL mapping [53]. The simplest of these methods is single marker regression, which tests the correlation of the genotype of an individual at single markers with variation of a phenotype of interest by analysis of variance (ANOVA). Interval mapping (IM) and composite interval mapping (CIM) overcome several disadvantages of single marker regression by including the use of a genetic map which adds marker location information to the analysis. In IM, the interval between two markers as well as individual markers are tested for the presence of a QTL (i.e. correlation between the genotype and phenotype).

Many QTL studies have been performed on sorghum biparental populations and association panels [32,33,48,50,59–66]. Early studies used a combination of RFLP and SSR markers to explore a wide range of agronomic traits, such as stem height, tiller height,

panicle width and weight, and maturity [48,66]. More recent studies have used genotyping by sequencing (GBS) markers and larger populations [32,33,50].

A RIL population derived from BTx623/IS3620c has been used to investigate the genetic basis of many traits such as plant height, panicle architecture, seed number, tiller number and height, leaf morphology, and disease resistance [48,58,66–68]. BTx623 is an elite grain inbred line, and IS3620c is a converted inbred line. The two are highly divergent as demonstrated by marker analysis [69,70]. The original $F_{2:6-8}$ RIL population consisted of 137 lines that were developed between 1991 and 1994 [67]. An additional 302 $F_{2:7}$ RILs were added to the populations between 2005 and 2008 by propagating additional F_2 individuals from the original cross [67]. Several genetic maps have been generated for BTx623/IS3620c [58,69,70], and many improvements have been made in the genetic map for this population. For example, the method for map construction as described by Truong et al. [71] accounts for excess heterozygosity, creating a more accurate representation of recombination events. Additionally, version 3 of the sorghum genome has been released with improved genome sequence order, 29.6 Mbp of additional genome sequence, and a 24% increase in annotated genes [28].

The current study explores the genetic basis of several agronomic and plant architecture traits that are relevant to grain and biomass yield. It also compares the flowering time/maturity and stem height QTL identified in the current study with QTL identified in previous studies of BTx623/IS3620c, which used genetic maps based on RLFP and/or SRR markers.

Results and Discussion

Descriptive statistics

The recombinant inbred line (RIL) population derived from BTx623/IS3620c is comprised of 398 individuals. BTx623 is an elite grain sorghum that has short stature and early flowering with a closed panicle. IS3620C is a converted line, a landrace sorghum that had a recessive *Maturity1* (*ma*$_1$) introgressed into IS3620 to confer less photoperiod sensitivity [72]. It has an open panicle and flowers slightly later than BTx623. The population was grown in two environments in the summer of 2014: a high-light greenhouse (HLG) (> 800 μMol/m^2/s PAR) and a lower-light greenhouse (LLG) (~400 μMol/m^2/s PAR).

Mean days to anthesis (DTA) were similar between both environments ($p = 0.8887$) (Figure 1). DTA is controlled by several environmental inputs such as photoperiod, temperature, and light intensity [73–77]; however, a major controlling factor for DTA is day length or photoperiod. Since day lengths were the same between the two environments, despite the light levels being different, the similar flowering times could be a reflection of the relatively stable day-length and temperature environment of the greenhouses. Flowering time has been recorded previously for this population by Hart et al and Brown et al [48,66]. Hart et al and Brown et al used the original BTx623/IS3620C RIL population that had only 137 individuals, so the means and ranges of the phenotypes recorded represent a sub-population of the population used in the current study. The latest flowering RILs in this study flowered earlier than the latest flowering RILs in Brown et al

[48] but were the same as the Lubbock, TX population of Hart et al [66]. The earliest flowering RILs in this study flowered significantly earlier than the RILs in Brown et al [48]. Light intensity and temperature fluctuations are just two of the many environmental variables present in the field that are less extreme under greenhouse conditions. It is likely that the spread of flowering times seen in Hart et al and Brown et al was influenced by the field environment.

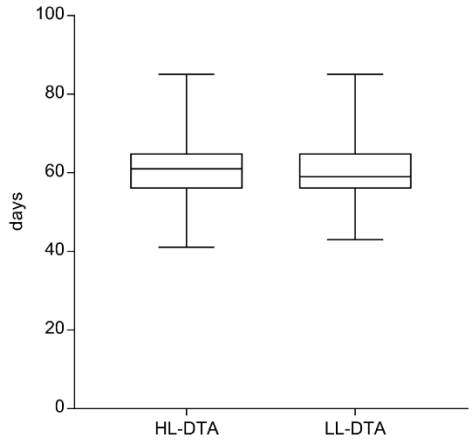

Figure 1. Box and whisker plot of days to anthesis (DTA) data in the high-light (HL) and low-light (LL) environments.

Leaf number (LN) is related to DTA since sorghum has a determinate growth pattern. After the initiation of flowering, no additional phytomers are formed. The HLG had a higher mean LN than the LLG ($p = 0.0005$). Though the flowering dates of this population were very similar in both environments, the slightly higher mean LN in the HLG indicates that the HLG plants' growth was slightly accelerated by greater availability of light or by small difference in temperature.

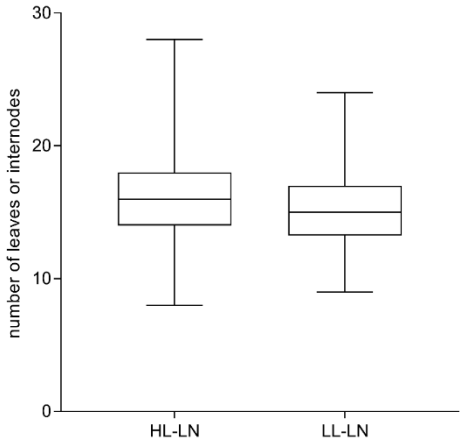

Figure 2. Box and whisker plot of leaf number (LN) data in the high-light (HL) and low-light (LL) environments.

Mean total stem length (TSL) was higher in the HLG than the LLG ($p = 0.0005$). Maximum TSL was 194.5 cm in HLG compared to 174 cm in LLG. Typically, lower light conditions trigger the shade avoidance response which elongates the stem; however, there is a trade off with growing time, as lower PAR reduces the energy and photosynthate availability required for growth. It is conceivable that within the time to flowering the plants in the LLG environment did not reach the same length as the plants in the HLG environment because of the lower availability of light.

The HLG population had a higher mean total fresh weight (TFW), total dry weight (TDW), stem fresh weight (SFW), and stem dry weight (SDW) ($p < 0.0001$ for TFW, TDW, SFW, and SDW). This is likely due again to the higher availability of light allowing for more photosynthesis to take place and thus the accumulation of more biomass in this

environment. While the difference in light levels between these two environments was not enough to affect time to flowering, it was sufficient to affect total biomass accumulation.

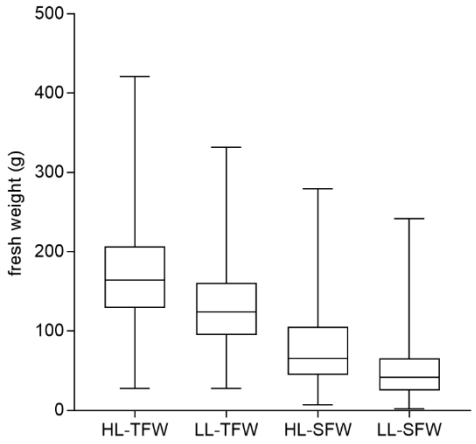

Figure 3. Box and whisker plot of total fresh weight (TFW) and stem fresh weight (SFW) data in the high-light (HL) and low-light (LL) environments.

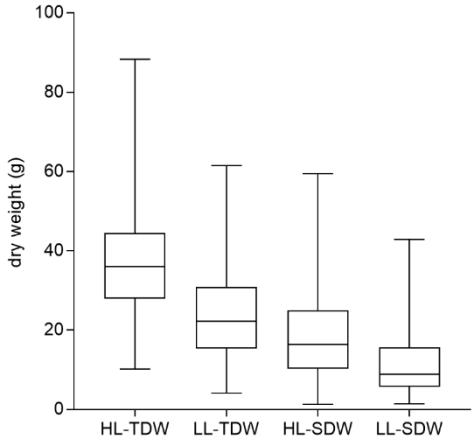

Figure 4. Box and whisker plot of total dry weight (TDW) and stem dry weight (SDW) data in the high-light (HL) and low-light (LL) environments.

Quantitative trait loci

QTL for were mapped for variation in several physiological characteristics in the whole population. The population was grown in two environments: high light greenhouse (HLG) and low light greenhouse (LLG). Six unique QTL were identified in this study across a variety of physiological characteristics. The largest effect QTL explained 41.7% of the variation for a given trait, and the smallest explained 6.1% of variation.

Table 1. Description of QTL for each trait measured in the high-light and low-light environments.

	High-light greenhouse			Low-light greenhouse		
Trait	Chromosome (peak location)*	LOD	%Var†	Chromosome (peak location)*	LOD	%Var †
DTA	1 (45.97)	10.8	14.1	1 (50.11)	6.95	9.6
	3 (61.49)	5.71	7.7	3 (62.76)	7.56	10.4
	8 (52.31)	11.8	15.3	-	-	-
IN	1 (43.99)	10.3	11.5	1 (45.75)	12.3	18.0
	3 (61.77)	5.30	6.1	-	-	-
	8 (52.04)	14.6	16.0	8 (52.04)	8.67	13.1
LN	1 (43.88)	9.74	13.0	1 (49.64)	11.15	17.0
	3 (61.77)	5.48	7.6	-	-	-
	8 (51.86)	14.1	18.3	8 (52.04)	10.61	16.2
TSL	6 (42.50)	9.25	10.3	6 (42.77)	17.32	18.4
	7 (59.78)	45.4	41.7	7 (59.83)	42.70	39.6
TFW	-	-	-	1 (56.49)	9.17	12.2
	4 (1.190)	5.88	7.72	-	-	-
	6 (50.94)	12.1	15.2	6 (50.42)	6.78	9.16
	7 (59.83)	7.53	9.78	-	-	-
	8 (59.34)	7.73	10.0	8 (59.54)	5.85	7.96
SFW	-	-	-	1 (56.49)	10.76	14.1
	-	-	-	6 (47.62)	5.77	7.8
	7 (59.83)	12.6	13.8	7 (59.83)	17.2	21.6
TDW	7 (59.86)	7.56	8.5	1 (56.73)	7.68	24.5
SDW	1 (25.52)	5.38	6.1	1 (56.51)	15.83	25.7
	-	-	-	6 (44.05)	5.05	9.1
	7 (59.78)	15.8	17.0	7 (59.83)	8.26	14.4
	8 (59.34)	8.10	9.1	8 (59.34)	6.11	10.8

*peak location is in Mbp

LOD, Logarithm of odds

†%Var is percent of variance explained by the QTL. %Var is calculated by Eq 1, where n = the number of individuals measured for the trait.

DTA, days to anthesis; IN, internode number; LN, leaf number; TSL, total stem length; TFW, total fresh weight; SFW, stem fresh weight; TDW, total dry weight; SDW, stem dry weight; LDW, leaf dry weight

Flowering time, internode number, and leaf number

QTL for flowering time were mapped to chromosome 1 (ch1), ch3, and ch8 in the HLG environment and to ch1 and ch3 in the LLG. In the HLG the three QTL accounted for a total of 37.1% of the variation in flowering time. The two QTL identified in the LLG explained 20% of the variation in flowering time. Brown et al [48] and Hart et al [66] also mapped flowering time in this population (Table 2). These studies identified QTL on ch9 and ch10 across multiple field environments; however, neither Brown et al [48] nor Hart et al [66] identified the flowering time QTL on ch1 and ch 3 that were identified in this study (Table 2). QTL instability across environments is a well-known phenomenon that is attributed to genotype x environment effects. It is possible that the QTL on ch9 and ch10 require the higher light intensity found in field environments. The expanded range of flowering times recorded by Brown et al [48] also suggests that the field environment creates additional variation that is not present in the greenhouse.

Some of the QTL identified in this study for flowering time co-localize with QTL for flowering time identified in other populations. Murphy et al [31] identified a QTL on ch 1 in BTx623/R07007 that colocalizes with our QTL on ch1. They also identified an additional two QTL on ch6 not found in this study. *Maturity1* is known to be on ch6 [30]; however, both BTx623 and IS3620C are recessive for *Maturity1*.

Table 2. QTL for DTA and TSL identified in this study and two previous studies of BTx623/IS3620C.

	Current study		Hart et al 2001		Brown et al 2006
Trait	HLG	LLG	CS	LB	
DTA/maturity	1*	1	-	-	-
	3	3	-	-	-
	8	-	-	-	8
	-	-	-	9	9
	-	-	10	10	-
TSL/plant height	-	-	1	1	-
	-	-	3	-	3
	6	6	-	-	6
	7	7	7	7	7
	-	-	-	-	8
	-	-	10	10	-

*Numbers represent the chromosome number that the QTL was identified on.
CS = College Station, TX; LB = Lubbock, TX

In each environment, the QTL identified for leaf number and internode number co-localized with QTL for days to flowering, due to sorghum's determinate growth pattern (Table 1). In LLG, the QTL on ch1 was identified for days to flowering, internode number, and leaf number. However, the QTL on ch3 was not identified for leaf or internode number. The QTL on ch8 was also identified for leaf number and internode number, but not for days to flowering in the LLG.

Stem characteristics

Total stem length (TSL) is an agronomically important trait as shorter stems prevent lodging and taller stems allows for more biomass accumulation. TSL was measured from the base of each plant to the base of the panicle for each individual in both

environments (note this would include the peduncle). Two QTL were identified in both environments (Table 1). The QTL on ch7 co-localized with the known location of *Dwarf3* (*Dw3*), a gene encoding an ABCB1 transporter of auxin [34]. This gene is a major determinant of stem height in sorghum [78]. The *Dw3* QTL accounts for nearly 40% of variation in TSL in this population, which is consistent with previous analysis of stem length in BTx623/IS3620c [32]. The QTL on ch6 co-localized with the location of *Dwarf2* (*Dw2*) [32]. *Dw2* is also an important regulator of stem length was recently identified as an AGC protein kinase [32].

In order to dissect the genetics of stem length further, QTL were mapped for the length of each internode. The internodes were originally numbered from the bottom to the top of the plant, but each plant flowered at a different time and thus had a different number of internodes, which caused a misalignment of internodes at different developmental stages. For example, in one plant, internode 10 might be the internode subtending the peduncle, while in another plant, internode 10 might be several internodes below the peduncle. In order to compare internodes at the same developmental stage, the internodes were aligned to the top of the plant and re-numbered from the top down for this analysis. In HLG, *Dw3* accounted for 26 – 33% of variation in internode length for each internode measured. *Dw2* accounted for 7.5 – 10% of the variation in internode length. In the LLG, the *Dw2* QTL accounted for 8 – 25% of internode length variation, while *Dw3* accounted for 20 – 34% of the variation. QTL analysis of the individual internode lengths revealed two additional QTL that contributed to variation in specific internodes. In the HLG,

besides the *Dw2* and *Dw3* QTL, a QTL on ch1 was identified for the length of internode (int) 1 (first internode below the peduncle). In the LLG, an additional QTL was found on ch3 for the length of int6 and int7.

Large effect QTL such as the *Dw3* QTL can mask the effect of smaller QTL, so a subpopulation was analyzed that was fixed for the *Dw3* QTL and used to map variation in internode length. Reducing population size also reduces the statistical power of QTL mapping, so it is important to make sure that the subpopulation does not become so small that it is no longer useful. The number of individuals in the subpopulations of the top 6 internodes were large enough to map QTL (Table 3). The lower internodes differ in subpopulation number because individuals in the population had a different number of total internodes. In this subpopulation, a QTL on ch5 was identified for the length of the third internode from the top; however, it is not present in any other internode. The *Dw2* QTL on ch6 is responsible for most of the variation in internode length in this sub-population.

Table 3. Sub-population numbers for each internode.

High-light		Low-light	
Internode	Sub-population n	Internode	Sub-population n
TD1	163	TD1	139
TD2	163	TD2	139
TD3	163	TD3	139
TD4	163	TD4	139
TD5	162	TD5	138
TD6	155	TD6	130
TD7	136	TD7	121
TD8	104	TD8	91
TD9	84	TD9	62
TD10	68	TD10	40

TD, top-down. TD1 corresponds to the first internode below the peduncle//

Table 4. QTL for internode lengths in the subpopulation fixed for Dw3.

	High-light greenhouse			Low-light greenhouse		
Internode	Chromosome (peak location)*	LOD	%Var†	Chromosome (peak location)*	LOD	%Var†
1	6 (42.74)	5.30	13.9	6 (42.74)	9.31	2.65
2	6 (44.90)	6.92	17.8	6 (42.72)	6.62	1.97
3	5 (59.82)	5.78	15.1	5 (58.33)	5.56	1.68
	6 (45.18)	6.00	15.6	6 (42.72)	6.34	1.89
4	-	-		6 (42.72)	12.49	3.39
5	-	-		-	-	-
6	6 (44.90)	5.21	14.3	6 (42.64)	11.00	3.23

*peak location is in Mbp

†%Var is percent of variance explained by the QTL. %Var is calculated by Eq 1, where n = the number of measured individuals the population.

Internodes are numbered from the internode below the peduncle down.

Biomass

Though the parents of this population are not sorghums bred specifically for bioenergy, it can be useful to evaluate the biomass accumulation of other types of sorghum as sustainable bioenergy production may include multi-use feedstocks. The waste products

of sorghum grown for grain or sugar could be redirected to biofuel production [37], so evaluating the biomass accumulation of sorghums grown primarily for grain could reveal a source of additional biomass for biofuel production. The fresh and dry weights of the entire plant and the stem were recorded and used to map QTL for variation in biomass.

Variation in the total fresh weight (TFW) mapped to several QTL that co-localized with *Dw2*, *Dw3*, and several flowering time QTL in both environments (Table 1). The *Dw2* QTL accounted for the most variation in TFW in both environments. Surprisingly, the *Dw3* QTL contributed to variation in TFW only in the HLG environment. It also accounted for only 9.6% of the variation, compared to 14% for the *Dw2* QTL. In contrast, the *Dw3* QTL was the only QTL identified for stem fresh weight (SFW) in the HLG, and it accounted for 13.8% of variation in SFW. In the LLG environment, an additional two QTL were identified for SFW, and the *Dw3* QTL accounted for 7.8% more variation in SFW in that environment.

Total dry weight (TDW) mapped to only one QTL in each environment, but they were not the same QTL. In the HLG, QTL for TDW co-localized with *Dw3* and accounted for only 8.5% of variation. In the LLG, variation for TDW mapped to a QTL on ch1, and accounted for a large amount of variation (25.7%). The light environment of the two greenhouses caused different genes to contribute to dry biomass accumulation. The QTL on ch1 also appears to affect dry biomass accumulation more strongly in a low light environment. It is possible that this QTL confers an improved ability to capture or utilize light energy in low light environments.

Stem dry weight (SDW) co-localized with flowering time QTL on ch1 and ch8 as well as with the *Dw2* and *Dw3* QTL. As with TDW, the QTL on ch1 accounted for the most variation in SDW (25.7%) in the LLG. Low light conditions appear to create greater variation in the population in terms of dry weight accumulation. Fewer QTL were mapped in the HLG (3 in HLG vs. 4 in LLG) and the QTL accounted for much less of the total variation in SDW (32.2% in HLG vs. 60% in LLG).

Conclusion

The current study surveyed several plant architecture and biomass traits in the large biparental population BTX623/IS3620C and identified six unique QTL across those traits. The expanded population as well as the improved genetic map allowed higher resolution mapping of these traits. Two QTL for flowering time were identified in this study that were not identified in previous studies of this population. Earlier studies identified several QTL that were not identified in this study, likely due to a difference in the environment in which the population was grown. Variation in flowering time and stem length (*Dw2*, *Dw3*) have a large influence on biomass accumulation in this sorghum population. All the QTL identified for biomass measurements co-localized with flowering time and *Dw* QTL. Manipulation of these key components of sorghum development could be useful for optimizing sorghum biomass yield.

Methods

Plant materials and growing conditions

Seed for the population BTx623/IS3620c was obtained from the Sorghum Breeding Program at Texas A&M University in College Station, TX. Plants were grown in greenhouses in the summer of 2014 in College Station, TX. Three seeds of each line were planted in 10-L pots filled with MetroMix 900 (SunGro Horticulture). Three plants per pot were allowed to grow until 30 days after planting (DAP) then pots were thinned to one plant per pot. One replicate of the whole population was grown in two environments. In the high-light greenhouse (HLG), > 800 µMol/m2/s of photosynthetically active radiation (PAR) was available for the growing plants. In the low-light greenhouse (LLG), ~400 µMol/m2/s of PAR was available. Plants were fertilized using Peters Professional 20-20-20 every two weeks after 30 DAP. Plants were watered as needed to avoid over-watering.

Genotyping

This population was genotyped previously by Digital Genotyping (DG) using the NgoMIV restriction enzyme [58,71]. The template fragments were sequenced on Illumina HiSeq 2500 with 72 samples per lane. A detailed description of the marker curation process is available in [71]. Once the v3 sorghum genome was released, the sequences were realigned to the new genome version, and the v3-aligned map was used for QTL mapping in this study.

Phenotyping

Leaves of each plant were numbered as they grew, and internode number corresponds to the internode above the leaf sheath attachment point. All plants were allowed to grow until anthesis. Upon first emergence of the panicle, the sorghum panicle was bagged and marked with the date. The number of days from planting until bagging was recorded as days to anthesis (DTA). After anthesis, each plant was cut at the base near the soil. Total fresh weight (TFW) was recorded by placing the intact plant on a balance. Leaves and leaf sheaths were removed from the stem and retained to later obtain dry weights. Internode number (IN) and leaf number (LN) were recorded. Total stem length (TSL) was measured from the base to the node just below the panicle. Each internode was measured from the midpoint of the inferior node to the midpoint of the superior node. Stem fresh weight (SFW) was recorded then stem tissue was dried for dry weights. The panicles were also dried and dry weights were recorded. Dry weights of each component part were summed to obtain total dry weights (TDW).

QTL mapping

All composite interval QTL mapping was performed in R/qtl [71,79] with the generation interval set to 7. Thresholds were calculated with 1000 permutations and $\alpha = 0.05$. Percent variance explained was calculated according to Equation 1 as described in [80],

$$\%Var = 1 - 10^{-(\frac{2}{n} \times LOD)} \quad (1)$$

where n = the population size and LOD = LOD score of the QTL.

In order to understand the contribution of additional QTL to internode length, individuals with the BTx623 haplotype (*dw3*) around the known location of *Dwarf3* (*Dw3*) (58,610,896 – 58,618,660 bp) were removed, and QTL were mapped in a reduced population that no longer segregated for *Dw3*.

Statistics

All descriptive statistics and unpaired t tests of the traits were performed in GraphPad Prism 7.03 for Windows (GraphPad Software La Jolla California USA, www.graphpad.com).

CHAPTER III

IDENTIFICATION AND CHARACTERIZATION OF *SBNAC_D*, A MAJOR GENE INFLUENCING AERENCHYMA FORMATION IN THE STEM *

Introduction

Plant stems are essential for the transport of water and nutrients to and from roots and leaves. Stems also serve as a temporary storage depot for excess photosynthate and nutrients not immediately needed for basic metabolism, growth, or grain filling [4]. In grasses, internode numbers and lengths impact canopy architecture, radiation use efficiency, nitrogen and water use [76,81]. High biomass sorghums develop 4-5m length stems that can account for up to 84% of harvested biomass [39]. Sorghum stems are comprised of phytomers produced approximately every 4 days by the shoot apical meristem [82]. Stem internode elongation is repressed during the juvenile phase but increases during the vegetative phase [82] or following floral initiation [83]. Several genes that regulate sorghum stem growth have been identified including an ABCB1 auxin transporter encoded by *Dw3* [34], a plasma membrane protein encoded by *Dw1* [33,84] that is involved in brassinosteroid signaling and cell proliferation [85], and an AGC kinase encoded by *Dw2* [32].

Sorghum stems have the capacity to accumulate high concentrations (~0.5M) and large amounts of sucrose (up to 40% by weight) similar to sugarcane [42,86–89]. Stem volume, juiciness, and sugar concentration have been under selection in sweet sorghum in order to maximize sugar yield [41,60,90]. Traits and QTL that modulate the accumulation of stem sugars have been identified [60,66,90–97] and the biochemical and molecular basis of stem sugar accumulation has been investigated [42,86,87,98,99]. The stems of sorghum genotypes that accumulate high levels of sucrose generally lack or accumulate low levels of aerenchyma and maintain functional pith parenchyma with large vacuoles where sucrose is sequestered [41,90,100]. Other sorghum genotypes have stems that convert pith parenchyma into aerenchyma to varying degrees [41,100,101].

Genetic variation in the extent of aerenchyma formation in stems and leaf midribs was noted ~100 years ago [100]. Genetic analysis of variation in stem juiciness (vs. pithiness) and the color of leaf midribs (green vs. white) revealed that in some populations the genetics of aerenchyma formation was simple and that a single dominant gene largely accounted for the appearance of 'pithy' stems and white midribs caused by accumulation of aerenchyma [100,102]. Swanson and Parker assigned the symbols *D* and *d* to this main effect genetic factor for stem and leaf mid-rib "dryness" [103]. Genotypes with pithy stems and white midribs encode an active *D*-allele that causes the accumulation of aerenchyma in stems and leaf midribs [100,101]. A major QTL corresponding to the *D*-gene was mapped to LG-06 by Xu et al. [104], a location subsequently confirmed using other

populations [66,96]and by GWAS [90]. A stem moisture QTL was also mapped to the same location on LG-06 [91].

Flooding and hypoxia induce formation of aerenchyma in sorghum roots [105]. Hypoxia increases the synthesis and accumulation of ethylene in roots and ethylene induces programmed cell death mediated root aerenchyma formation [106]. In species such as rice, some root aerenchyma form in the absence of flooding and additional aerenchyma form in response to flooding or ethylene treatment [107]. In maize, accumulation of aerenchyma is induced by hypoxia [106,108]. In contrast the wetland species *Juncus effusus* form root aerenchyma in the absence of ethylene and flooding [109]. Programmed cell death is also induced in response to pathogens through the hypersensitive response [110]. Moreover, programmed cell death occurs in many tissues such as root caps [111], xylem [112], the endosperm [113,114] and during emergence of adventitious roots [115] and organ senescence [116]. The molecular events and genes involved in developmental and abiotic stress induced programmed cell death have been identified in a number of plants [117–120].

The yield and composition of stem biomass is of critical importance for high biomass sorghum since stems account for ~80% of the harvested biomass [39]. Variation in stem aerenchyma formation can impact stem biomass yield and composition by altering the capacity of stems to accumulate sucrose. Therefore, in this study we characterized when and where stem aerenchyma form during stem and internode development and cloned the sorghum *D*-gene to better understand the genetic network that modulates this

important stem trait. The analysis showed that the *D*-gene encodes a NAC transcription factor expressed in stems that regulates the expression of a gene involved in developmental programmed cell death and the formation of aerenchyma.

Materials and methods

Plant materials and growing conditions

All plants were grown in a greenhouse under 14 h long days in 14.5 L pots with MetroMix 900 (Sun Gro Horticulture) fertilized as needed with Peters 20-20-20. Plants were thinned to one plant per pot and grown at 10-20 cm spacing. Seed for all genotypes and populations was obtained from the Texas A&M Sorghum Breeding Program (College Station, TX).

Plant phenotyping and visualization of aerenchyma by microscopy and CT scanning

The extent of aerenchyma (AER) formation was rated by visual inspection of cross sections of internodes taken approximately at the midpoint between nodes. Ratings were based on a scale from 1 – 5 (1, no AER; 5, high AER) in a manner similar to prior studies (S-Figure 1) [41,121]. Results from visual rating was similar to ratings based on a percent of area calculation when data was used for QTL analysis [41]. All AER ratings are the mean of at least 3 biological replicates unless otherwise indicated.

Multi-slice computed tomography (MSCT) scans of five R07020 stems were collected at the Diagnostic Imaging & Cancer Treatment Center of the Texas A&M Veterinary Medicine & Biomedical Sciences facilities in College Station, Texas. A

SOMATOM Definition AS+ (SIEMENS) was used with settings of 120 kVp, 1.024 pixels per mm, at a 0.6 mm slice thickness [122].

For microscopic imaging, hand sections were made of representative R07020 internodes at different stages of internode development to investigate changes in the distribution and extent of aerenchyma accumulation in stems. The sections were imaged under white light using a Zeiss M2Bio Fluorescence Combination Zoom Stereo/Compound Microscope coupled with a Zeiss AxioCam digital camera (Kramer Scientific). Photographs were captured using the Zeiss AxioVision 3.0.6 software.

Quantitative trait locus (QTL) analysis and fine mapping

Stem aerenchyma QTL were mapped in three populations: BTx623 x IS3620c (RILs, n = 380), BTx623 x R07007 (F5, n = 215), BTx623 x Standard Broomcorn (F2, n = 133). All plants were grown to anthesis and phenotyped by cutting the internode below the peduncle at a midpoint between the upper and lower node. The cross-sections were visually inspected and rated on a scale from 1 (no aerenchyma) to 5 (high aerenchyma) (Figure 5). One plant of each RIL from the BTx623*IS3620c population was grown in the summer of 2014 under long days. Three plants of each line of the BTx623*R07007 population were grown in a greenhouse in the spring of 2015 under long days. Genotyping and genetic map construction were performed as described in Truong et al. [71] and DG-marker sequences were mapped to version 3.1 of the sorghum reference genome assembly (Sorghum bicolor v3.1 DOE-JGI, http://phytozome.jgi.doe.gov/), using BWA [123]. INDEL realignment and joint variant calling were performed with the GATK using the

naïve pipeline of the RIG workflow [124]. QTL mapping was performed in R/qtl using composite interval mapping (CIM) with 1000 permutations and $\alpha = 0.05$. BTx623*IS3620c and BTx623*R07007 populations were mapped as RILs while BTx623*Standard Broomcorn was mapped as an F2. The BTx623*IS3620c population was used to fine map *SbNAC_D*. Five plants from each line that had recombination breakpoints in the 2-LOD region spanning the main effect QTL on SBI-06 were grown to anthesis and phenotyped to confirm aerenchyma ratings.

DNA sequencing and sequence analysis

Coding region sequence variants between the four parental lines used for QTL analysis were obtained from publicly available whole genome sequences (Phytozome v12.1.6). The effect of protein sequence variants was analyzed using PROVEAN [125]. Genomic DNA was extracted from leaf tissue of diverse sorghum lines according to the Quick-DNA™ Plant/Seed Kit (Zymo Research) instructions. *SbNAC_D* was amplified by polymerase chain reaction (PCR) and was sequenced by Sanger sequencing. All template amplification and sequencing primers are listed in Table A-1.

Protein sequence analysis

Protein sequences of the closest homologs of *SbNAC_D* in grass species with available genome sequences were identified using BLAST (Phytozome v12). To identify the sugarcane homolog of *SbNAC_D*, the sorghum cDNA sequence was used in a BLAST search of the sugarcane transcriptome (SUCEST-FUN Database, http://sucest-fun.org). Protein sequences were aligned using MUSCLE [126] and visualized using Jalview [127].

Evolutionary trees were inferred using the Neighbor-Joining method [128] in MEGA7 [129]. All positions containing gaps and missing data were eliminated.

cDNA sequencing and qRT-PCR

Total RNA was extracted from all samples using the Direct-Zol™ RNA Miniprep Kit (Zymo Research), and cDNA was synthesized using SuperScript™ III First-Strand Synthesis SuperMix for qRT-PCR (ThermoFisher Scientific). For analysis of gene expression in *SbNAC_D* and *SbNAC_d1* RILs, the internode subtending the peduncle was collected before aerenchyma formation and 7 – 10 days later when aerenchyma were visible. Internodes were collected from 3 biological replicates. For the developmental analysis of gene expression in R07020, internodes were collected from 3 biological replicates at midday. Tissue was collected from the two short non-elongated internodes immediately below the shoot apex that were approximately 1-2 cm in length (int1, int2). Internodes 3 and 4 (int3-X, int4-X) were sub-sectioned into 1 cm sections. For the fully elongated internodes 5 – 7 (int5-X, int6-X, int7-X), 2-cm sections were isolated from the top and base of each internodes for analysis. A list of gene targets and primers used for the analyses is provided in Table A-2. For all qRT-PCR experiments, relative expression was determined using the comparative cycle threshold (Ct) method. Raw Ct values for each sample were normalized to Ct values for the reference gene SbUBC (Sobic.001G526600). Reference gene stability across all samples was determined by NormFinder (Andersen et al. 2004). ΔΔCt values were calculated relative to the sample with the highest expression (lowest Ct value). Relative expression values were calculated

with the $2^{-\Delta\Delta Ct}$ method [130]. Fold change in gene expression was calculated based on ΔCt values between the samples with the lowest and highest expression according to the equation $FC = 2^{(\Delta Ct(max) - \Delta Ct(min))}$. Primer specificity was tested by dissociation curve analysis and gel electrophoresis.

For cDNA sequencing, total RNA was extracted from the leaf sheaths of BTx623 and IS3620c and used for cDNA synthesis as described above. The cDNA corresponding to *SbNAC_D* was amplified and sequenced using Sanger sequencing. The primers used for template amplification and sequencing are listed in Table A-1.

Statistical analyses

All statistical analyses were performed in GraphPad Prism version 7 for Windows (GraphPad Software, La Jolla California USA, www.graphpad.com).

Results

Sorghum genotypes vary in stem aerenchyma accumulation

Sorghum genotypes were grown in a greenhouse and screened for variation in the extent of stem aerenchyma accumulation in order to identify genotypes for further study and parental lines of populations for QTL-assisted identification of genes involved in stem aerenchyma formation. The extent of aerenchyma formation was rated by visual inspection of stem cross sections taken mid-internode on a scale from 1 (no aerenchyma) to 5 (high aerenchyma) (Figure 5). This screen showed that the stems of BTx623 contained low levels of aerenchyma whereas the stems of IS3620c, R07007, R07020, and Standard

Broomcorn had higher levels of stem aerenchyma (Figure 6). For most genotypes, aerenchyma levels were similar in fully elongated internodes along the length of the stem. IS3620c was more variable and showed a trend towards higher aerenchyma in the internodes immediately below the peduncle (Int1-3). Standard Broomcorn stems contained the highest levels of stem aerenchyma among the genotypes screened.

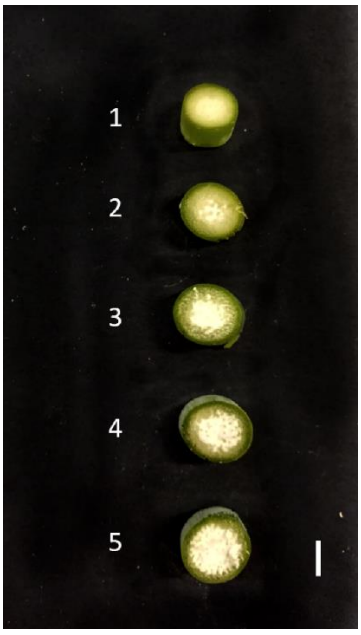

Figure 5. Stem aerenchyma rating scale, 1 (no aerenchyma) to 5 (high aerenchyma). Reprinted under the Creative Commons Attribution License from [35].

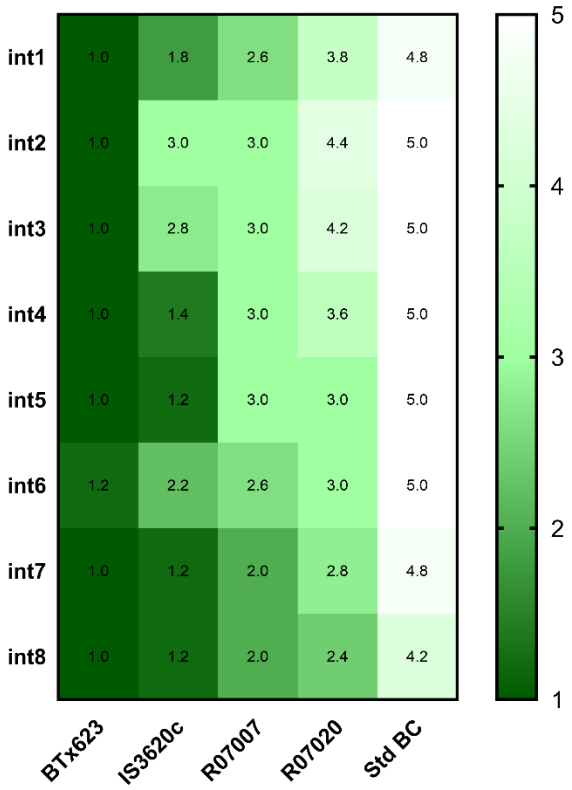

Figure 6. Variation in the extent of stem aerenchyma (AER) in diverse sorghum genotypes. Internodes of five sorghum genotypes were screened for the extent of AER formation in eight fully elongated internodes using a scale from 1 (minimal AER, dark green) to 5 (maximum AER, white) (heat map scale to the right). Internode 1 (int1) is the internode at the top of the stem below the peduncle. BTx623, and elite grain line, showed low levels of AER formation in all internodes. Standard Broomcorn (Std BC) had the highest levels of AER across all internodes among all screened genotypes. AER ratings are the average of at least 3 biological replicates. Reprinted under the Creative Commons Attribution License from [35].

QTL-assisted identification of the **D***-gene*

The genetic network that regulates the extent of stem aerenchyma accumulation was investigated using QTL analysis. The results of genotype screening (Figure 6) indicated that populations derived from crosses involving BTx623*IS3620c, BTx623*R07007, and BTx623*Standard Broomcorn would be useful for mapping QTL for variation in stem aerenchyma accumulation. Populations derived from each of these crosses were grown in a greenhouse and scored for variation in stem aerenchyma accumulation in internodes of fully elongated stem internodes. The stem aerenchyma phenotypes and pre-existing genetic maps for these populations [32,33,71] were used to identify QTL that affect the extent of aerenchyma formation in stems. A main effect QTL that modulates variation in stem aerenchyma accumulation was identified on SBI06 at approximately 50 Mbp in all three populations (Figure 7). The LOD scores of the QTL on SBI06 were >20 in the BTx623*IS3620c and BTx623*R07007 RIL populations and ~7.4 in the BTx623*Standard Broomcorn F2 population (Table 5). Two additional QTL with sub-threshold LOD scores located on SBI-01 and SBI-02 respectively were detected using the Standard Broomcorn population (Figure 7C). In all three populations, lines with homozygous BTx623 haplotypes spanning the QTL on SBI06 had low stem aerenchyma ratings and lines lacking BTx623 haplotypes accumulated the higher levels of stem aerenchyma (Figure 8).

Table 5. 2 LOD QTL intervals from three biparental mapping populations. Reprinted under the Creative Commons Attribution License from [35].

Population	Chr	2-LOD interval (Mb)	Peak position (Mb)	LOD score
BTx623/IS3620c	6	50820292 - 51044545	50.8	30.1
BTx623/R07007	6	48273018 - 51329068	50.8	43.6
BTx623/Std Broomcorn	6	50954277 - 51787450	51.5	7.37

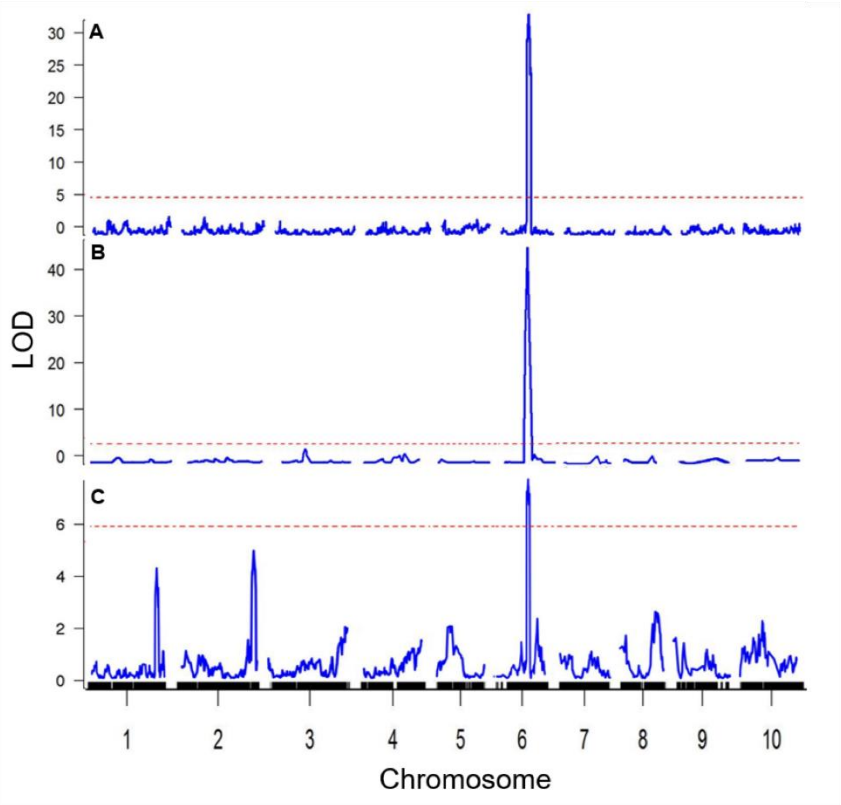

Figure 7. QTL associated with variation in aerenchyma (AER) formation were mapped in three biparental populations: BTx623*IS3620c (A, n = 380, RIL), BTx623*R07007 (B, n = 215, F5), and BTx623*Standard Broomcorn (C, n = 133, F2). The extent of AER formation in each line was rated visually in cross-sections at a midpoint of fully elongated internodes. Main effect QTL colocalized to chromosome 6 in all three populations where the BTx623 allele reduced stem AER formation. QTL analyses were performed in R/qtl using composite interval mapping. AER ratings in BTx623*IS3620c were from 1 biological replicate per RIL. AER ratings in BTx623*R07007 3 biological replicates per individual. Reprinted under the Creative Commons Attribution License from [35].

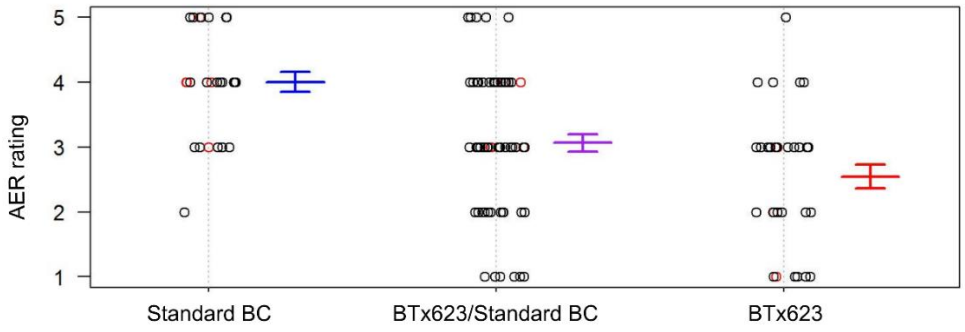

Figure 8. Phenotype-genotype plot of stem aerenchyma ratings in the F2 population BTx623*Standard Broomcorn at the marker with the highest LOD score in the QTL interval. The open circles represent AER ratings of individuals in the population. The average AER ratings of individuals genotyped as Standard BC at this marker were higher than the average rating of individuals genotyped as BTx623. Heterozygous individuals had intermediate phenotypes. Reprinted under the Creative Commons Attribution License from [35].

The QTL located on SBI06 that modulates stem aerenchyma accumulation in the current study is aligned with a QTL on LG-06 previously identified as encoding the D-gene [66]. The prior study was based analysis of 132 RILs derived from BTx623*IS3620c and a low-resolution genetic map. To fine map the D-locus, the current study utilized 380 RILs derived from BTx623*IS3620c [67] and a high-resolution genetic map constructed by scoring the segregation of ~10,091 DG-markers in the expanded RIL population [71]. A SNP marker located at 50,878,497 on SBI06 was tightly linked to the aerenchyma QTL (Figure 9A). Four RILs with breakpoints immediately flanking this stem aerenchyma QTL marker on SBI06 were identified (Figure 9B). SNP and INDEL markers useful for further fine mapping breakpoints within the D-locus in the four RILs were identified from an alignment of the IS3620c whole genome sequence and the BTx623 reference sequence

(Figure 9B) [28]. Sequence analysis of the markers in the four RILs delimited the D-locus to a 73kbp region (Figure 9B and 9C, Table 6). Based on gene annotation information in Phytozome (Sbiv3.1) the 73kbp region encodes a NAC (NAM, ATAF1/2, CUC) transcription factor, three genes encoding threonine aldolase, a MYB-like DNA-binding factor, and three genes encoding predicted (or unknown) proteins that lack functional annotation (Table 6).

Table 6. Genes in the fine-mapped QTL region. Reprinted under the Creative Commons Attribution License from [35].

Gene name	Description	Location	
Sobic.006G147400	No apical meristem protein (NAM)	Chr06:50896169..50898604	forward
Sobic.006G147450	Threonine aldolase	Chr06:50902561..50905501	forward
Sobic.006G147500	Predicted protein	Chr06:50914667..50916048	forward
Sobic.006G147600	Threonine aldolase	Chr06:50922630..50925976	reverse
Sobic.006G147700	Threonine aldolase	Chr06:50929529..50933978	reverse
Sobic.006G147800	Predicted protein	Chr06:50941894..50945361	forward
Sobic.006G147900	Predicted protein	Chr06:50948145..50951111	forward
Sobic.006G148000	Myb protein	Chr06:50959448..50962420	reverse

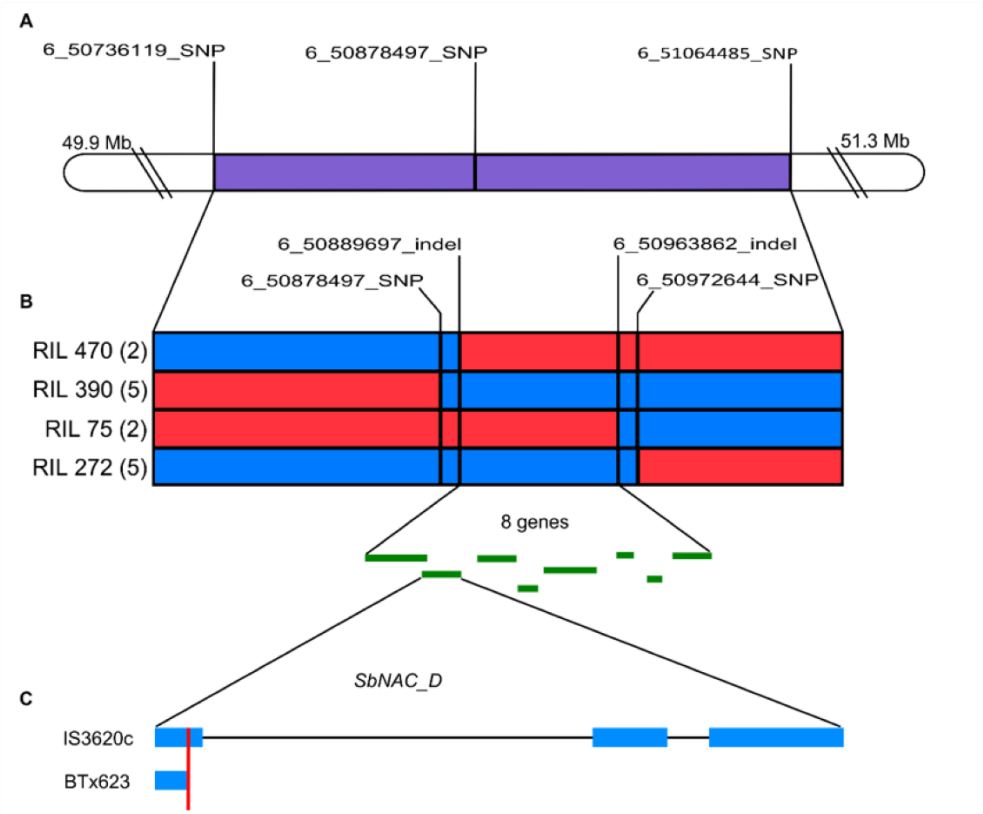

Figure 9. Positional cloning of SbNAC_D. (A) The marker 6_50878497_SNP had the highest LOD score in the QTL interval and was used to define a 2-LOD interval spanning the QTL (purple). (B) RILs from BTx623*IS3620c (RIL 470, RIL 390, RIL 75, RIL 272) with breakpoints within the 2-LOD interval were identified and phenotyped for the AER formation at anthesis. AER ratings are indicated in parentheses next to the RIL label to the left of the diagram. Blue haplotypes correspond to IS3620c alleles (+AER), and red haplotypes correspond to BTx623 alleles (-AER). The fine mapped interval contained 8 genes. (C) Sobic.006G147400 (SbNAC_D) was identified as the best candidate for the "D" gene. The BTx623 allele contains a T to C polymorphism that causes a Gln to STOP (Q39X) change in the first exon of BTx623 (red line). Reprinted under the Creative Commons Attribution License from [35].

The genes in the delimited *D*-locus were analyzed for the presence of sequence variants in coding regions to help identify candidate genes that could correspond to the *D*-gene. Gene intron/exon annotations were checked using RNAseq data (Phytozome v12)

and most of the gene annotations were supported by this analysis [28]. However, Sobic.006G147400 was annotated with 3 introns although RNAseq data and information from direct cDNA sequencing (Figure 10) provided evidence for only the two larger introns. Moreover, the incorrectly annotated intron in Sobic.006G147400 encodes amino acid sequences (V32 – E44) that are present in the sorghum cDNA and in homologs of this gene **(**Figure 10, Figure 11). Therefore, an annotation of Sobic.006G147400 that contains two introns was used for subsequent analysis of coding variants. A proposed intron in Sobic.006G147800 was also not well supported by RNAseq data possibly because expression of this gene was low resulting in incomplete coverage of the gene (Phytozome v12). A gene annotation lacking introns resulted in a continuous coding region through the predicted intron consistent with the annotation of a homolog of Sobic.006G147800 in Setaria that lacks an intron (Phytozome). Based on these results, the annotation of Sobic.006G147800 with and without the proposed intron was analyzed for coding variants.

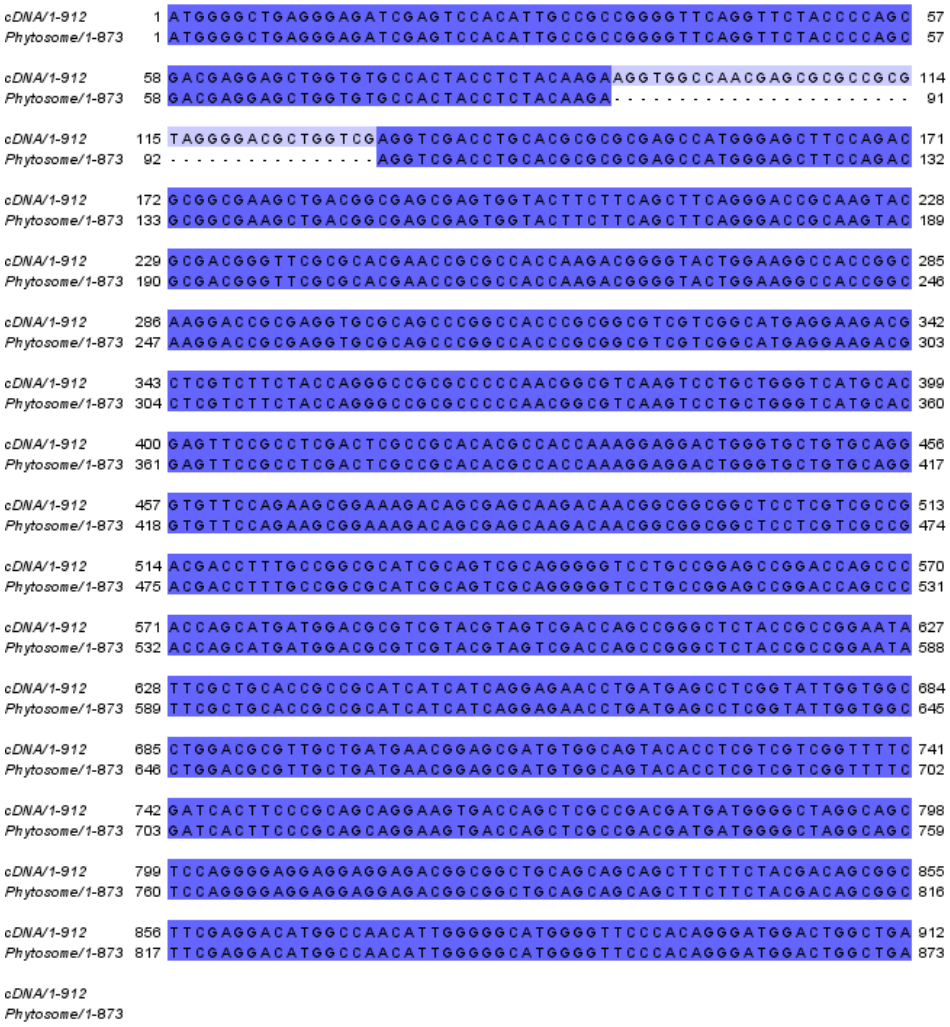

Figure 10. Transcript sequences of the candidate gene Sobic.006G147400 obtained from Sanger sequencing (top) and from Phytozome (bottom). The first exon of the Sobic.006G147400 transcript sequence available on Phytozome contains a 39-nucleotide deletion because those bases were misannoted as an intron, likely because of the stop codon mutation (at 115 nt) in BTx623. Reprinted under the Creative Commons Attribution License from [35].

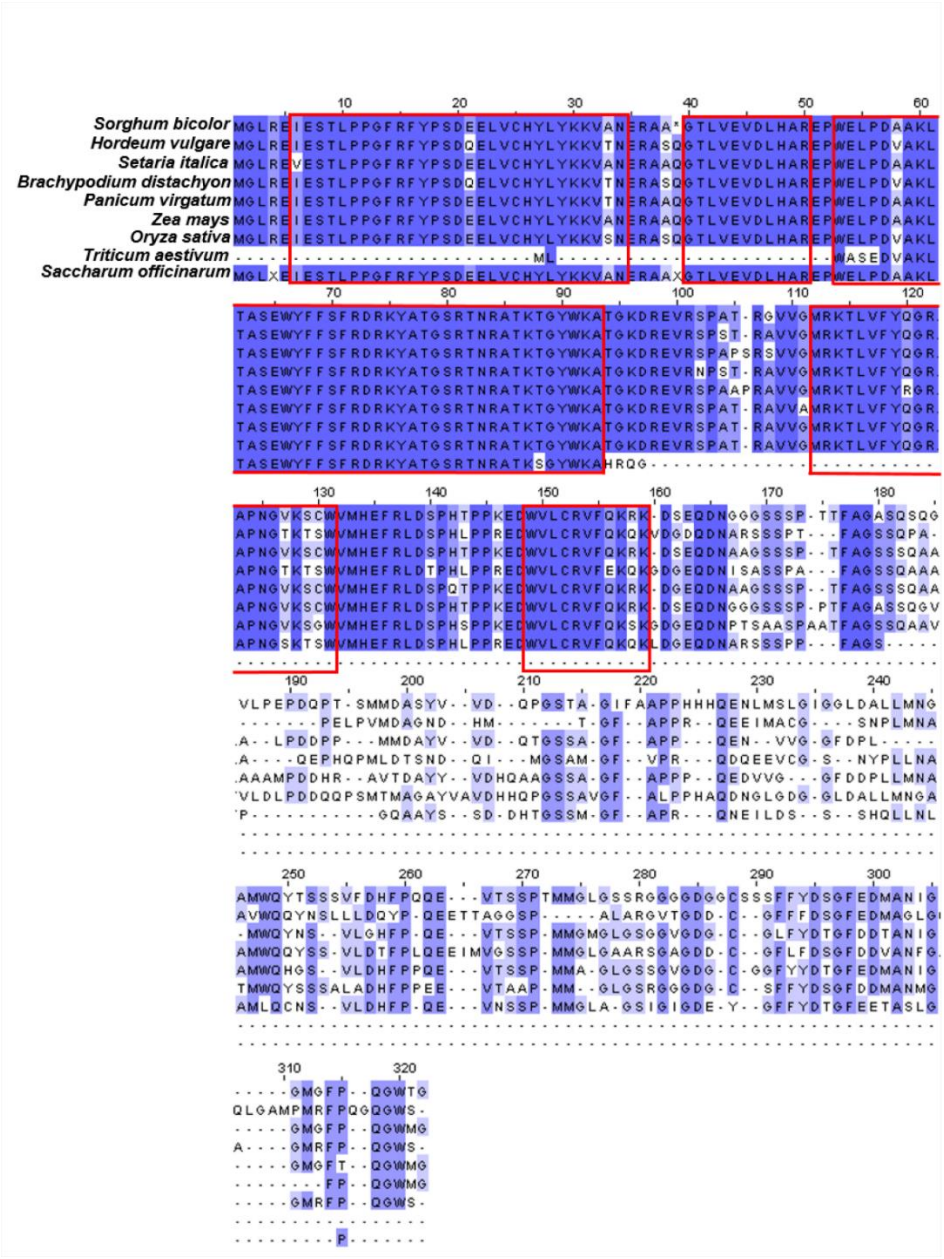

Figure 11. Protein sequence alignment of SbNAC_D homologs in select grass species. N-terminal sequences (1-170) of the SbNAC_D homologs are highly conserved, including protein motifs conserved in NAC transcription factors (highlighted by red boxes). Protein sequences were aligned using MUSCLE and visualized using Jalview. Reprinted under the Creative Commons Attribution License from [35].

Genetic analysis in the three populations is consistent with the prediction that there is a sequence variant (or variants) in BTx623 within the *D*-locus that inactivates a gene or regulatory element that normally increases stem aerenchyma formation and that this sequence variant is not present in IS3620c, R07007, and Standard Broomcorn. Polymorphisms were identified in the coding regions of the 8 genes within the delimited *D*-locus that distinguish the parental lines of the QTL mapping populations (BTx623, IS3620c, R07007, and Standard Broomcorn). In addition, coding variants were identified in Rio and SC170, genotypes that have low levels of stem aerenchyma similar to BTx623 (see below and Table 7). Coding variants were further analyzed to determine if they are predicted to alter protein function (Table 8) [125]. This analysis showed that three genes in the delimited *D*-locus lacked deleterious coding polymorphisms (Sobic.006g147450, Sobic.006g147500, Sobic.006g148000) that distinguished BTx623, Rio, SC1470, IS3620c, R07007, and Standard Broomcorn (Table 8). However, the analysis of gene sequences identified a sequence variant in the low aerenchyma lines BTx623, Rio, and SC170 that creates a stop codon (Q39X) in the NAC transcription factor encoded by Sobic.006g147400 (Table 8). In contrast, genotypes that accumulate higher levels of stem aerenchyma (IS3620c, R07007, and Standard Broomcorn) encoded full-length NAC proteins. The stop codon (Q39X) in Sobic.006G147400 present in BTx623, Rio, and SC170 removes motifs B – D of the NAC domain [131] and the rest of the C-terminal portion of the protein. Sobic.006G147600, a gene that encodes threonine aldolase, contained a change in amino acid sequence (S218P) in IS3620c and SC170 that was

predicted to cause a deleterious change in protein function by PROVEAN analysis [125]. SC170 has low aerenchyma while IS3620c has high aerenchyma, so this deleterious change in amino acid sequence is unlikely to be associated with variation aerenchyma formation (Table 7). Two other sequence variants in Sobic.006G147700 present only in Standard Broomcorn were predicted to cause deleterious changes in protein sequence. Sequence variants in Sobic.006G147800 were predicted to cause deleterious changes in protein function in R07007 and Standard Broomcorn, but these changes were also present Rio and SC170. Additionally, no coding region mutations in Sobic.006G147800 were identified in IS3620c, so variation coding variants in Sobic.006G147800 were not associated with the stem aerenchyma QTL in the population derived from BTx623*IS3620c. In summary, the analysis identified Sobic.006G147400 (NAC transcription factor) as the only gene with a deleterious coding variant in the low aerenchyma lines (BTx623, Rio, and SC170) that was not present in lines that accumulate higher levels of stem aerenchmya. Taken together, the analysis of coding variants indicated that Sobic.006g147400, a gene encoding a NAC transcription factor, contained a stop-codon in BTx623, Rio, and SC170 that was not present in IS3620c, R07007, and Standard Broomcorn, consistent with genetic prediction that BTx623 contains a mutation in a gene that is involved in induction of stem aerenchyma formation. As such we designated the two alleles of this gene as *SbNAC_D* and *SbNAC_d1* (the allele in BTx623) to distinguish this gene from other NAC-genes in sorghum.

Table 7. *SbNAC_D* alleles identified in various genotypes of *Sorghum bicolor*. Aerenchyma (AER) ratings were scored on a scale from 1 – 5 (1, no AER; 5, high AER) at a midpoint of an internode. Ratings are an average of at least 3 biological replicates. Reprinted under the Creative Commons Attribution License from [35].

Genotype	Allele	Amino acid change	AER rating
R07007	*SbNAC_D*	-	3-5*
R07020	*SbNAC_D*	-	4-5
IS3620c†	*SbNAC_D*	-	3-5
Standard Broomcorn	*SbNAC_D*	-	5
Feterita	*SbNAC_D*	-	3-4
TX08001 (BTx623/R07007)	*SbNAC_D/d1*	-	3
BTx623†	*SbNAC_d1*	Q39X	1-2
Blackhull Kafir	*SbNAC_d1*	Q39X	1-2
Tx3197	*SbNAC_d1*	Q39X	1-2*
RTx430	*SbNAC_d1*	Q39X	1
SC170	*SbNAC_d1*	Q39X	1
Hegari	*SbNAC_d1*	Q39X	1
Sweet sorghums			
Rio	*SbNAC_d1*	Q39X	1-2*
Wray	*SbNAC_d1*	Q39X	1
Della	*SbNAC_d1*	Q39X	1-2*
Collier	*SbNAC_d1*	Q39X	1-2
Grassl	*SbNAC_d1*	Q39X	1

†Allele sequence was determined by whole genome sequences. All others were determined by Sanger sequencing. *AER ratings were adapted from Carvalho et al. 2017.

Table 8. Coding region sequence variants in genes in the NAC_D QTL interval among the parental lines used in QTL mapping. Deleterious alleles are highlighted in red. Variant calls were taken from publicly available whole genome sequence data (Phytozome v12.1.6). The effect of each amino acid change was determined by PROVEAN analysis. Reprinted under the Creative Commons Attribution License from [35].

Gene ID	Position (Chr_Mb)	BTx623	IS3620c	R07007	Standard Broomcorn	Rio	SC170	Amino acid change	Effect
Sobic.006G147400	6_50896283	**T**	C	C	C	**T**	**T**	X39Q	Deleterious
Sobic.006G147450	6_50903023	A	A	A	C	C	C	T95P	Neutral
Sobic.006G147500									No variants
Sobic.006G147600	6_50924009	A	**G**	A	A	A	**G**	S218P	Deleterious
Sobic.006G147600	6_50924916	G	G	T	T	G	T	P128Q	Neutral
Sobic.006G147700	6_50931496	T	T	T	**A**	T	T	K267I	Deleterious
Sobic.006G147700	6_50932111	G	G	G	C	G	G	Q164E	Neutral
Sobic.006G147700	6_50932356	G	G	G	C	G	G	L122V	Neutral
Sobic.006G147700	6_50932467	C	C	C	**G**	C	C	F293L	Deleterious
Sobic.006G147800	6_50944184	T	C	C	C	C	C	F208S	Neutral
Sobic.006G147800	6_50944189	G	A	A	A	A	A	A210T	Neutral
Sobic.006G147800	6_50944209	A	C	C	C	C	C	L216F	Neutral
Sobic.006G147800	6_50944295	T	T	**C**	**C**	**C**	**C**	I245T	Deleterious
Sobic.006G147800	6_50944541	C	C	C	C	A	A	T327L	Neutral
Sobic.006G147800	6_50944760	T	T	**A**	**A**	**A**	**A**	V400D	Deleterious
Sobic.006G147800	6_50944922	T	T	C	C	C	C	L454S	Neutral
Sobic.006G147800	6_50945034	C	C	A	A	A	A	S491R	Neutral
Sobic.006G147800	6_50945071	C	C	**A**	**A**	C	**A**	P504T	Deleterious
Sobic.006G147900	6_50948305	G	G	C	C	G	C	A11P	Neutral
Sobic.006G147900	6_50948336	T	T	C	C	T	C	F21S	Neutral
Sobic.006G147900	6_50948375	T	T	C	C	T	C	V32A	Neutral
Sobic.006G147900	6_50948380	G	G	T	T	G	T	A36S	Neutral
Sobic.006G147900	6_50948386	G	G	T	T	G	T	A38S	Neutral
Sobic.006G147900	6_50950605	C	C	G	G	C	G	L111V	Neutral
Sobic.006G147900	6_50950719	A	A	G	G	G	G	N149D	Neutral
Sobic.006G147900	6_50950813	C	C	C	C	C	**T**	S180F	Deleterious
Sobic.006G148000	6_50960848	T	G	T	G	G	G	Y238S	Neutral

D-*alleles in sorghum germplasm*

Sequence analysis of Sobic.006g147400 from R07007, R07020, Standard Broomcorn and Feterita showed that these genotypes encode *SbNAC_D*, consistent with the accumulation of relatively high levels of stem aerenchyma in the stems of each genotype (Table 7). The energy sorghum hybrid, TX08001 derived by crossing A/BTx623*R07007 and therefore heterozygous for *NAC_D* alleles (*NAC_D*:*NAC_d1*) had an intermediate aerenchyma rating (Table 7). The recessive *SbNAC_d1* allele found in BTx623 was also present in other lines used in grain sorghum breeding including Blackhull Kafir, Tx3197, RTx430, SC170, and Hegari, genotypes that accumulate relatively low levels of aerenchyma (Table 7). The sweet sorghums Rio, Della, Wray, Collier and Grassl, genotypes that accumulate low levels of stem aerenchyma, also encode *SbNAC_d1* (Table 7). In this limited survey of sorghum germplasm, the presence of *SbNAC_D* in genotypes was correlated with high stem aerenchyma and *SbNAC_d1* with lower levels of stem aerenchyma.

***Phylogenetic analysis of* SbNAC_D**

NAC transcription factors were named for three genes initially discovered to contain NAC domains that contain N-terminal conserved motifs; NAM (no apical meristem), ATAF1 and -2, and CUC (cup-shaped cotyledon) [131,132]. NAC transcription factors are unique to plants and most plant species including sorghum encode >100 different NAC family members [133]. Phylogenetic analysis of grass genes encoding the *SbNAC-D*-like factors generated C4-grass and C3-grass groups (Figure 12). Protein

alignment showed that the grass NAC_D-like proteins were highly conserved across their N-terminal 170 amino acids that spans the NAC domain motifs A-E (Figure 11). Greater divergence was observed across the C-terminal portion of the grass NAC_D-like proteins consistent with what has been found for other NAC-proteins. The large N-terminal deletion present in a NAC_D homolog of wheat may be a consequence of selection for solid stems in this grain crop [121]. A NAC_D homolog of sugarcane was also truncated which may contribute to reduced aerenchyma formation in stems of sugarcane (Figure 11).

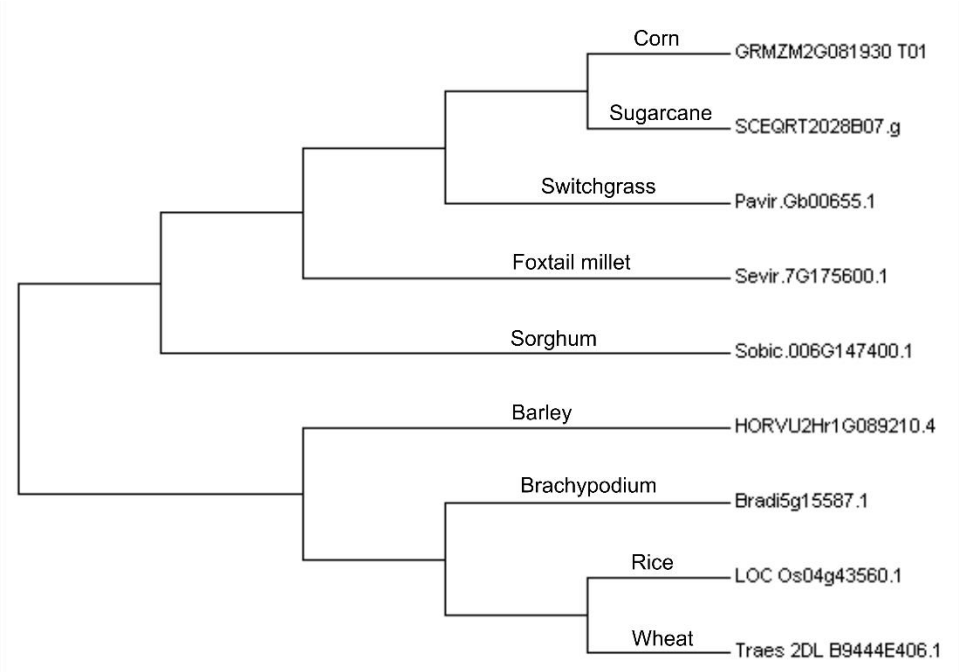

Figure 12. Phylogenetic analysis of grass homologs of SbNAC_D. Unroote phylogenetic analysis grouped SbNAC_D homolgs into C4 (i.e. sorghum, corn) and C3 (i.e. barley, rice) grasses. Relationships were inferred using hte Neighbor-Joining method. All positions containing gaps and missing data were eliminiated. Evolutionary analyses were conducted in MEGA7. Reprinted under the Creative Commons Attribution License from [35].

NAC family genes related to *SbNAC_D* were identified in sorghum (Figure 13A) and other plants (Figure 13B) to see if this would provide information about *SbNAC_D* function. The most closely related NAC-gene family member in Arabidopsis, *AtNAC074* (At4G28530), is differentially expressed to some extent in xylem compared to phloem cambium [134]. NAC factors such as *VND6/7*, *SND1*, and *NST1/2* play key roles in secondary cell wall formation and xylem differentiation, developmental processes that involve programmed cell death (PCD) [135]. NAC-genes also play a role in PCD-

associated with senescence [116]. Therefore, sorghum NAC-gene homologs of NAC-genes involved in secondary cell wall formation and senescence were identified and compared with *SbNAC_D* homologs from grasses (Figure 14). This showed that grass *NAC_D*-like genes grouped in a cluster distinct from NAC-genes involved in secondary cell wall formation and senescence. The analysis also showed that SbNAC_D was closely related to ZmNAC22 (GRMZM2G081930), a member of the ZmNAC-h (NAC1) subfamily [136]. This subfamily is distinct from the ZmNAC-j (VND6) and ZmNAC-I (SND/NST1) subfamilies involved in secondary cell wall formation. These results indicate that the function of *SbNAC_D* differs from NAC transcription factors involved in xylem secondary cell wall formation/xylem differentiation and senescence.

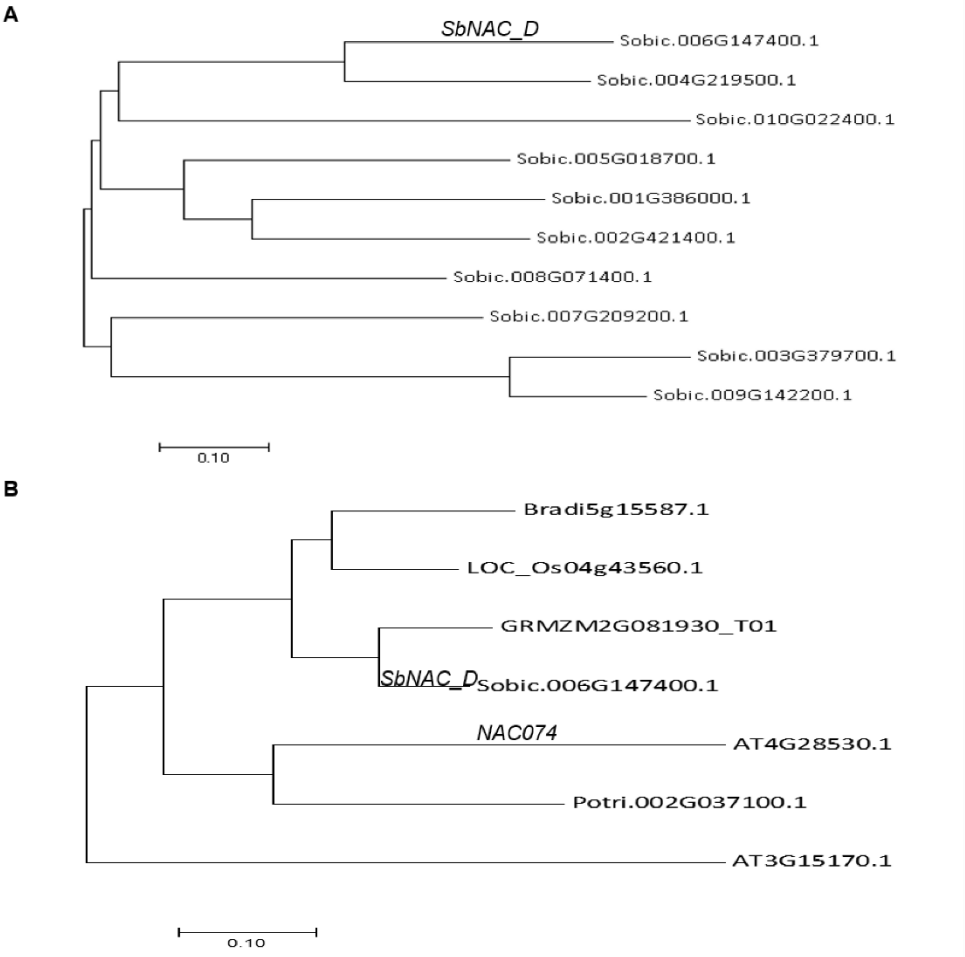

Figure 13. Phylogenetic relationships of SbNAC_D to (A) SbNAC_D paralogs in the sorghum genome and (B) SbNAC_D orthologs in other plant species. Relationships were inferred using the Neighbor-Joining method. All positions containing gaps and missing data were eliminated. Evolutionary analyses were conducted in MEGA7. Reprinted under the Creative Commons Attribution License from [35].

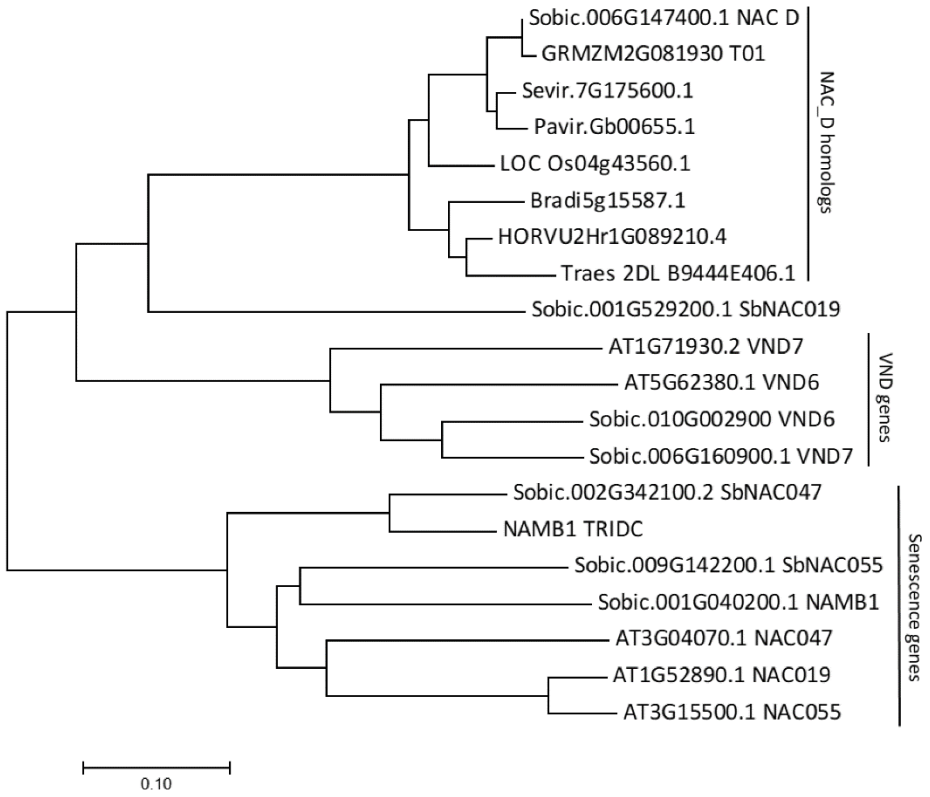

Figure 14. Phylogenetic relationships of SbNAC_D homologs in grasses to NAC-family VND genes involved in vascular differentiation and NAC-family genes involved in senescence. Relationships were inferred using the Neighbor-Joining method. All positions containing gaps and missing data were eliminated. Evolutionary analyses were conducted in MEGA7. Reprinted under the Creative Commons Attribution License from [35].

Expression of D-locus genes in RILs with BTx623 or IS3620c haplotypes

Sobic.006G147400 was identified as the best candidate gene for the *D*-gene among the genes located in the *D*-locus based on an analysis of coding sequence variants. However, it is possible that there are sequence variants among the parental genotypes used for QTL analysis that modify non-coding regulatory elements in the *D*-locus that affect

aerenchyma formation by altering the expression of one or more the 8-genes in the delimited *D*-locus. To check this possibility, the expression of the genes in the delimited *D*-locus was analyzed by qRT-PCR using RNA from stems of 3 pairs of RILs derived from BTx623*IS3620c with similar flowering times and dwarfing genotypes that either encoded the IS3620c *D*-locus haplotype or the BTx623 *d1*-locus haplotype. RNA was extracted from the top internode of the stem from each RIL shortly after full internode elongation for qRT-PCR analysis. The analysis showed no significant difference in expression of the 8 genes in the three pairs of RILs indicating that polymorphisms distinguishing BTx623 (*d1*) and IS3620c (*D*) in non-coding regions of the *D*-locus were not altering the expression of genes in the *D*-locus in this tissue (data not shown).

Developmental timing of stem aerenchyma accumulation and* SbNAC_D *expression

During the analysis above, it was observed that aerenchyma levels were low in internodes that were elongating, however 7-10 days after internodes reached full elongation, higher levels of aerenchyma were observed in *D*-genotypes. This indicated that stem aerenchyma formation is suppressed in nascent and growing internodes of sorghum stems and that the induction of aerenchyma formation occurs following internode elongation. If *SbNAC_D* is involved in stem aerenchyma formation, then expression of Sobic.006G147400 may increase following internode elongation in parallel with the formation of stem aerenchyma. To test this possibility, internodes from the three pairs of RILs with contrasting *D*-locus/*d1*-locus haplotypes were collected when they first reached full length (time point 1, low aerenchyma) and again 7-10 days later (time point 2) when

higher levels of aerenchyma were visible in genotypes encoding *SbNAC_D* (but not in *SbNAC_d1*). The results showed that Sobic.006G147400 expression was significantly higher in internodes that were forming stem aerenchyma (time point 2) in *SbNAC_D* genotypes (Figure 15). In addition, increased expression of Sobic.006G147400 was also observed 7-10 days after internode elongation in *SbNAC_d1* genotypes even though aerenchyma levels did not increase in these genotypes. This indicates that increased expression of Sobic.006G147400 was induced by development and not caused by the formation of aerenchyma. To complete the survey of gene expression, RNA extracted from the internodes from one pair of RILs (RIL49, *SbNACd1*; RIL392, *SbNAC_D*) was used to analyze expression the 7 other genes in the *D*-locus by qRT-PCR (Table 6). Only one of these genes, Sobic.006g148000, showed an increase in expression in stems between time point 1 and time point 2 in RIL392 (*SbNAC_D*) (Table 6). However, no deleterious coding sequence variants were identified in the parental lines used to fine map *D*-locus QTL (Table 8) suggesting that Sobic006G148000 plays some other role in internode development.

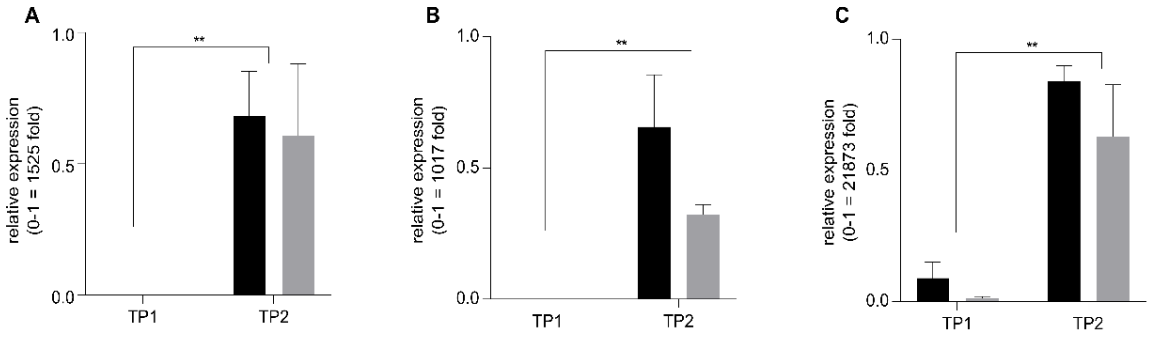

Figure 15. *SbNAC_D* expression in three pairs of *SbNAC_D*/*SbNAC_d1* RILs: (A) RIL226/RIL212, (B) RIL119/RIL474, (C) RIL392/RIL49. Expression was analyzed in the internode subtending the peduncle via qRT-PCR at time points before aerenchyma (AER) formation (TP1) and 7-10 days later when AER were visible in the internode (TP2). In all three RIL pairs, expression of *SbNAC_D* increased in parallel with AER formation. All gene expression was normalized to the expression of *SbUBC* in each sample. Relative expression was calculated via the $2^{-\Delta\Delta Ct}$ method relative to the internode section with the highest expression of the gene in each RIL pair. Expression values are the average of 3 biological replicates. Differences in expression values between TP1 and TP2 were analyzed by ANOVA in GraphPad Prism v7. **, $p < 0.005$. Reprinted under the Creative Commons Attribution License from [35].

The developmental timing of *SbNAC_D* expression and aerenchyma accumulation in sorghum stems was examined in more detail during vegetative phase stem growth using R07020, a tall, photoperiod-sensitive, late-flowering, high-biomass genotype. Growth of R07020 in 14h day-lengths represses floral initiation allowing analysis of stem aerenchyma formation during an extended period of vegetative growth after completion of the juvenile phase. R07020 internode elongation begins between ~45-55 days after germination and aerenchyma were observed in fully elongated internodes produced during vegetative development. Qualitative visual inspection of cross-sections showed that

aerenchyma were also present in the leaf midribs and leaf sheaths of R07020 that developed during the vegetative phase.

The timing of aerenchyma formation during growth and development of stem internodes was examined using cross-section analysis of R07020 stems (Figure 16A). The stems of rapidly growing R07020 plants contained two short nascent internodes immediately below the shoot apex (Int1, Int2), followed by an elongating internode (Int3) and several older fully elongated internodes (Int4-7) (Figure 16A, left). The two nascent non-elongated internodes (Int1, Int2) and the partially elongated internode 3 contained very low levels of aerenchyma (Figure 16B). Aerenchyma were present in the upper end of fully elongated Internode 4 (sections Int4-1 to Int4-6) but not at the basal end of this internode (Int4-9 to Int4-11). Aerenchyma were observed throughout older internodes (Int5-7), although at lower abundance near the base of internode 5 (Figure 16A and 16B).

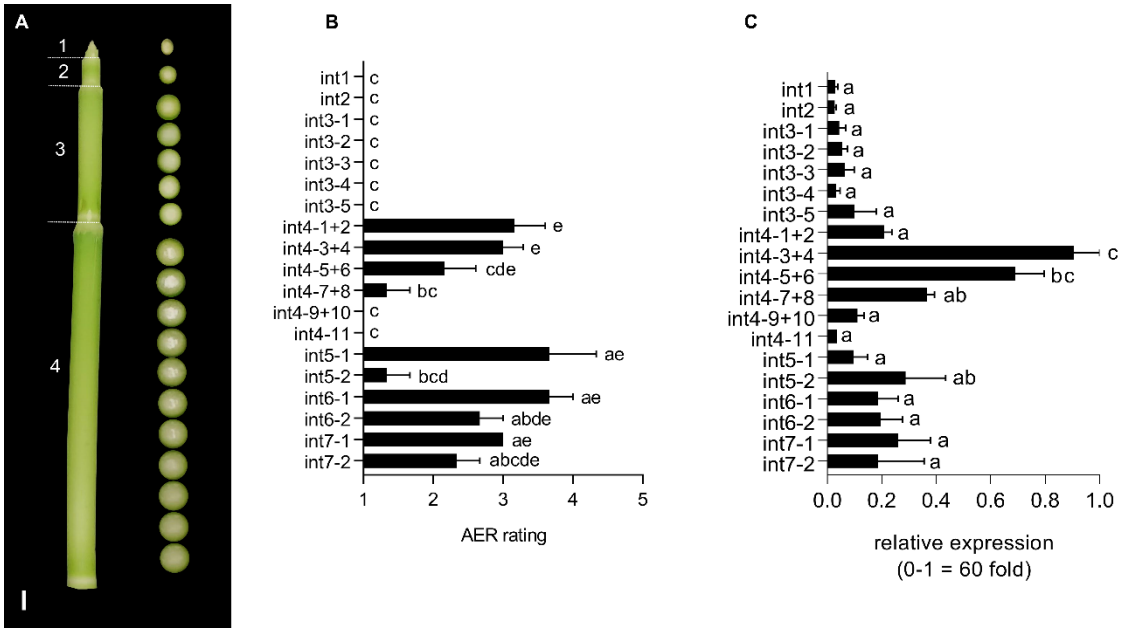

Figure 16. Aerenchyma (AER) formation and expression of *SbNAC_D* in stem sections of R07020. (A) Photograph of the top of a R07020 stem. Internode numbers are indicated to the left of the stem. Since R07020 remained in the vegetative phase throughout these experiments, internode 1 was defined as the youngest visible internode below the apical dome. Photographs of cross sections of each internode taken every 1 cm are shown to the right of the stem in (A). No AER is visible in internodes 1 - 3 or in the lower portion of internode 4. AER formation begins in the top of internode 4 in the center of the internode. AER forms in a basipetal manner and spreads radially outward. Scale bar = 1 cm. (B) AER was rated on a scale from 1 (no AER) to 5 (high AER). Ratings of adjacent pairs of sections of internode 4 (int4) were averaged. AER formation of internodes 5 - 7 was rated 2 cm from the top node of each internode (int5-1, int6-1, int7-1) and 2 cm from the bottom node of each internode (int5-2, int6-2, int7-2). Each rating is the average of 5 biological replicates. Tukey's multiple comparison test was used to compare ratings of each internode section. Sections with different letters are significantly different from each other, $p < 0.05$ (C) Relative expression of *SbNAC_D* across internode sections of R07020. Expression of *SbNAC_D* is correlated with AER formation (Spearman $r = 0.709$). Relative expression was calculated via the $2^{-\Delta\Delta Ct}$ method relative to the internode section with the highest expression. Expression values are the average of 3 biological replicates. Reprinted under the Creative Commons Attribution License from [35].

Expression of *SbNAC_D* during internode development in R07020 was analyzed using qRT-PCR (Figure 16B). Stem internode sections used for expression analysis were identical to those used for analysis of aerenchyma accumulation (Figure 16B). *SbNAC_D* was expressed at relatively low levels in the two youngest non-elongated internodes located immediately below the shoot apex that lacked aerenchyma (Int1, Int2) (Figure 16C). *SbNAC_D* expression increased to a small extent in Internode 3 and reached maximal levels in the upper portion of internode 4 (Figure 16C, sections Int4-1+2 to Int4-7+8), a region where aerenchyma were visible (Figure 16, Int4). Aerenchyma levels and *SbNAC_D* expression were low at the basal end of internode 4 (i.e., Int4, sections Int4-9 – Int4-11). *SbNAC_D* expression was also observed in tissue of older internodes (Int5-7). There was a strong correlation between *SbNAC_D* expression and aerenchyma levels during vegetative stem internode development ($\rho = 0.709$).

SbXCP1 *expression is correlated with aerenchyma formation and regulated by* SbNAC_D

The formation of stem aerenchyma could be mediated by programmed cell death (PCD). Induction of cysteine proteases is associated with PCD in rice [137] and has been correlated with developmental PCD in Arabidopsis [138]. Therefore, we identified the sorghum homolog (Sobic.007g172100, *SbXCP1*) of a rice PCD-regulated gene (Os02g48450) that encodes a cysteine protease and characterized its expression to see if it would be a useful marker for sorghum stem aerenchyma formation. RNA extracted from internodes of the pairs of RILs encoding either *SbNAC_D* or *SbNAC_d1* (RIL392/RIL49,

RIL474/RIL119, RIL226/RIL212) were used for qRT-PCR analysis of *SbXCP1* gene expression. In the RILs that encode *SbNAC_D, SbXCP1* expression increased in parallel with appearance of aerenchyma in internodes (Figure 17A, right). However, *SbXCP1* expression was not significantly induced in RILs that encode *SbNAC_d1* (Figure 17A, left) indicating that *SbNAC_D* is required for induction of *SbXCP1* expression in stems during aerenchyma formation.

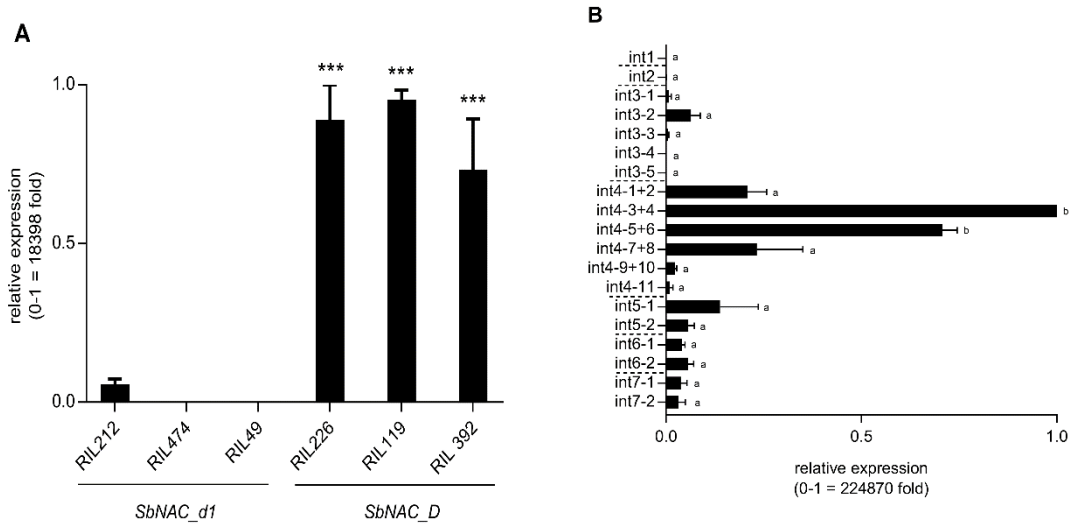

Figure 17. Expression of *SbXCP1* in (A) *SbNAC_d1/SbNAC_D* RILs and (B) R07020 internodes. SbXCP1 is a developmental programmed cell death (PCD) marker gene. (A) The expression of SbXCP1 is elevated in *SbNAC_D* RILs (RIL119, RIL392, RIL226) compared to *SbNAC_d1* RILs (RIL212, RIL474, RIL49). The uppermost internode of each RIL was sampled at a time point after aerenchyma (AER) formation had started. Expression differences between RILs was compared by Tukey's multiple comparison test, $p < 0.05$. (B) Expression of *SbXCP1* is low in internode sections where no AER formation is occurring (int1, int2, int3-X). Expression increases in sections of internode 4 (int4-X) where AER are forming. Tukey's multiple comparison test was used to compare ratings of each internode section. Sections with different letters are significantly different from each other, $p < 0.05$. All gene expression was normalized to the expression of *SbUBC* in each sample. Relative expression was calculated via the $2^{-\Delta\Delta Ct}$ method relative to the internode section with the highest expression of each gene. Expression values are the average of 3 biological replicates. Reprinted under the Creative Commons Attribution License from [35].

The correlation between internode development, aerenchyma formation, and expression of the PCD-marker gene *SbXCP1* was also investigated during R07020 stem development (Figure 17B). The relative expression of *SbXCP1* was low in internodes 1 and 2, with a small increase in internode 3, and highest expression in the mid to upper

portion of internode 4 (Int4, sections 3-8) (Figure 17B). Elevated but lower expression was observed in Internodes 5, 6 and 7. In general, variation in the expression of the PCD-marker gene *SbXCP1* was similar to *SbNAC_D* expression and matched the time course and location of aerenchyma formation in stems during internode development. Expression of six of the other genes in the delimited *D*-locus were also analyzed during internode development in R07020 (Figure 18).

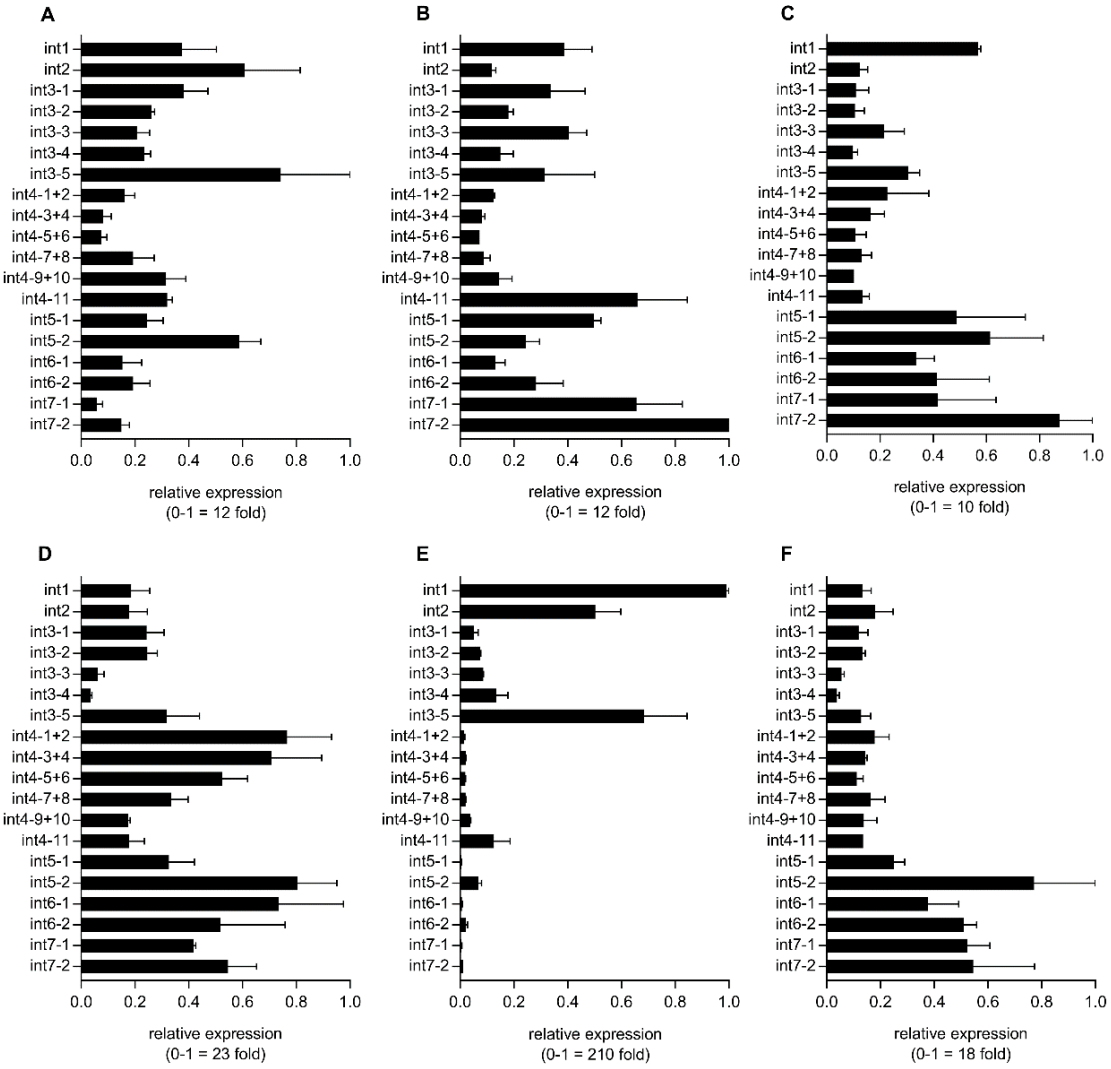

Figure 18. Expression of genes in the NAC_D QTL interval during the development of internodes of R07020 stems: (A) Sobic.006g147450, (B) Sobic.006g147500, (C) Sobic.006g147600, (D) Sobic.006g147800, (E) Sobic.006g147900, (F) Sobic.006g148000. Internodes were numbered from the apex of the plant downward. Internode sections are described in the Methods. Expression of the genes in the NAC_D QTL interval was not well correlated with AER formation. No expression could be detected from Sobic.006g147700 in R07020 tissues. Reprinted under the Creative Commons Attribution License from [35].

Spatial analysis of aerenchyma accumulation in sorghum internodes

Microscopic analysis of internode cross-sections and CT analysis showed that aerenchyma form first in pith parenchyma located between vascular bundles (VBs) in the center of the internode (Figure 19). As internode development proceeds, aerenchyma accumulation spreads radially, with relative few aerenchyma forming in the rind (Figure 19). Pith parenchyma cells located between VBs were initially converted to aerenchyma leaving islands of VBs and closely associated cells (Figure 19). At later stages of internode development, nearly all of the cells surrounding VBs located in the central region of the stem had been converted to aerenchyma (Figure 19).

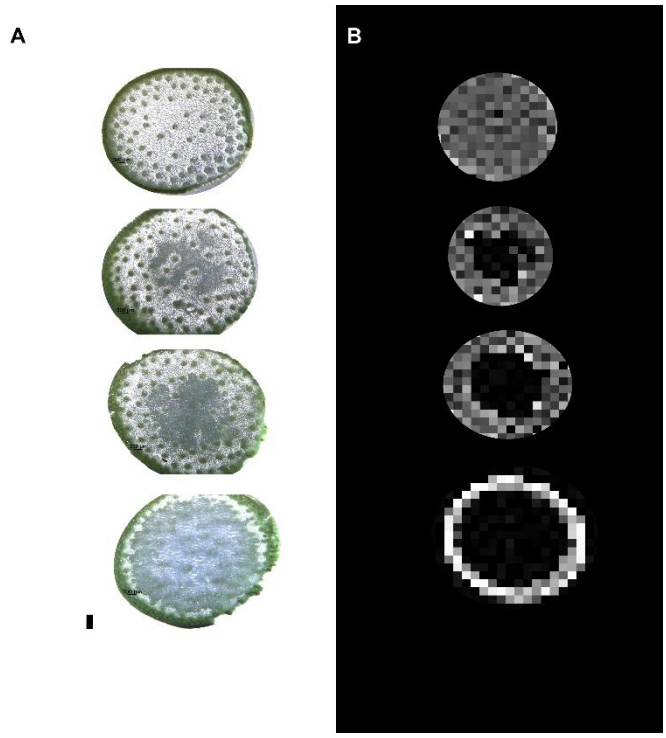

Figure 19. Microscopic (A) and CT scan (B) images of R07020 stem cross-sections at progressive stages of aerenchyma (AER) formation. From the top down, solid stem with no AER formation to high levels of AER formation across the stem cross section. AER formation begins in the pith parenchyma cells at the center of the stem and spreads radially outward toward the rind. Scale bar = 0.1 cm. Reprinted under the Creative Commons Attribution License from [35].

Discussion

Genotypic variation in sorghum stem aerenchyma formation was initially documented in 1916 [100]. This genetic determinant of variation in stem/leaf midrib aerenchyma accumulation was named the *D*-locus and further characterized in subsequent studies [101–103]. Sorghum genotypes with opaque white leaf midribs were reported to have drier, "pithy" stems whereas genotypes with green midribs had stems with low levels

of aerenchyma that accumulated higher amounts of sugars post flowering. More recent research identified QTL that modulate sorghum leaf-midrib color and stem moisture content [62,66,90,91,93,94,96,97]. In the current study, a main effect QTL on SBI06 was identified in three populations that has a large impact on the accumulation of stem aerenchyma. This QTL aligned with a leaf midrib color QTL previously designated as the *D*-locus ("D" for dry stems) [62,66,102]. QTL-assisted candidate gene identification and analysis of candidate gene coding variants and gene expression identified Sobic.006g147400, a gene encoding a NAC-transcription factor, as the best candidate for the *D*-gene (*SbNAC_D*).

Numerous important sorghum genes involved in fertility restoration, flowering time, seed shattering, and the regulation of stem growth have been cloned using QTL-assisted map-based gene identification [31–33,139–141]. This well-established approach was used in the current study to identify a gene and alleles that could explain variation in phenotypes associated with the *D*-locus. QTL that modulate the extent of stem aerenchyma accumulation were mapped in three sorghum populations derived from parents that differed in stem aerenchyma ratings. A single large effect QTL on SBI06 was identified in all three populations. Genetic analysis showed that BTx623, the female parent of each population, was the source of sequence variation that reduced aerenchyma accumulation in stems. Fine mapping delimited the *D*-locus to 73kbp, a region encoding 8 genes. The sequences of the 8 genes in the delimited *D*-locus derived from the four parental lines used for QTL analysis were examined to identify sequence variants that

could affect protein coding regions. This survey identified a disruptive sequence variant in one of the eight genes in BTx623 that was not present in the other parental lines used for QTL analysis. The key variant introduced a stop codon in the coding region of Sobic.006G147400 resulting in truncation of the NAC transcription factor encoded by this gene. No other disruptive coding variants were identified in BTx623 in genes in the *D*-locus. Disruptive mutations were not present in the coding regions of four of the genes in the delimited *D*-locus and disruptive coding variants found in the other 3 genes were present in one or two of the parental lines crossed to BTx623 (IS3620c, R07007, Standard Broomcorn) but not all three genotypes. Based on the analysis of coding variants, Sobic.006G147400 was tentatively identified as a candidate for the *D*-gene and the two alleles were named *SbNAC_D* and *SbNAC_d1*. The association of *SbNAC_D* genotype and stem aerenchyma phenotype was extended by screening a panel of diverse sorghum genotypes for *SbNAC_D* alleles and stem aerenchyma accumulation. This analysis showed that genotypes encoding *SbNAC_D* had higher levels of stem aerenchyma relative to genotypes encoding *SbNAC_d1*. Further evidence for the involvement of *SbNAC_D* in the formation of stem aerenchyma was obtained by analyzing the expression of *SbNAC_D* during stages of internode development before and after the onset of aerenchyma accumulation. Analysis of the internode located immediately below the peduncle in plants at flag leaf stage showed that *SbNAC_D* expression was low prior to the accumulation of aerenchyma and much higher 7-10 days later when internodes were accumulating aerenchyma. The correlation between increases in *SbNAC_D* expression and onset of

aerenchyma formation was also high in vegetative phase internodes of R07020. In contrast, the expression of the other 7 genes located in the delimited *D*-locus showed lower or no correlation with stem aerenchyma formation compared to *SbNAC_D*. In addition, expression of a PCD-marker gene encoding a cysteine protease (*SbXCP1*) was highly correlated with expression of *SbNAC_D* and the accumulation of aerenchyma in stems and expression was not induced in stems of genotypes encoding *SbNAC_d1*. The involvement of a NAC-transcription factor in stem aerenchyma formation is not surprising since NAC transcription factors have been associated with other types of PCD. For example, the NAC transcription factors VND6/7 play key roles in vascular tissue differentiation a developmental process that includes PCD [142]. Similarly, NAC transcription factors mediate PCD that occurs during senescence [113,116], root cap development (ANAC033) [111], and PCD associated with oxidative stress (ANAC13) [117]. Phylogenetic analysis showed that the grass homologs of *SbNAC_D* formed a cluster that was differentiated from clusters of NAC transcription factors involved in secondary cell wall formation and senescence. Taken together, these results are consistent with the identification of *SbNAC_D* (Sobic.006g147400) as the *D*-gene, a major modulator of aerenchyma formation in stems.

Sorghum genotypes vary in the timing and extent of aerenchyma formation in leaf midribs and stems during development [100]. In the current study, aerenchyma were not detected in non-elongated nascent internodes of vegetative phase plants and the onset of aerenchyma accumulation was associated with completion of internode elongation.

Intercalary meristems and zones of cell elongation are located at the base of elongating internodes in sorghum [143]. In vegetative phase plants, aerenchyma were first observed in the more fully developed upper portion of recently elongated internodes. As internode development proceeds, aerenchyma spread radially from the center of the stem towards the rind/epidermis and in a basipetal manner until most of the internode was filled with aerenchyma. The stem nodes generally lacked or had low levels of aerenchyma throughout development. Rice shows a similar pattern of aerenchyma accumulation during stem and internode development [144]. The timing of expression of *SbNAC_D* during stem and internode development was consistent with the gene's proposed role in stem aerenchyma formation. *SbNAC_D* was expressed at very low levels in the top two nascent internodes, internode growing zones, and nodes of developing internodes, tissues that accumulate only low levels of aerenchyma. *SbNAC_D* expression increased in the upper portion of fully expanded internodes in the same location where aerenchyma form first during development. Moreover, analysis of internodes derived from *SbNAC_D* and *SbNAC-d1* RILs that differ in the extent of stem aerenchyma formation differentially expressed a gene encoding a sorghum homolog of a cysteine protease similar to vignain (*SbXCP1*) that is involved in developmental programmed cell death in rice [137]. The expression of this gene increased in parallel with *SbNAC_D* expression and the appearance of stem aerenchyma in stem internodes during development. In contrast, expression of *SbXCP1* remained low in internodes of a sorghum genotype encoding *SbNAC_d*1 indicating that *SbNAC_D* regulates the expression of *SbXCP1* directly or indirectly. Taken together, these

results are consistent with the identification of *SbNAC_D* (Sobic.006g147400) as the *D*-gene, a major modulator of aerenchyma formation in stems.

Programmed cell death plays an important role in plant development, response to pathogens and abiotic stress [145,146]. The accumulation of aerenchyma in stem pith parenchyma cells provides a way to remobilize nutrients from these cells to support rapid growth of stems and leaves during the vegetative phase. This may be one reason aerenchyma formation is correlated with rapid stem elongation during the vegetative phase. The selective destruction of pith parenchyma cells located in the stem also reduces maintenance respiration. However, the destruction of these stem cells decreases the capacity of stems for sucrose storage that occurs in some genotypes when plants transition to the reproductive phase. By anthesis, the demand for nutrients to support stem and leaf growth has declined and excess sugars accumulate in the vacuoles of pith parenchyma stem cells. A large supply of stem sugars is useful during the grain filling phase, especially as a buffer during periods of water deficit and other stresses that inhibit photosynthesis.

The recessive *SbNAC_d1* allele was present in Blackhull Kafir, Hegari, and Collier, genotypes that were imported from Africa into the U.S. between 1881-1908 [147]. This indicates that the *SbNAC_d1* allele originated in Africa and was introduced at the onset of sorghum use in the U.S. consistent with the wide distribution of this allele in U.S. sorghum germplasm [147]. Following introduction into the U.S., Blackhull Kafir (*SbNAC_d1*) was used in the development of lines such as BTx3197 (*SbNAC_d1*) and BTx623 (*SbNAC_d1*) useful for grain sorghum hybrid production. SC170, a genotype

used in breeding R-lines such as RTx430, also encodes *SbNAC_d1*. Therefore, the historically important grain hybrid derived by crossing ATx623*RTx430 was homozygous for *SbNAC_d1*. The identification of *SbNAC_d1* in genotypes used for grain sorghum breeding suggests that reduced levels of stem aerenchyma is beneficial in these crops, possibly by reducing lodging or by increasing stem sugars that could provide a source of carbohydrate for grain filling under conditions of drought or other adverse environmental conditions.

The sweet sorghum genotypes Rio, Della, Wray, Collier and Grassl all encode *SbNAC_d1* alleles and the stems of these genotypes had lower levels of aerenchyma compared to genotypes encoding *SbNAC_D*. Reducing stem aerenchyma is important in sweet sorghum because retention of stem pith parenchyma increases the capacity for sucrose accumulation in the large vacuoles of these cells. Two previous sorghum genetic studies that characterized QTL for stem juiciness and sugar yield did not identify QTL aligned with *SbNAC_D* [60,92]. The parental genotypes used in these studies (BTx623*Rio, BTx3197*Rio) all encode *SbNAC_d1* consistent with lack of stem moisture or sugar content QTL aligned with the *D*-locus [60,92]. These previous studies identified other QTL that affect stem sugar content, concentration, and juiciness indicating that the genetic basis of these traits is complex [60,92].

Analysis of NAC_D homologs in other grasses showed that in the genotypes of wheat and sugarcane analyzed NAC_D homologs contained large deletions suggesting that breeders may have selected for inactive versions of NAC_D in these species as

occurred in sorghum. In sugarcane, selection for inactive alleles of NAC_D is consistent with low levels of stem aerenchyma, high stem juiciness, and high stem sucrose content in that crop [148]. Selection for more solid stems has also occurred in wheat in part to increase resistance to insects that damage stems although the genetic basis of insect resistance is complex [121].

The identification of alleles of *SbNAC_D* will make marker-assisted selection or genome editing of sorghum genotypes for various end uses easier, although screening for white/green leaf midribs as an indicator of *D*-gene activity is reliable and easy to implement. More importantly, the identification of *SbNAC_D* provides an entry point for the identification of factors that regulate the tissue specific expression of *SbNAC_D* during development and genes regulated by *SbNAC_D* that modulate aerenchyma formation. Variation in the extent of stem aerenchyma formation was observed in genotypes that encode *SbNAC_D* and low levels of variation in stem 'pithiness' were observed in populations derived from parental lines containing *SbNAC-d1* [92]. Follow on studies are needed to more fully elucidate the factors that regulate *SbNAC_D* expression, the developmental PCD pathway that leads to stem aerenchyma formation and other genes that modulate stem and leaf midrib aerenchyma formation.

CHAPTER IV

MATURITY2, A NOVEL REGULATOR OF FLOWERING TIME IN *SORGHUM BICOLOR,* INCREASES EXPRESSION OF *SBPRR37* AND *SBCO* IN LONG DAYS DELAYING FLOWERING

Introduction

Sorghum bicolor is a drought resilient, short-day C4 grass that is grown globally for grain, forage and biomass [40,76,149,150]. Precise control of flowering time is critical to achieve optimal yields of sorghum crops in specific target production locations/environments. Sorghum genotypes that have delayed flowering in long days due to high photoperiod sensitivity are high-yielding sources of biomass for production of biofuels and specialty bio-products [39,76]. In contrast, grain sorghum was adapted for production in temperate regions by selecting genotypes that have reduced photoperiod sensitivity resulting in earlier flowering and reduced risk of exposure to drought, heat, or cold temperatures during the reproductive phase. A range of flowering times are found among forage and sweet sorghums [151]. Sweet sorghum genotypes with longer vegetative growth duration have larger stems that have greater potential for sucrose accumulation [151–153].

Flowering time is regulated by development, day length, phytohormones, shading, temperature, and the circadian clock [73,154,155]. In the long-day plant *Arabidopsis thaliana*, circadian and light signals are integrated to increase the expression of *FLOWERING LOCUS T* (*FT*) and flowering in long days. *FT* encodes a signaling protein

synthesized in leaves that moves through the phloem to the shoot apical meristem (SAM) where it interacts with *FLOWERING LOCUS D* (*FD*) and reprograms the vegetative shoot apical meristem for reproductive development [156,157]. Expression of circadian clock genes such as *LATE ELONGATED HYPOCOTYL* (*LHY*) and *TIMING OF CAB1* (*TOC1*) regulate the expression of the clock output gene *GIGANTEA* (*GI*) and genes in the flowering time pathway [158–160]. Photoperiod and circadian clock signals are integrated to control the expression and stability of CONSTANS (CO) an activator of *FT* expression [161]. Under inductive long day (LD) photoperiods, CO promotes the expression of *FT* which induces flowering in *Arabidopsis* [162].

Many of the genes in the *Arabidopsis* flowering time pathway are found in sorghum and other grass species such as *Oryza sativa* (rice) [73] and maize [163], however, the regulation of flowering time in these grasses has diverged from Arabidopsis in several important ways. Both rice and sorghum are facultative short-day (SD) plants. In rice, the expression of the *FT-like* gene *Heading date 3a* (*Hd3a*) is promoted in SD [164]. In sorghum, expression of two different *FT-like* genes, *SbCN8* and *SbCN12*, is induced when plants are shifted from LD to SD [30,165]. In contrast to *Arabidopsis*, the rice and sorghum homologs of *CO* (rice *Heading date1, OsHd1; SbCO*) repress flowering in LD [73,166]. Rice and sorghum encode two additional grass-specific regulators of flowering *Ehd1* and *Ghd7*. *Early heading date1* (*Ehd1*) activates the expression of *FT*-like genes, and *Grain number, plant height and heading date7* (*Ghd7*) represses the expression of

EHD1 and flowering [167,168]. When sorghum is grown in short days, *SbEhd1* and *SbCO* induce the expression of *SbCN8* and *SbCN12*, leading to floral induction [30,31,165,169].

Under field conditions, time to flowering in sorghum varies from ~50 to >150 days after planting (DAP) depending on genotype, planting location (latitude/day-length), and the environment. A tall and "ultra-late" flowering sorghum variety called Milo Maize was introduced to the United States in the late 1800s [170]. Shorter and earlier flowering Milo genotypes such as Early White Milo and Dwarf Yellow Milo were selected from the introduced Milo genotype to promote improved grain yield in temperate regions of the US [40,170,171]. Genetic analysis determined that mutations in three independently segregating *Maturity* (*Ma*) loci (Ma_1, Ma_2, Ma_3) were responsible for early flowering times in the Milo genotypes. A cross between Early White Milo ($ma_1Ma_2Ma_3$) and Dwarf Yellow Milo ($Ma_1ma_2ma_3$) was used to construct a set of Milo maturity standards, a series of nearly isogenic lines that differ at one or more of the *Maturity* loci (Quinby and Karper 1945, Quinby 1966, Quinby, 1967). A fourth *Maturity* locus (Ma_4) was discovered in crosses of Milo (Ma_4) and Hegari (ma_4) [44]. More recent studies identified Ma_5 and Ma_6 [45]. Subsequent research showed that all of the Milos are dominant for Ma_5 and recessive for ma_6 [31,166]. In addition to these six *Ma* loci, many other flowering time quantitative trait loci (QTL) have been identified in sorghum [61,64,66,149,172]. Additional research has linked several of these QTL to genes such as *SbEhd1* and *SbCO* that are activators of *SbCN8* and *SbCN12* expression, sources of florigen in sorghum.

The genes corresponding to four of the six *Maturity* loci have been identified. *Ma_1*, the locus with the greatest influence on flowering time photoperiod sensitivity, encodes *SbPRR37*, a pseudo-response regulator that inhibits flowering in LD [30]. *Ma_3* encodes phytochome B [29], *Ma_5* encodes phytochrome C [166], and *Ma_6* encodes *Ghd7* a repressor of flowering in long days [31]. The genes corresponding to *Ma_2* and *Ma_4* have not been identified but recessive alleles at either locus results in early flowering in long days in genotypes that are photoperiod sensitive (*Ma_1*) [170]**.** Prior studies also noted that genotypes recessive for *Ma_2* flower later in genotypes that are photoperiod insensitive and recessive for *Ma_1* and *Ma_6* [170].

In this study, the impact of *Ma_2* alleles on the expression of genes in the sorghum flowering time pathway was characterized. A QTL corresponding to *Ma_2* was mapped and a candidate gene for *Ma_2* identified by fine mapping and genome sequencing. The results show that *Ma_2* enhances *Ma_1* (*SbPRR37*) and *SbCO* expression consistent with the impact of *Ma_2* alleles on flowering time in genotypes that vary in *Ma_1* alleles.

Methods

Plant growing conditions and populations

The cross of 100M and 80M was carried out by the Sorghum Breeding Lab at Texas A&M University in College Station, TX. F_1 plants were grown in the field in Puerto Rico and self-pollinated to generate the F_2 population used in this study. The 100M/80M

F_2 population was planted in the spring of 2008 at the Texas Agricultural Experiment Station in College Station, TX

The cross of Hegari and 80M was made in the greenhouse at Texas A&M University in College Station, TX. F_1 plants were confirmed and self-pollinated to generate the F_2 population used in this study. The Hegari/80M F_2 population (n = 432) was planted in the spring of 2011 in the greenhouse in 18 L nursery pots in a 2:1 mixture of Coarse Vermiculite (SunGro Horticulture, Bellevue, WA) to brown pasture soil (American Stone and Turf, College Station, TX). All subsequent generations of Hegari/80M for fine mapping were grown in similar conditions. Greenhouse-grown plants were watered as needed and fertilized every two weeks using Peters general purpose 20-20-20 (Scotts Professional).

For circadian gene expression experiments, 100M and 80M genotypes were planted in MetroMix 900 (Sungro Agriculture) in 6 L pots, and thinned to 3 plants/pot after 2 weeks. Plants were grown in the greenhouse under 14 h days until 30 days after planting (DAP). After 30 days, the plants were moved into growth chambers and allowed to acclimate for 3 days. The growth chamber was set to 30°C and 14/10h L/D for the 3 days of entrainment and the first 24 h of tissue collection. The lights were changed to constant light for the second 24 h of tissue collection.

QTL mapping and multiple-QTL analysis

DNA was extracted from leaf tissue for all individuals described above as described in the FastDNA Spin Kit manual (MP Biomedicals). All individuals in each

mapping or HIF population were genotyped by Digital Genotyping using FseI digestion enzyme as described in Morishige et al [58]. DNA fragments were sequenced using the Illumina GAII platform and the reads were mapped back to the sorghum reference genome (v1.0, Phytozome v6). Genetic maps were created using MapMaker 3.0B with the Kosambi function [173]. QTL were mapped using WinQTLCartographer (v2.5.010) using composite interval mapping with a 1.0 cM walk speed and forward and backward model selection [174]. The threshold was set using 1000 permutations and $\alpha = 0.05$. Upon release of v3.1 of the sorghum reference genome, the QTL coordinates were updated [28].

To look for possible gene interactions multiple-QTL analysis was used in the Hegari/80M F_2 population. A single QTL analysis using the EM algorithm initially identified two primary additive QTL which were used to seed model selection. The method of Manichaikul et al. [175] was employed for model selection as implemented in R/qtl for multiple-QTL analysis [79]. Computational resources on the WSGI cluster at Texas A&M were used to calculate the penalties for main effects, heavy interactions, and light interactions. These penalties were calculated from 24,000 permutations for flowering time to find a significance level of 5% in the context of a two-dimensional, two-genome scan.

Fine mapping of the Ma_2 QTL

All fine mapping populations for the *Ma_2* QTL were derived from F_2 individuals from the Hegari/80M population. The genetic distance spanning the *Ma_2* locus is 2 cM corresponding to a physical distance of ~1.8 Mbp, so 1000 progeny would be required to

obtain 20 recombinants within the *Ma2* QTL region. Six individuals that were heterozygous across the *Ma2* QTL were self-pollinated to generate six heterogeneous inbred families (HIFs) totaling 1000 F_3 individuals. These individuals were grown out in the greenhouse, and flowering time was recorded. They were genotyped by Digital Genotyping as described above [176]. Two F_3 individuals that had useful breakpoints with a heterozygous genotype on one side of the breakpoint were grown and self-pollinated to generate an additional round of HIFs (F_4, n = 150) that were planted in the spring of 2013 and analyzed as described above. No new breakpoints were identified in the F_4 generation, so this process was repeated again to generate F_5 plants in the spring of 2014.

Circadian gene expression analysis

For the circadian gene expression analysis, 30-day-old plants were placed in a growth chamber set to 14 h days for the first 24 h and constant light for the second 24 h at 30°C. Plants were entrained for 3 d before beginning tissue collection. Leaf tissue was collected and pooled from 3 plants every 3 h for 48 h. The experiment was repeated three times for a total of three biological replicates. RNA was extracted from each sample using the Direct-Zol™ RNA Miniprep Kit (Zymo Research) according to the kit instructions. cDNA was synthesized using SuperScript III kit for qRT-PCR (Invitrogen) according to the kit instructions. Primers for sorghum flowering pathway genes were developed previously, and primer sequences are available in Murphy et al [30]. Primer sequences for *Ma2* are available in Table A-3. Relative expression was determined using the comparative cycle threshold (C_t) method. Raw C_t values for each sample were normalized to C_t values

for the reference gene *SbUBC* (Sobic.001G526600). Reference gene stability was determined previously [35]. $\Delta\Delta C_t$ values were calculated relative to the sample with the highest expression (lowest C_t value). Relative expression values were calculated with the $2^{-\Delta\Delta Ct}$ method [177]. Primer specificity was tested by dissociation curve analysis and gel electrophoresis of qRT-PCR products.

Ma$_2$ phylogenetic analysis

Protein sequences of the closest homologs of Ma$_2$ were identified using BLAST analysis. Protein sequences were aligned using MUSCLE [126] and visualized using Jalview [127]. Evolutionary trees were inferred using the Neighbor-Joining method [178] in MEGA7 [129]. All positions containing gaps and missing data were eliminated.

Ma$_2$ DNA sequencing and whole genome sequence analysis

Whole genome sequence reads of 52 sorghum genotypes including 100M and 80M were obtained from Phytozome v12. Base quality score recalibration, INDEL realignment, duplicate removal, joint variant calling, and variant quality score recalibration were performed using GATK v3.3 with the RIG workflow [124]. Sobic.002G302700 was sequenced via Sanger sequencing in the genotypes in Table 9 according to the BigDye Terminator Kit (Applied Biosystems). Primers for template amplification and sequencing are provided in Table A-3.

Results

Effects of Ma_2 alleles on flowering pathway gene expression

The recessive ma_2 allele in 80M ($Ma_1\mathbf{ma_2}Ma_3Ma_4Ma_5ma_6$) was previously reported to cause 80M to flower earlier than100M ($Ma_1\mathbf{Ma_2}Ma_3Ma_4Ma_5ma_6$) in long days [170]. To help elucidate how Ma_2 modifies flowering time, we investigated the impact of Ma_2 alleles on the expression of genes in sorghum's flowering time pathway. Gene expression was analyzed by qRT-PCR using RNA isolated from 100M (Ma_2) and 80M (ma_2) leaves collected every 3 hours for one 14h light/10h dark cycle and a second 24-hour period of constant light.

SbPRR37 is a central regulator of photoperiod sensitive flowering in sorghum that acts by repressing the expression of *SbCN* (*FT*-like) genes in LD [30]. *SbPRR37* expression in 100M and 80M grown in long days peaked in the morning and again in the evening as previously observed [30] (Figure 20). The amplitude of both peaks of *SbPRR37* expression was reduced in 80M (ma_2) compared to 100M (Ma_2) (Figure 20A). *SbCO* also shows peaks of expression in the morning (dawn) and in the evening (~14h) [30] (Figure 20B). Analysis of *SbCO* expression in 100M and 80M showed that both peaks of *SbCO* expression were reduced in 80M compared to 100M (Figure 20B).

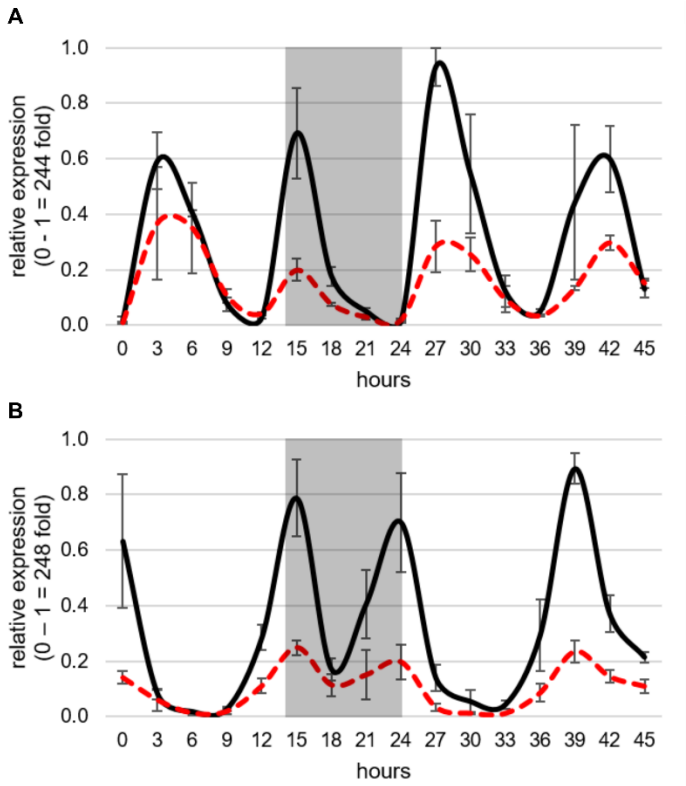

Figure 20. Circadian expression of genes regulating flowering in *S. bicolor* in 100M and 80M under long days. (A) Expression of *SbPRR37* in 100M (solid black lines) and 80M (dashed red lines). The expression peaks of *SbPRR37* are reduced in 80M. This is consistent with earlier flowering in 80M because *SbPRR37* represses the expression of the sorghum *FT*-like genes. (B) Expression of *SbCO* in 100M and 80M. Expression peaks of *SbCO* are also reduced in 80M. This is consistent with earlier flowering in 80M because under long days *SbCO* is a repressor of flowering. All expression values are normalized to *SbUBC* and are the mean of 3 biological replicates.

SbCN8, *SbCN12*, and *SbCN15* are homologs of *AtFT* that encode florigens in sorghum [165]. Expression of *SbCN8* and *SbCN12* increases when sorghum plants are shifted from LD to SD, whereas *SbCN15* shows minimal response to day length [30,31]. SbPRR37 and SbCO are co-repressors of the expression of *SbCN8* and *SbCN12* in long

days, therefore, the influence of *Ma2* alleles on *SbCN8/12/15* expression was investigated [30,169]. When plants were grown in long days, expression of *SbCN12* was ~4 fold higher in 80M compared to 100M consistent with earlier flowering in 80M (Figure 21).

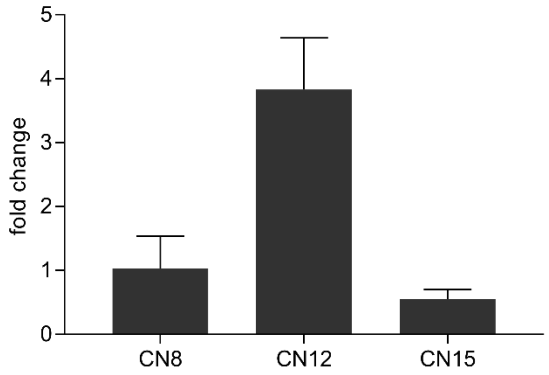

Figure 21. Expression of the *S. bicolor FT*-like genes *SbCN8*, *SbCN12*, and *SbCN15* in long days at the expected peak of expression. Expression of *SbCN12* was elevated in 80M, which is consistent with earlier flowering in that genotype. All expression values are normalized to *SbUBC* and are the mean of 3 biological replicates. Fold change was calculated as $2^{-[Ct(100M)-Ct(80M)]}$. Error bars are SEM.

Previous studies showed that SbGHD7 represses *SbEHD1* expression and that alleles of *SbGHD7* differentially affect *SbCN8* expression (>SbCN12) [31]. Analysis of *SbEHD1* and *SbGHD7* expression in 100M and 80M showed that *Ma2* alleles modify the expression of these genes only to a small extent (Figure 22C and 22D).

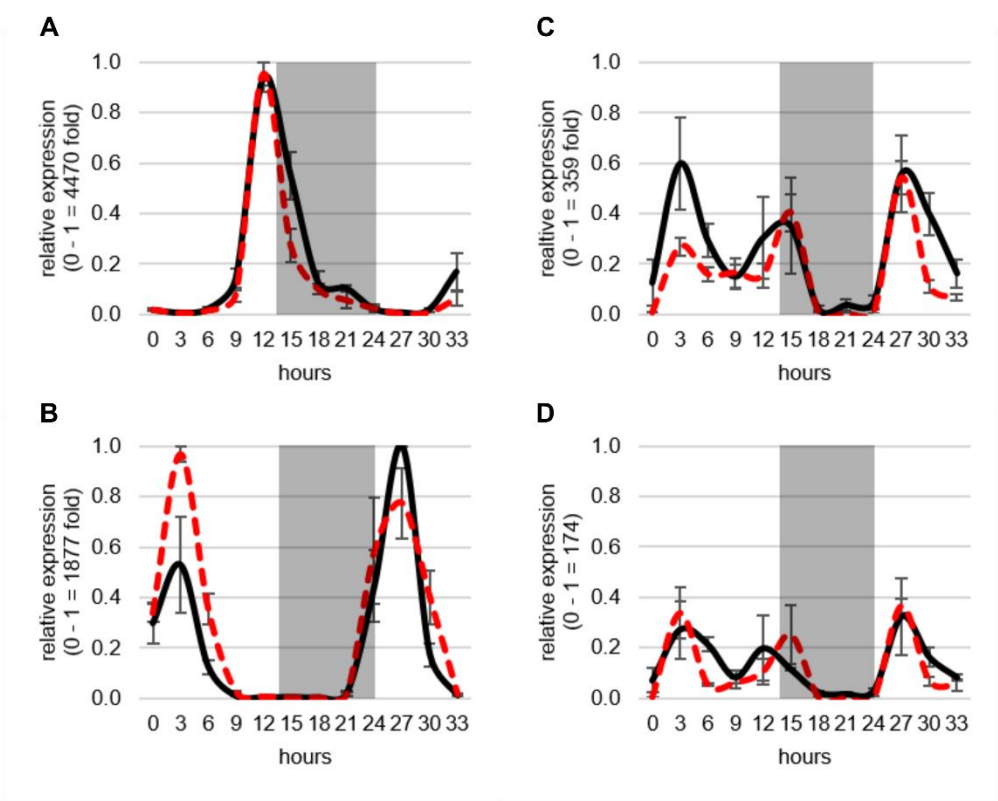

Figure 22. Circadian expression of *SbTOC1, SbLHY, SbGhd7*, and *SbEhd1*. There were no consistent differences in expression of (A) *SbTOC1*, (B) *SbLHY*, (C) *SbGhd7*, and (D) *SbEhd1* between 100M (solid black line) and 80M (dashed red line).

The timing of the two daily peaks of *SbPRR37* and *SbCO* expression in sorghum is regulated by the circadian clock [30,31]. Therefore, it was possible that *Ma2* modifies *SbPRR37/SbCO* expression by altering clock gene expression. However, expression of the clock genes *TOC1* and *LHY* was similar in 100M and 80M (Figure 22A and 22B). Taken together, these results show that Ma_2 is an activator of *SbPRR37* and *SbCO* expression in long days. Prior studies showed that co-expression of *SbPRR37* and *SbCO* in long days inhibits expression of *SbCN12* and floral initiation [169]. Later flowering in sorghum

genotypes that are *Ma1Ma2* vs. *Ma1ma2* in long days is consistent with lower *SbCN12* expression in *Ma1Ma2* genotypes.

Genetic analysis of Ma2 and Ma4

An F_2 population derived from a cross of 100M (*Ma2*) and 80M (*ma2*) was generated to map the *Ma2* locus. Because 100M and 80M are nearly isogenic lines that differ at *Ma2*, only *Ma2* alleles were expected to affect flowering time in this population [170]. The F_2 population (n = ~1100) segregated for flowering time in a 3:1 ratio as expected. The parental lines and F_2 individuals were genotyped by Digital Genotyping (DG) which identifies single nucleotide polymorphism (SNP) markers in thousands of sequenced sites that distinguish the parents of a population [176]. The near isogenic nature of the parental lines resulted in a very sparse genetic map that lacked coverage of large regions of the sorghum genome including all of the long arm of SBI02. In retrospect, no *Ma2* QTL for flowering time was identified using this genetic map because the gene is located on the long arm of SBI02 (see below).

To overcome the limitations associated with the 80M/100M population, a second mapping population was created to identify the genetic locus associated with *Ma2*. An F_2 population (n = 215) that would segregate for *Ma2* and *Ma4* was constructed by crossing Hegari (*Ma1**Ma2**Ma3**ma4**Ma5ma6*) and 80M (*Ma1**ma2**Ma3**Ma4**Ma5ma6*) [44,179]. The population was grown in a greenhouse under long day conditions and phenotyped for days to flowering. QTL for flowering time were identified on SBI02 and SBI10 (Figure 23). Recessive alleles of *Ma2* and *Ma4* result in earlier flowering when plants are grown in

long days. The Hegari haplotype across the QTL on SBI10 was associated with early flowering therefore this QTL corresponds to *Ma*$_4$ (Figure 24). The 80M haplotype across the QTL on SBI02 was associated with early flowering therefore the QTL on SBI02 corresponds to *Ma*$_2$.

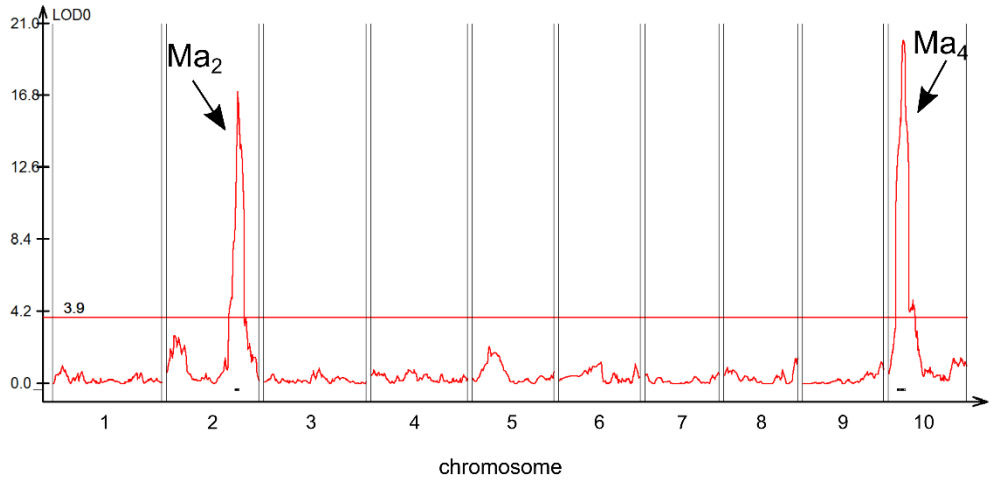

Figure 23. Quantitative trail locus (QTL) map of flowering time in the Hegari/80M F_2 population. Two QTL were identified for variation in flowering time in the F_2 population derived from Hegari (*Ma*$_1$*Ma*$_2$*Ma*$_3$*ma*$_4$) and 80M (*Ma*$_1$*ma*$_2$*Ma*$_3$*Ma*$_4$). This population was expected to segregate for *Ma*$_2$ and *Ma*$_4$. Each recessive *Ma* allele causes earlier flowering. The QTL on LG10 corresponds to *Ma*$_4$ because F_2 individuals carrying the Hegari allele contributed to accelerated flowering. F_2 individuals carrying the 80M allele at the QTL on LG02 flowered earlier, so this QTL corresponds to *Ma*$_2$.

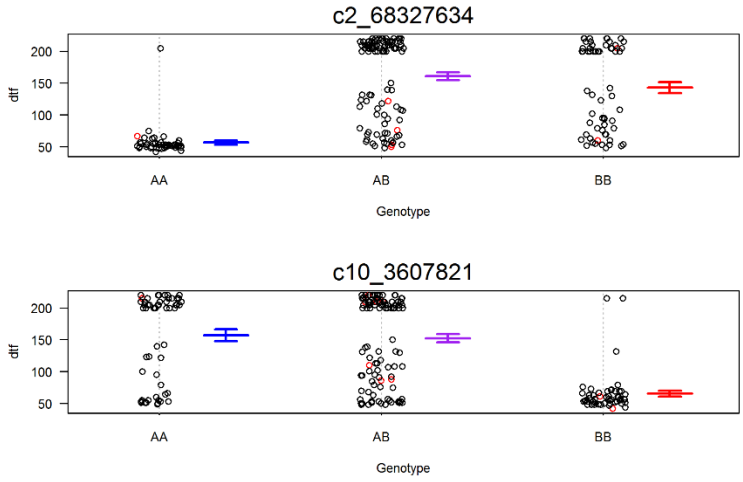

Figure 24. Genotype x phenotype plots for the QTL on SBI02 and SBI10. Recessive alleles of *Maturity* genes contribute to earlier flowering. 80M (AA) is recessive for *ma*$_2$, while Hegari (BB) is dominant. Individuals genotyped AA for the QTL on SBI02 (represented by marker c2_68327634) flowered ~100 d earlier than those genotyped BB. 80M is dominant for *Ma*$_4$, and individuals genotyped AA at the QTL on SBI10 (represented by marker c10_3607821) flowered ~100 d earlier than those genotyped BB.

Epistatic interactions between Ma$_2$ *and Ma*$_4$

Previous studies indicated an epistatic interaction exists between *Ma*$_2$ and *Ma*$_4$ [170]. Therefore, Multiple QTL Mapping (MQM) analysis [180] was employed, using data from the Hegari x 80M F_2 population, to identify additional flowering time QTL and interactions amongst the QTL as previously described [50]. MQM analysis identified the QTL for flowering time on SBI02 and SBI10 and an additional QTL on SBI09. Additionally, an epistatic interaction was identified between *Ma*$_2$ and *Ma*$_4$ (pLOD = 42). Interaction plots showed that in a dominant *Ma*$_4$ background, a dominant allele at *Ma*$_2$ delays flowering, while in a recessive *ma*$_4$ background, *Ma*$_2$ has a minimal impact on

flowering time (Figure 25). The interaction between *Ma2* and *Ma4* identified by MQM analysis is consistent previous observations that in a recessive *ma4* background flowering is early regardless of allelic variation in *Ma2* [170].

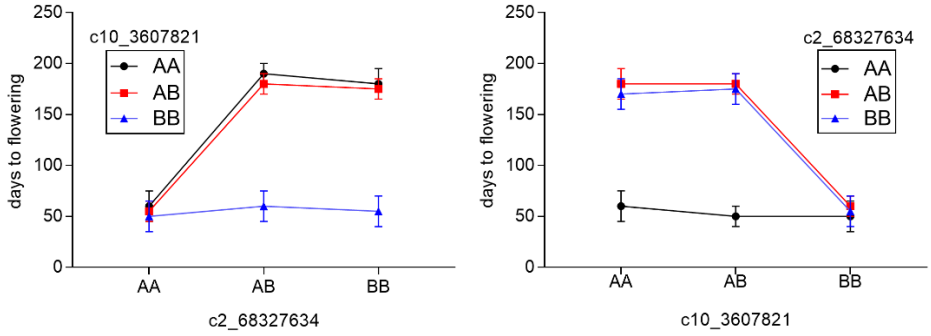

Figure 25. Interaction plots for the Ma2 QTL and the Ma4 QTL. There is a known interaction between *Ma2* (represented by marker c2_68327634) and *Ma4* (represented by marker c10_3607821). This interaction was identified by multiple QTL mapping (MQM). Dominant alleles of the *Ma* genes delay flowering. In a recessive ma4 background (AA at c2_68327634), the effect of *Ma2* on days to flowering is reduced. A represents the 80M allele and B represents the Hegari allele at each QTL. Reciprocal plots are shown.

Ma2 candidate gene identification

The Hegari*80M F_2 population located *Ma2* on SBI02 between 67.3 Mbp to 69.1 Mbp (Figure 26). To further delimit the *Ma2* locus, six lines from the Hegari x 80M population that were heterozygous across the *Ma2* QTL but fixed across the *Ma4* locus (*Ma4Ma4*) were selfed to create heterogeneous inbred families (HIFs) (n=1000 F_3 plants) [181]. Analysis of these HIFs narrowed the region encoding *Ma2* to ~600 kb (67.72 Mb-68.33 Mb) (Figure 26). Genotypes that were still heterozygous across the delimited locus

were selfed and 100 F_4 plants were evaluated for differences in flowering time. This process narrowed the *Ma*$_2$ locus to a region spanning ~500 kb containing 76 genes (67.72Mb-68.22Mb) (Figure 26, Table A-4).

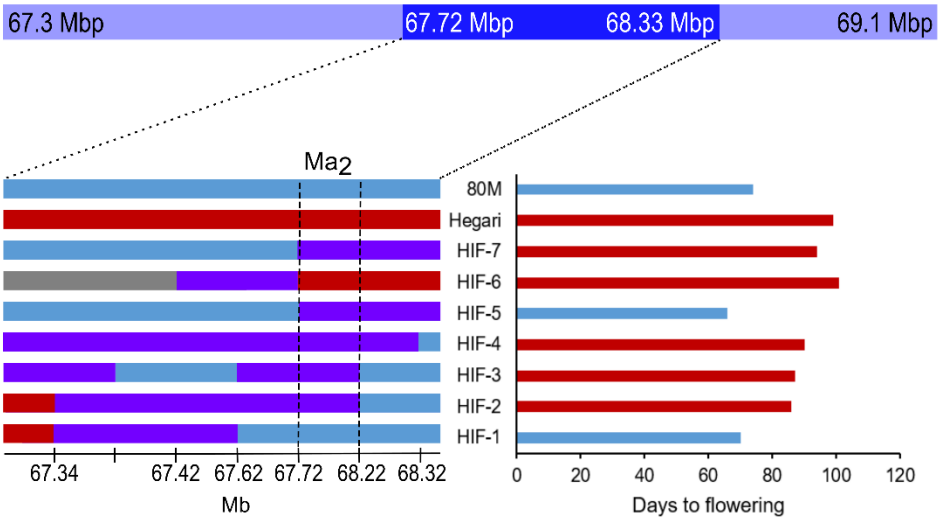

Figure 26. Fine-mapping of the Ma2 QTL. The *Ma*$_2$ QTL spans from 67.3 Mpb to 69.1 Mbp (light blue bar). Five F_2 individuals that were heterozygous across the *Ma*$_2$ QTL were self-pollinated to generate heterogeneous inbred families (HIFs) totaling 1000 F_3 individuals. Genotype and phenotype analysis of these HIFs narrowed the QTL region to ~600 kb (darker blue bar). Two additional rounds of fine-mapping narrowed the QTL region to ~500 kb (vertical dashed lines). This region contained 76 genes. The genotypes of relevant HIFs and the parents are shown to the left and their corresponding days to flowering are shown to the right. Blue regions correspond to the 80M genotype and red regions correspond to the Hegari genotype. Purple regions are heterozygous.

The low rate of recombination across the *Ma*$_2$ locus led us to utilize whole genome sequencing in conjunction with fine mapping to identify a candidate gene for *Ma*$_2$. Since 100M and 80M are near isogenic lines that have very few sequence differences along the

long arm of SBI02 where the *Ma₂* QTL is located, whole genome sequences (WGS) of 100M and 80M were generated in collaboration with JGI (sequences available at www.phytozome.jgi.doe.gov). The genome sequences were scanned for sequence differences within the 500 kb locus spanning *Ma₂*. Only one T to A single nucleotide polymorphism (SNP) located in Sobic.002G302700 was identified that distinguished 100M and 80M within the region spanning the *Ma₂* locus. The T → A mutation causes a Lys141* change in the third exon, resulting a truncated protein. A 500 bp DNA sequence spanning the T to A polymorphism in Sobic.002G302700 was sequenced from 80M and 100M to confirm the SNP identified by comparison of the whole genome sequences (Table 9). The T → A point mutation was present in 80M (*ma₂*) whereas 100M (*Ma₂*) encoded a functional version of Sobic.002G302700 that encodes a full-length protein. Since this mutation was the only sequence variant between 100M and 80M in the fine-mapped locus, Sobic.002G302700 was identified as the best candidate gene for *Ma₂*.

Sobic.002G302700 is annotated as a SET (Suppressor of variegation, Enhancer of Zeste, Trithorax) and MYND (Myeloid-Nervy-DEAF1) (SMYD) domain-containing protein. SMYD domain family proteins in humans have been found to methylate histone lysines and non-histone targets and have roles in regulating chromatin state, transcription, signal transduction, and cell cycling [182,183]. The SET domain in SMYD-containing proteins is composed of two sub-domains that are divided by the MYND zinc-finger domain. The SET domain includes conserved sequences involved in methyltransferase activity including nine cysteine residues that are present in the protein encoded by

Sobic.002G303700 (Fig 6) [184]. The MYND domain is involved in binding DNA and is enriched in cysteine and histidine residues [185]. Protein sequence alignment of Sobic.002G302700 homologs revealed that the SYMD protein candidate for Ma2 is highly conserved across flowering plants (Figure 27).

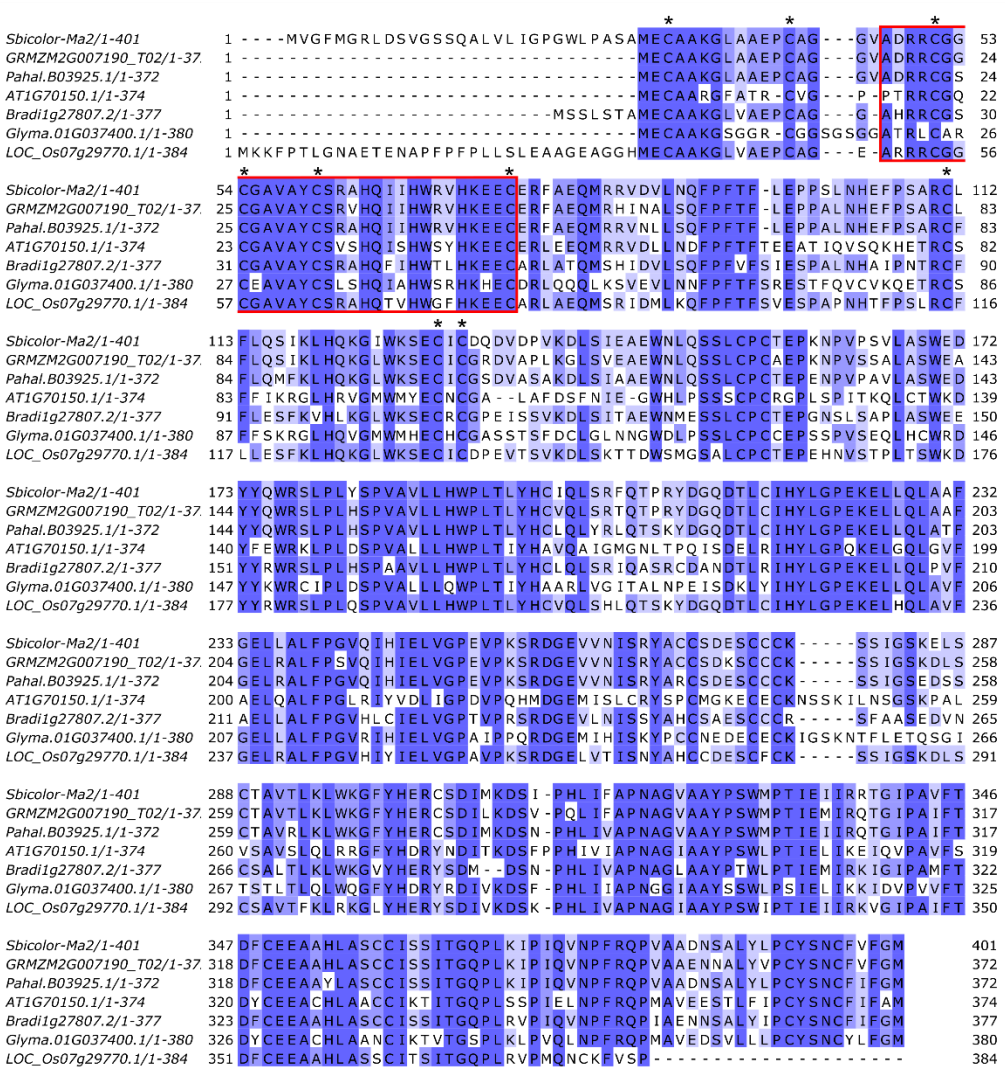

Figure 27. Alignment of Sobic.002G302700 with its closest homologs in several plant species. Sobic.002G302700 is highly conserved across plant species. It is annotated as a Set and MYND (SMYD) protein. SMYD proteins have lysine methyltransferase activity. The MYND region is highlighted in red. The nine conserved Cys residues typical of SMYD proteins are indicated by asterisks.

To learn more about *Ma2* regulation, the expression of Sobic.002G302700 in 100M and 80M was characterized during a 48h L:D/L:L cycle. *Ma*$_2$ showed a small

increase in expression from morning to evening and somewhat higher expression in 100M compared to 80M during the evening (Figure 28).

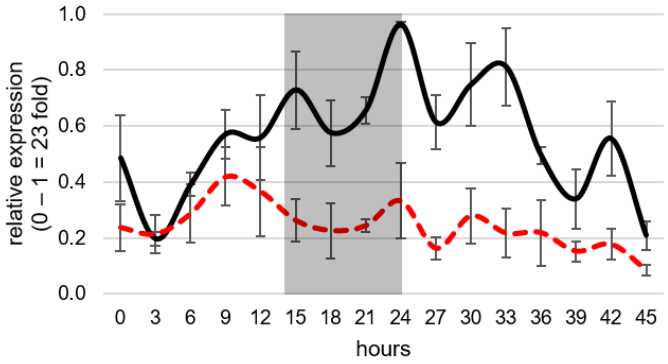

Figure 28. Circadian expression of Sobic.002G302700 in 100M and 80M. The expression of Sobic.002G302700 does not cycle diurnally in 100M (solid black line) or 80M (dashed red line). There was no difference in expression between 100M and 80M in the first day. Expression was slightly elevated in 100M compared to 80M during the night and through the following morning.

Distribution of Ma_2 alleles in the sorghum germplasm

Recessive ma_2 was originally found in the Milo background and used to construct Double Dwarf Yellow Milo ($Ma_1\mathbf{ma_2}ma_3Ma_4Ma_5ma_6$) [170]. Double Dwarf Yellow Milo was crossed to Early White Milo ($ma_1\mathbf{Ma_2}Ma_3Ma_4Ma_5ma_6$) and the progeny selected to create 100M, 80M and the other Milo maturity standards [40,49,170]. Several of the Milo maturity standards were recorded as recessive Ma_2 (80M, 60M, SM80, SM60, 44M, 38M) and others as Ma_2 dominant (100M, 90M, SM100, SM90, 52/58M). In order to confirm the Ma_2 genotype of the maturity standards, the 500 bp sequence spanning the Lys141*

mutation in Sobic.002G302700 was obtained from most of these genotypes (Table 9). Kalo was also identified as carrying a recessive allele of *Ma2*. Kalo was derived from a cross of Dwarf Yellow Milo (*ma2*), Pink Kafir (*Ma2*), and CI432 (*Ma2*), therefore it was concluded that DYM is the likely source of recessive *ma2* [170]. Sequence analysis showed that the genotypes previously identified as *ma2* including Kalo, 80M, SM80, 60M, 44M, 38M, and 58M carry the recessive mutation in Sobic.002G302700 identified in 80M. 100M, SM100, and Hegari that were identified as *Ma2* did not contain the mutated version of Sobic.002G302700 (Table 9). Additionally, sequences of *Ma2* from 52 sorghum genotypes with publicly available genome sequences were compared [28]. Sobic.002G302700 was predicted to encode functional proteins in all except one of these sorghum genotypes. A possible second recessive *Ma2* allele was found in IS3614-2 corresponding to an M83T missense mutation that was predicted to be deleterious by PROVEAN [125].

Table 9. Sequence variants of Sobic.002G203700 and their predicted effect on protein function

Genotype	Historical *Ma2* allele	Sequence variant	Effect on protein function
100M	Ma_2	-	-
SM100	Ma_2	-	-
SM90	Ma_2	-	-
Hegari	Ma_2	-	-
80M	ma_2	L141*	Deleterious
SM80	ma_2	L141*	Deleterious
60M	ma_2	L141*	Deleterious
44M	ma_2	L141*	Deleterious
38M	ma_2	L141*	Deleterious
58M	ma_2	L141*	Deleterious
Kalo	ma_2	L141*	Deleterious
IS3614-2	-	M83T	Deleterious

*Sequenced by Sanger sequencing

Discussion

Sorghum is a facultative short-day plant. In photoperiod sensitive sorghum genotypes, following the vegetative juvenile phase, day length has the greatest impact on flowering time under normal growing conditions. The development of early flowering grain sorghum adapted to temperate regions of the US was based on the selection of mutations in numerous genes that reduced photoperiod sensitivity. Genetic analysis of the loci and genes containing these mutations beginning in the 1940's [49,179] identified six *Maturity* loci (Ma_1-Ma_6) that resulted in earlier flowering time when plants were grown in long days. Recessive alleles at each of the six *Ma* loci reduces photoperiod sensitivity [44,45,49]. Molecular identification of the genes corresponding to Ma_1, Ma_3, Ma_5 and Ma_6 and other genes in the sorghum flowering time pathway (i.e., *SbCO*, *SbEHD1*, *SbCN8/12*) and an understanding of their regulation by photoperiod and the circadian

clock led to the model of the flowering time pathway shown in Figure 7 [38]. The current study showed that Ma2 represses flowering in long days by increasing the expression of SbPRR37 (*Ma$_1$*) and *SbCO*. The study also located QTL for *Ma$_2$* and *Ma$_4$*, confirmed an epistatic interaction between *Ma$_2$* and *Ma$_4$*, and identified a candidate gene for *Ma$_2$*.

The recessive *ma$_2$* allele characterized in this study arose in a highly photoperiod sensitive Milo genotype that was introduced into the US in the late 1800's and then selected for early flowering to enhance grain production. Quinby and Karper [171] created near isogenic Milo maturity genotypes with allelic variation at specific *Ma* loci to facilitate genetic and physiological analysis of flowering time regulation. In the current study, we utilized two of these maturity genotypes, 100M (*Ma$_1$**Ma$_2$**Ma$_3$Ma$_4$Ma$_5$ma$_6$*) and 80M (*Ma$_1$**ma$_2$**Ma$_3$Ma$_4$Ma$_5$ma$_6$*), to characterize how allelic variation in *Ma$_2$* affects the expression of genes in the sorghum photoperiod regulated flowering time pathway (Fig 7). This analysis showed that mutation of *ma$_2$* (80M) significantly reduced the amplitude of the morning and evening peaks of *SbPRR37* and *SbCO* expression compared to 100M (*Ma$_2$*) without altering the timing of their expression. In addition, the expression of *SbCN12* (*FT*-like) increased 8-fold in leaves of 80M compared to 100M consistent with earlier flowering in 80M. In contrast, expression of clock genes (*TOC1, LHY*) and other genes (i.e., GHD7, EHD1) in the photoperiod regulated flowering time pathway were modified to only a small extent by allelic variation in *Ma$_2$*. Based on these results, we tentatively place *Ma$_2$* in the flowering time pathway downstream of day length sensing

phytochromes and circadian clock regulation and identify *Ma*$_2$ as a factor that enhances *SbPRR37* and *SbCO* expression (Figure 29).

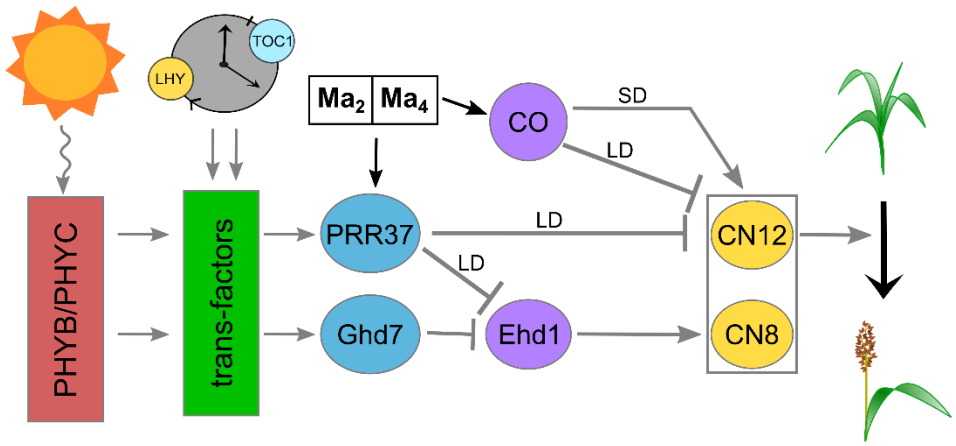

Figure 29. A model of the flowering time regulatory pathway in *S. bicolor*. *Ma*$_2$ and *Ma*$_4$ work codependently to enhance the expression of *SbPRR37* and *SbCO*. In LD, SbPRR37 and SbCO in turn repress the expression of the *SbCN* genes, especially *SbCN12*, to repress the floral transition.

The differential increase in *SbCN12* expression in 80M (vs. 100M) is consistent with differential inhibition of *SbCN12* expression in long days by the concerted action of SbPRR37 and SbCO which has been previously shown to inhibit *SbCN12* expression [169]. Prior studies showed that 100M (*Ma*$_2$) flowers later than 80M (*ma*$_2$) in long days [170]. The impact of *Ma*$_2$ alleles on the expression of *SbPRR37* and *SbCO* is consistent with the effect of these alleles on flowering time in long days. Genetic studies showed that floral repression mediated by SbPRR37 requires SbCO as a co-repressor [169]. Therefore,

enhanced expression of both *SbPRR37* (*Ma_1*) and *SbCO* by *Ma_2* in *Ma_1Ma_2* genotypes in long days is consistent with delayed flowering under these conditions relative to genotypes such as 80M that are *Ma_1ma_2*. Molecular genetic studies also showed that SbCO is an activator of *SbCN12* expression and flowering in *ma_1* genetic backgrounds [169]. This is consistent with the observation that *ma_1Ma_2* genotypes flower earlier than *ma_1ma_2* genotypes when grown in long days [170].

Interactions between Ma_2 and Ma_4

Multiple QTL (MQM) analysis of results from a population derived from Hegari/80M identified an interaction between *Ma_2* and *Ma_4* as well as one additional flowering QTL on SBI09. Flowering time QTL on SBI09 have been identified in other mapping populations, but the gene(s) involved have not been identified [61,64]. The interaction between *Ma_2* and *Ma_4* confirmed previous observations that recessive *ma_4* causes accelerated flowering in long days in *Ma_1Ma_2* genotypes [170]. Interestingly, the influence of *Ma_2* and *Ma_4* alleles on flowering time is affected by temperature [77,170]. The influence of temperature on flowering time pathway gene expression in 80M and 100M in the current study was minimized by growing plants at constant 30C. However, analysis of the temperature dependence of Ma2 and Ma4 on flowering time may help elucidate interactions between photoperiod and flowering time that have been previously documented [170,186]. Positional cloning of *Ma_4* is underway to better understand the molecular basis of *Ma_2* and *Ma_4* interaction and their impact on flowering time.

Identification of a candidate gene for Ma_2

A mapping population derived from Hegari/80M that segregated for Ma_2 and Ma_4 enabled localization of the corresponding flowering time QTL in the sorghum genome (SBI02, Ma_2; SBI10, Ma_4). The Ma_2 QTL on SBI02 was fine-mapped using heterozygous inbred families (HIFs) from Hegari/80M. Several rounds of fine-mapping delimited the QTL to a ~500kb region containing 76 genes. Low recombination rates in this region of SBI02 made it difficult to delimit the QTL further using break point analysis therefore comparison of genome sequences from 80M and 100M was used to help identify a candidate gene for *Ma2*. The recessive ma_2 allele present in 80M arose in a Milo genotype similar to 100M [170] and genetic analysis of 100M and 80M showed that these near isogenic genotypes lacked DNA markers on the long arm of SBI02 where Ma_2 is located. Indeed, a scan of the whole genome sequences of 100M and 80M identified only a single T to A mutation in the 500 kb region spanning the fine-mapped Ma_2 locus. This mutation caused a Lys141* change in the third exon of Sobic.002G302700 resulting in protein truncation. Based on this information Sobic.002G302700 was tentatively identified as the best candidate gene for Ma_2.

Sobic.002G302700 encodes a SET (Suppressor of variegation, Enhancer of Zeste, Trithorax) and MYND (Myeloid-Nervy-DEAF1) (SMYD) domain containing protein. In humans, SMYD proteins act as lysine methyltransferases, and the SET domain is critical to this activity. Therefore, Ma_2 could be altering the expression of *SbPRR37* and *SbCO* by modifying histones associated with these genes. The identification of this SMYD family

protein's involvement in flowering in sorghum as well as the identification of highly conserved homologs in other plant species suggests that *Ma2* may correspond to a novel regulator of sorghum flowering. While a role for SYMD-proteins (lysine methyltransferases) as regulators of flowering time has not been previously reported, genes encoding histone lysine demethylases (i.e., JMJ30/32) have been found to regulate temperature modulated flowering time in Arabidopsis [187].

J.R. Quinby [179] identified only one recessive allele of Ma_2 among the sorghum genotypes used in the Texas sorghum breeding program. The maturity standard lines including 80M that are recessive for ma_2 and the genotype Kalo were reported to be derived from the same recessive ma_2 Milo genotype [170]. To confirm this, Ma_2 alleles in the relevant maturity standards and Kalo were sequenced confirming that all of these ma_2 genotypes carried the same mutation identified in 80M (Table 9). Among the 52 sorghum genotypes with available whole genome sequences, only 80M carried the mutation in Ma2 [28]. One possible additional allele of ma_2 was identified in IS36214-2, which contained a M83T missense mutation that was predicted to be deleterious to protein function by PROVEAN [125].

In conclusion, we have shown that Ma_2 represses flowering in long days by promoting the expression of the floral repressor *SbPRR37* and *SbCO*, a gene that acts as a co-repressor in long days (Figure 29). Sobic.002G302700 was identified as the best candidate for the sorghum *Maturity* locus Ma_2 although further validation such as targeted mutation of Sobic.002G302700 in a Ma_1Ma_2 sorghum genotype or complementation of

Ma$_1$*ma*$_2$ genotypes will be required to confirm this gene assignment. The identification of this gene and its interaction with *Ma*$_4$ begins to elucidate a new element of the photoperiod flowering regulation pathway in sorghum.

CHAPTER V

CONCLUSIONS

The research described in thesis contribute to a greater understanding of genes affecting traits relevant to sorghum development in general as well as to the improvement of sorghum as a dedicated bioenergy crop. In the first study, variation for flowering, stem traits, and biomass were phenotyped in BTx623/IS3620C in two environments. In the controlled greenhouse environment, two new flowering time QTL were identified in this population. Flowering time and stem length regulated by *Dw2* and *Dw3* contribute to variation in biomass yield since most of the biomass QTL co-localized with QTL for flowering time and/or the *Dw* loci.

In the second study, the sorghum *D* gene, which was originally described in the early 1900s as a gene controlling white or dry leaf midribs and stems, was identified to be a NAC (NAM, ATAF1/2, CUC) family transcription factor, here referred to as *SbNAC_D*. The dry stems are due to aerenchyma formation in the stem parenchyma cells through programmed cell death (PCD). *SbNAC_D* expression in the vegetative stem was correlated with the PCD-related gene *SbXCP1*. PCD is a complex process that is triggered environmentally and developmentally by a variety of molecules such as reactive oxygen species, fatty acids, and phytohormones [110,145,146,188]. The progression of *SbNAC_D* expression from no expression in the young internodes that lack aerenchyma to high expression in the fully developed internode that are filled with aerenchyma suggests that

stem aerenchyma formation is developmentally triggered. The mechanism that induces *SbNAC_D* expression during stem development is an area for further study as well as a deeper understanding of the advantages associated with stem pith parenchyma PCD. Perhaps the plant is reducing the energy requirement to maintain live parenchyma cells in mature internodes that are mainly structural or recycling carbohydrates and nitrogen to support further growth. The nutrient recycling hypothesis is supported by observations of maize roots where aerenchyma formation enhanced growth under phosphorus, nitrogen, and potassium limitation [189].

In the third study, the sorghum flowering time gene *Maturity2* (*Ma_2*) was isolated from a QTL identified in Hegari/80M through fine-mapping and analysis of whole genome sequences. The best candidate gene for *Ma_2* was a lysine methyltransferase, which had a stop codon mutation in 80M, the earlier flowering genotype. This mutation is also present in several genotypes that were previously established to be recessive for *Ma_2*. Circadian expression studies of genotypes that differed at *Ma_2* showed that *Ma_2* enhances the expression of *SbPRR37* and *SbCO*, which are repressors of flowering time in long days. Further study is required to understand how *Ma_2* affects the expression of *SbPRR37* and *SbCO*. Lysine methyltransferases affect chromatin structure, so it is possible that *Ma_2* affects DNA accessibility around the promoters of *SbPRR37* and *SbCO*. *Ma_2* also epistatically interacts with *Ma_4*, so the identification of *Ma_4* could reveal a unique molecular mechanism for the regulation of flowering by these two genes.

Overall, the identification of these genes is an important step toward understanding the genetics controlling biomass accumulation, internode growth, and flowering time in sorghum. The identification of genes that contribute to bioenergy traits of sorghum may provide points of eventual transgenic manipulation to optimize sorghum as a bioenergy crop. Direct manipulation of one or more genes through transgenic approaches may offer an option for further gains in yield. These options are important to explore as the global human population grows and demand for food, fiber, and fuel increases.

REFERENCES

1. Tilman D, Balzer C, Hill J, Befort BL. Global food demand and the sustainable intensification of agriculture. Proc Natl Acad Sci. 2011;108(50):20260–4.

2. Duvick DN. Genetic Progress in Yield of United States Maize (Zea mays L.). Maydica. 2005;50:193–205.

3. Evenson RE, Gollin D. Assessing the Impact of the Green Revolution, 1960 to 2000. Science (80-). 2003;300(5620):758–62.

4. Slewinski TL. Non-structural carbohydrate partitioning in grass stems: a target to increase yield stability, stress tolerance, and biofuel production. J Exp Bot. 2012;63(13):4647–70.

5. Grassini P, Eskridge KM, Cassman KG. Distinguishing between yield advances and yield plateaus in historical crop production trends. Nat Commun. 2013;4:2918.

6. Foley JA, DeFries R, Asner GP, Barford C, Bonan G, Carpenter SR, et al. Global Consequences of Land Use. Science (80-). 2005;309(5734):570–4.

7. Höök M, Xu T. Depletion of fossil fuels and anthropogenic climate change – a review. Energy Policy. 2013;52:797–809.

8. Wheeler T, von Braun J. Climate Change Impacts on Global Food Security. Science (80-). 2013;341(6145):508–13.

9. NOAA Earth System Research Laboratory, Global Monitoring Division. NOAA Research. 2019.

10. Jemma G, Richard B, Eleanor B, Robin C, Joanne C, Kate W, et al. Implications of climate change for agricultural productivity in the early twenty-first century. Philos Trans R Soc B Biol Sci. 2010;365(1554):2973–89.

11. Lobell DB, Schlenker W, Costa-Roberts J. Climate Trends and Global Crop Production Since 1980. Science (80-). 2011;333:616–20.

12. Holou RAY, Kindomihou VM. The Biofuel Crops in Global Warming Challenge: Carbon Caputure by Corn, Sweet Sorghum, and Switchgrass Biomass Grown for Biofuel Production in the USA. In: Jacob-Lopes E, Zepka LQ, editors. Frontiers in Bioenergy and Biofuels. INTECH; 2017. p. 139–51.

13. Schmer MR, Vogel KP, Varvel GE, Follett RF, Mitchell RB, Jin VL. Energy Potential and Greenhouse Gas Emissions from Bioenergy Cropping Systems on Marginally Productive Cropland. PLoS One. 2014;9(3):e89501.

14. Hoekman SK, Broch A. Environmental implications of higher ethanol production and use in the U.S.: A literature review. Part II – Biodiversity, land use change, GHG emissions, and sustainability. Renew Sustain Energy Rev. 2018;81:3159–77.

15. Cui X, Kavvada O, Huntington T, Scown CD. Strategies for near-term scale-up of cellulosic biofuel production using sorghum and crop residues in the U.S. Environ Res Lett. 2018;13:124002.

16. Wu X, Staggenborg S, Propheter JL, Rooney WL, Yu J, Wang D. Features of sweet sorghum juice and their performance in ethanol fermentation. Ind Crops Prod. 2010;31:164–70.

17. Menetrez MY. An Overview of Algae Biofuel Production and Potential Environmental Impact. Environ Sci Technol. 2012;46(13):7073–85.

18. Energy USD of, Perlack RD, Eaton LM, Turhollow AFJ, Langholtz MH, Brandt CC, et al. U.S. Billion-ton Update: Biomass Supply for a Bioenergy and Bioproducts Industry. Agric Biosyst Eng Tech Reports White Pap. 2011;16.

19. Houghton J, Weatherwax S, Ferrell J. Breaking the Biological Barriers to Cellulosic Ethanol: A Joint Research Agenda. In: A Research Roadmap Resulting from the Biomass to Biofuels Workshop Sponsored by the US Department of Energy. Rockville, MD: U.S. Department of Energy; 2006. p. 39–83.

20. Wyman CE. What is (and is not) vital to advancing cellulosic ethanol. Trends Biotechnol. 2007;25(4):153–7.

21. Barrière Y, Cédric Riboulet B, Lapierre C, Thomas Lübberstedt B, Jean-Pierre Martinant B. Genetics and Genomics of Lignification in Grass Cell Walls Based on Maize as Model Species. Genes, Genomes and Genomics. 2007;1(2):133–56.

22. Dien BS, Sarath G, Pedersen JF, Sattler SE, Chen H, Funnell-Harris DL, et al. Improved Sugar Conversion and Ethanol Yield for Forage Sorghum (Sorghum bicolor L. Moench) Lines with Reduced Lignin Contents. BioEnergy Res. 2009;2(3):153–64.

23. Van Acker R, Leplé J-C, Aerts D, Storme V, Goeminne G, Ivens B, et al. Improved saccharification and ethanol yield from field-grown transgenic poplar deficient in cinnamoyl-CoA reductase. Proc Natl Acad Sci. 2014 Jan 14;111(2):845 LP-850.

24. Mottiar Y, Vanholme R, Boerjan W, Ralph J, Mansfield SD. Designer lignins: harnessing the plasticity of lignification. Curr Opin Biotechnol. 2016;37:190–200.

25. Rooney WL. Sorghum. In: Cellulosic Energy Cropping Systems. 2014. p. 109–29.

26. FAOSTAT [Internet]. Available from: www.fao.org/faostat

27. Kimber CT, Dahlberg JA, Kresovich S. The Gene Pool of Sorghum bicolor and Its Improvement. In: Paterson AH, editor. Genomics of the Saccharinae. 1st ed. New York: Springer-Verlag; 2013. p. 23–41.

28. McCormick RF, Truong SK, Sreedasyam A, Jenkins J, Shu S, Sims D, et al. The Sorghum bicolor reference genome: improved assembly, gene annotations, a transcriptome atlas, and signatures of genome organization. Plant J. 2018;93(2):338–54.

29. Childs KL, Miller FR, Cordonnier-Pratt MM, Pratt LH, Morgan PW, Mullet JE. The sorghum photoperiod sensitivity gene, Ma3, encodes a phytochrome B. Plant Physiol. 1997;113:611–9.

30. Murphy RL, Klein RR, Morishige DT, Brady JA, Rooney WL, Miller FR, et al. Coincident light and clock regulation of controls photoperiodic flowering in sorghum. Proc Natl Acad Sci U S A. 2011;37:1–6.

31. Murphy RL, Morishige DT, Brady J a., Rooney WL, Yang S, Klein PE, et al. Ghd7 (Ma6) Represses Sorghum Flowering in Long Days: Alleles Enhance Biomass Accumulation and Grain Production. Plant Genome. 2014;7(2):1–10.

32. Hilley JL, Weers BD, Truong SK, McCormick RF, Mattison AJ, McKinley B, et al. Sorghum Dw2 encodes a protein kinase regulator of stem internode length. Sci

Rep. 2017;7:1–13.

33. Hilley J, Truong S, Olson S, Morishige D, Mullet J. Identification of Dw1 , a Regulator of Sorghum Stem Internode Length. PLoS One. 2016;11(3):1–16.

34. Multani DS, Briggs SP, Chamberlin MA, Blakeslee JJ, Murphy AS, Johal GS. Loss of an MDR transporter in compact stalks of maize br2 and sorghum dw3 mutants. Science (80-). 2003;302(5642):81–4.

35. Casto AL, McKinley BA, Yu KMJ, Rooney WL, Mullet JE. Sorghum stem aerenchyma formation is regulated by SbNAC_D during internode development. Plant Direct. 2018;2(11):e00085.

36. Wu E, Lenderts B, Glassman K, Berezowska-Kaniewska M, Christensen H, Asmus T, et al. Optimized Agrobacterium-mediated sorghum transformation protocol and molecular data of transgenic sorghum plants. Vitr Cell Dev Biol Plant. 2014;50(1):9–18.

37. Mathur S, Umakanth A V., Tonapi VA, Sharma R, Sharma MK. Sweet sorghum as biofuel feedstock: Recent advances and available resources. Biotechnol Biofuels. 2017;10:1–19.

38. Mullet JE. High-Biomass C4 Grasses − Filling the Yield Gap. Plant Sci. 2017;261:10–7.

39. Olson SN, Ritter K, Rooney WL, Kemanian A, McCarl BA, Zhang Y, et al. High biomass yield energy sorghum: developing a genetic model for C4 grass bioenergy crops. Biofuels, Bioprod Biorefining. 2012;6(3):246–56.

40. Quinby JR. The Genetic Control of Flowering and Growth in Sorghum. Adv Agron. 1974;25:125–62.

41. Carvalho G, Rooney WL. Assessment of Stalk Properties to Predict Juice Yield in Sorghum. BioEnergy Res. 2017;10(3):657–70.

42. McKinley B, Rooney W, Wilkerson C, Mullet J. Dynamics of biomass partitioning, stem gene expression, cell wall biosynthesis, and sucrose accumulation during development of Sorghum bicolor. Plant J. 2016;88(4):662–80.

43. Morgan PW, Finlayson SA. Physiology and Genetics of Maturity and Height. In: Smith CW, Rederiksen RA, editors. Sorghum: Origin, Technology, and Production. New York: Wiley & Sons; 2000. p. 227–59.

44. Quinby JR. Fourth maturity gene locus in sorghum. Crop Sci. 1966;(6):516–8.

45. Rooney WL, Aydin S. Genetic Control of a Photoperiod-Sensitive Response in Sorghum bicolor (L.) Moench. Crop Sci. 1999;39:397–400.

46. Murphy R, Morishige D, Brady J, Rooney W, Yang S, Klein P, et al. Ghd7 (Ma6) Represses Flowering in Long Days: A Key Trait in Energy Sorghum Hybrids. Vol. 7. 2013. 0-34 p.

47. Hmon KPW, Shehzad T, Okuno K. Variation in inflorescence architecture associated with yield components in a sorghum germplasm. Plant Genet Resour. 2013/06/11. Cambridge University Press; 2013;11(3):258–65.

48. Brown PJ, Klein PE, Bortiri E, Acharya CB, Rooney WL, Kresovich S. Inheritance of inflorescence architecture in sorghum. Theor Appl Genet. 2006;113(5):931–42.

49. Quinby JR, Karper RE. Effect of Different Alleles on the Growth of Sorghum Hybrids. 1948;255–9.

50. Truong SK, McCormick RF, Rooney WL, Mullet JE. Harnessing genetic variation in leaf angle to increase productivity of sorghum bicolor. Genetics. 2015;201(3):1229–38.

51. Fernandez MGS, Hamblin MT, Li L, Rooney WL, Tuinstra MR, Kresovich S. Quantitative Trait Loci Analysis of Endosperm Color and Carotenoid Content in Sorghum Grain. Crop Sci. Madison, WI: Crop Science Society of America; 2008;48:1732–43.

52. Upadhyaya HD, Wang Y-H, Sharma S, Singh S, Hasenstein KH. SSR markers linked to kernel weight and tiller number in sorghum identified by association mapping. Euphytica. 2012;187(3):401–10.

53. Bernardo R. Breeding for Quantitative Traits in Plants. Woodbury: Stemma Press; 2010. 85-112 p.

54. Beckmann JS, Soller M. Restriction fragment length polymorphisms and genetic improvement of agricultural species. Euphytica. 1986;35(1):111–24.

55. Welsh J, McClelland M. Fingerprinting genomes using PCR with arbitrary primers. Nucleic Acids Res. 1990 Dec 25;18(24):7213–8.

56. Tautz D. Hypervariability of simple sequences as a general source for polymorphic DNA markers. Nucleic Acids Res. 1989 Aug 25;17(16):6463–71.

57. Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Hornes M, et al. AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res. 1995 Nov 11;23(21):4407–14.

58. Morishige DT, Klein PE, Hilley JL, Sahraeian SME, Sharma A, Mullet JE, et al. Digital genotyping of sorghum - a diverse plant species with a large repeat-rich genome. BMC Genomics. 2013;14(448):1–19.

59. Klein RR, Rodriguez-Herrera R, Schlueter JA, Klein PE, Yu ZH, Rooney WL. Identification of genomic regions that affect grain-mould incidence and other traits of agronomic importance in sorghum. Theor Appl Genet. 2001;102:307–19.

60. Murray SC, Sharma A, Rooney WL, Klein PE, Mullet JE, Mitchell SE, et al. Genetic Improvement of Sorghum as a Biofuel Feedstock: I. QTL for Stem Sugar and Grain Nonstructural Carbohydrates. Crop Sci. 2008;48(6):2165–79.

61. Zhao J, Mantilla Perez MB, Hu J, Salas Fernandez MG. Genome-Wide Association Study for Nine Plant Architecture Traits in Sorghum. Plant Genome. 2016;9(2):1–14.

62. Xu W, Subudhi PK, Crasta OR, Rosenow DT, Mullet JE, Nguyen HT. Molecular mapping of QTLs conferring stay-green in grain sorghum (Sorghum bicolor L. Moench). Genome. 2000;43:461–9.

63. Morris GP, Ramu P, Deshpande SP, Hash CT, Shah T, Upadhyaya HD, et al. Population genomic and genome-wide association studies of agroclimatic traits in sorghum. Proc Natl Acad Sci. National Academy of Sciences; 2013 Jan 8;110(2):453–8.

64. Higgins RH, Thurber CS, Assaranurak I, Brown PJ. Multiparental mapping of plant

height and flowering time QTL in partially isogenic sorghum families. G3. 2014;4:1593–602.

65. Brown PJ, Rooney WL, Franks C, Kresovich S. Efficient Mapping of Plant Height Quantitative Trait Loci in a Sorghum Association Population With Introgressed Dwarfing Genes. Genetics. 2008;180(1):629–37.

66. Hart GE, Schertz KF, Peng Y, Syed NH. Genetic mapping of Sorghum bicolor (L.) Moench QTLs that control variation in tillering and other morphological characters. Theor Appl Genet. 2001;103(8):1232–42.

67. Burow GB, Klein RR, Franks CD, Klein PE, Schertz KF, Pederson GA, et al. Registration of the BTx623/IS3620C Recombinant Inbred Mapping Population of Sorghum. J Plant Regist. 2011;5(1):141.

68. Feltus FA, Hart GE, Schertz KF, Casa AM, Kresovich S, Abraham S, et al. Alignment of genetic maps and QTLs between inter- and intra-specific sorghum populations. Theor Appl Genet. 2006;112(7):1295–305.

69. Xu G-W, Magill CW, Schertz KF, Hart GE. A RFLP linkage map of Sorghum bicolor (L.) Moench. Theor Appl Genet. Springer; 1994;89(2–3):139–45.

70. Menz MA, Klein RR, Mullet JE, Obert JA, Unruh NC, Klein PE. A high-density genetic map of Sorghum bicolor (L.) Moench based on 2926 AFLP®, RFLP and SSR markers. Plant Mol Biol. Springer; 2002;48(5–6):483–99.

71. Truong SK, McCormick RF, Morishige DT, Mullet JE. Resolution of genetic map expansion caused by excess heterozygosity in plant recombinant inbred populations. G3. 2014 Oct;4(10):1963–9.

72. Stephens JC, Miller FR, Rosenow DT. Conversion of alien sorghums to early combine genotypes. Crop Sci. 1967;7(4):396.

73. Tsuji H, Taoka KI, Shimamoto K. Florigen in rice: Complex gene network for florigen transcription, florigen activation complex, and multiple functions. Curr Opin Plant Biol. Elsevier Ltd; 2013;16:228–35.

74. Wigge PA. Ambient temperature signalling in plants. Curr Opin Plant Biol. Elsevier Ltd; 2013;16(5):661–6.

75. Pearce S, Vanzetti LS, Dubcovsky J. Exogenous Gibberellins Induce Wheat Spike Development under Short Days Only in the Presence of VERNALIZATION1. Plant Physiol. 2013;163(3):1433–45.

76. Mullet J, Morishige D, McCormick R, Truong S, Hilley J, McKinley B, et al. Energy sorghum--a genetic model for the design of C4 grass bioenergy crops. J Exp Bot. 2014;65(13):3479–89.

77. Major DJ, Rood SB, Miller FR. Temperature and Photoperiod Effects Mediated by the Sorghum Maturity Genes. Crop Sci. 1990;30(2):305–10.

78. Quinby JR. Sorghum Improvement and the Genetics of Growth. Texas Agricultural Experiment Station; 1974.

79. Broman KW, Wu H, Sen Ś, Churchill GA. R/qtl: QTL mapping in experimental crosses. Bioinformatics. 2003;19(7):889–90.

80. Broman K, Sen S. A Guide to QTL Mapping with R/qtl. Springer-Verlag New York; 2009.

81. George-Jaeggli B, Jordan DR, van Oosterom EJ, Broad IJ, Hammer GL. Sorghum dwarfing genes can affect radiation capture and radiation use efficiency. F Crop Res. 2013;149:283–90.

82. Kebrom TH. A Growing Stem Inhibits Bud Outgrowth – The Overlooked Theory of Apical Dominance. 2017;8(October):1–7.

83. Quinby JR. The genetics of sorghum improvement. J Hered. 1975;66:56–62.

84. Yamaguchi M, Fujimoto H, Hirano K, Araki-Nakamura S, Ohmae-Shinohara K, Fujii A, et al. Sorghum Dw1, an agronomically important gene for lodging resistance, encodes a novel protein involved in cell proliferation. Sci Rep. 2016;6(February):28366.

85. Hirano K, Kawamura M, Araki-Nakamura S, Fujimoto H, Ohmae-Shinohara K, Yamaguchi M, et al. Sorghum DW1 positively regulates brassinosteroid signaling by inhibiting the nuclear localization of BRASSINOSTEROID INSENSITIVE 2. Sci Rep. Springer US; 2017;7(126):1–10.

86. Lingle SE. Sucrose Metabolism in the Primary Culm of Sweet Sorghum during Development. Crop Sci. 1987;27(6):1214–9.

87. Gutjahr S, Vaksmann M, Dingkuhn M, Thera K, Trouche G, Braconnier S, et al. Grain, sugar and biomass accumulation in tropical sorghums. I. Trade-offs and effects of phenological plasticity. Funct Plant Biol. 2013;40(4):342–54.

88. Godoy JG V, Tesso TT. Analysis of Juice Yield, Sugar Content, and Biomass Accumulation in Sorghum. Crop Sci. 2013;53(4):1288–97.

89. Bihmidine S, Hunter CT, Johns CE, Koch KE, Braun DM. Regulation of assimilate import into sink organs: update on molecular drivers of sink strength. Front Plant Sci. 2013;4(177):1–15.

90. Burks PS, Kaiser CM, Hawkins EM, Brown PJ. Genomewide Association for Sugar Yield in Sweet Sorghum. Crop Sci. 2015;55(5):2138–48.

91. Han Y, Lv P, Hou S, Li S, Ji G, Ma X, et al. Combining next generation sequencing with bulked segregant analysis to fine map a stem moisture locus in sorghum (Sorghum bicolor L. Moench). PLoS One. 2015;10(5):1–14.

92. Felderhoff TJ, Murray SC, Klein PE, Sharma A, Hamblin MT, Kresovich S, et al. QTLs for Energy-related Traits in a Sweet x Grain Sorghum [Sorghum bicolor (L.) Moench] Mapping Population. Crop Sci. 2012;52(5):2040–9.

93. Xiao-ping L, Jin-feng Y, Cui-ping G, Acharya S. Quantitative trait loci analysis of economically important traits in Sorghum bicolor × S. sudanense hybrid. Can J Plant Sci. Canadian Science Publishing; 2011;91(1):81–90.

94. Guan Y, Wang H, Qin L, Zhang H, Yang Y, Gao F, et al. QTL mapping of bio-energy related traits in Sorghum. Euphytica. 2011;182(3):431.

95. Shiringani AL, Frisch M, Friedt W. Genetic mapping of QTLs for sugar-related traits in a RIL population of Sorghum bicolor L. Moench. Theor Appl Genet. 2010;121(2):323–36.

96. Srinivas G, Satish K, Madhusudhana R, Reddy RN, Mohan SM, Seetharama N, et al. Identification of quantitative trait loci for agronomically important traits and their association with genic-microsatellite markers in sorghum. Theor Appl Genet.

2009;118(8):1439–54.

97. Mocoeur A, Zhang YM, Liu ZQ, Shen X, Zhang LM, Rasmussen SK, et al. Stability and genetic control of morphological, biomass and biofuel traits under temperate maritime and continental conditions in sweet sorghum (Sorghum bicolour). Theor Appl Genet. 2015;128(9):1685–701.

98. Calviño M, Messing J. Sweet sorghum as a model system for bioenergy crops. Curr Opin Biotechnol. 2012;23(3):323–9.

99. Milne RJ, Reinders A, Ward JM, Offler CE, Patrick JW, Grof CPL. Contribution of sucrose transporters to phloem unloading within Sorghum bicolor stem internodes. Plant Signal Behav. Taylor & Francis; 2017;12(5):e1319030.

100. Hilson GR. A note on the inheritance of certain stem characters in Sorghum. Agr J India. 1916;11:150–5.

101. Stephens JCC, Quinby JRR. The D Rs P linkage group in sorghum. J Agric Res. 1939;59(10):725–30.

102. Ayyangar GNR, Ayyare MAS, Rao VP, Nambiar AK. Mendilian segregations for juciness and sweetness in sorghum stalks. Madras Agric J. 1936;24:247–8.

103. Swanson AF, Parker JH. Inheritance of Smut Resistance and Juiciness of Stalk in the Sorghum Cross, Red Amber X Feterita. J Hered. 1931;22(2):51–6.

104. Xu W, Subudhi PK, Crasta OR, Rosenow DT, Mullet JE, Nguyen HT. Molecular mapping of QTLs conferring stay-green in grain sorghum (Sorghum bicolor L. Moench). Genome. 2000;43(3):461–9.

105. Drew MC, Jackson MB, Giffard S. Ethylene-promoted adventitious rooting and development of cortical air spaces (aerenchyma) in roots may be adaptive responses to flooding in Zea mays L. Planta. 1979;147(1):83–8.

106. Drew MC, He CJ, Morgan PW. Programmed cell death and aerenchyma formation in roots. Trends Plant Sci. 2000;5(3):123–7.

107. Yamauchi T, Shimamura S, Nakazono M, Mochizuki T. Aerenchyma formation in

crop species: A review. F Crop Res. 2013;152:8–16.

108. Yamauchi T, Tanaka A, Mori H, Takamure I, Kato K, Nakazono M. Ethylene-dependent aerenchyma formation in adventitious roots is regulated differently in rice and maize. Plant Cell Env. 2016;39(10):2145–57.

109. Visser EJ, Bogemann GM. Aerenchyma formation in the wetland plant Juncus effusus is independent of ethylene. New Phytol. 2006;171(2):305–14.

110. Liang H, Yao N, Song JT, Luo S, Lu H, Greenberg JT, et al. Ceramides modulate programmed cell death in plants. Genes Dev. 2003;17(21):2636–41.

111. Fendrych M, Van Hautegem T, Van Durme M, Olvera-Carrillo Y, Huysmans M, Karimi M, et al. Programmed cell death controlled by ANAC033/SOMBRERO determines root cap organ size in Arabidopsis. Curr Biol. 2014;24(9):931–40.

112. Pesquet E, Zhang B, Gorzsas A, Puhakainen T, Serk H, Escamez S, et al. Non-cell-autonomous postmortem lignification of tracheary elements in Zinnia elegans. Plant Cell. 2013;25(4):1314–28.

113. Uauy C, Distelfeld A, Fahima T, Blechl A, Dubcovsky J. A NAC Gene regulating senescence improves grain protein, zinc, and iron content in wheat. Science (80-). 2006;314(5803):1298–301.

114. Young TE, Gallie DR. Regulation of programmed cell death in maize endosperm by abscisic acid. Plant Mol Biol. 2000;42(2):397–414.

115. Mergemann H, Sauter M. Ethylene induces epidermal cell death at the site of adventitious root emergence in rice. Plant Physiol. 2000;124(2):609–14.

116. Kim HJ, Hong SH, Kim YW, Lee IH, Jun JH, Phee BK, et al. Gene regulatory cascade of senescence-associated NAC transcription factors activated by ETHYLENE-INSENSITIVE2-mediated leaf senescence signalling in Arabidopsis. J Exp Bot. 2014;65(14):4023–36.

117. De Clercq I, Vermeirssen V, Van Aken O, Vandepoele K, Murcha MW, Law SR, et al. The membrane-bound NAC transcription factor ANAC013 functions in mitochondrial retrograde regulation of the oxidative stress response in Arabidopsis. Plant Cell. 2013;25(9):3472–90.

118. Yamauchi T, Yoshioka M, Fukazawa A, Mori H, Nishizawa NK, Tsutsumi N, et al. An NADPH Oxidase RBOH Functions in Rice Roots during Lysigenous Aerenchyma Formation under Oxygen-Deficient Conditions. Plant Cell. 2017;29(4):775–90.

119. Rajhi I, Yamauchi T, Takahashi H, Nishiuchi S, Shiono K, Watanabe R, et al. Identification of genes expressed in maize root cortical cells during lysigenous aerenchyma formation using laser microdissection and microarray analyses. New Phytol. 2011;190(2):351–68.

120. Takahashi H, Yamauchi T, Rajhi I, Nishizawa NK, Nakazono M. Transcript profiles in cortical cells of maize primary root during ethylene-induced lysigenous aerenchyma formation under aerobic conditions. Ann Bot. 2015;115(6):879–94.

121. Nilsen KT, N'Diaye A, MacLachlan PR, Clarke JM, Ruan Y, Cuthbert RD, et al. High density mapping and haplotype analysis of the major stem-solidness locus SSt1 in durum and common wheat. PLoS One. 2017;12(4):e0175285.

122. Gomez FE, Carvalho G, Shi F, Muliana AH, Rooney WL. High throughput phenotyping of morpho-anatomical stem properties using X-ray computed tomography in sorghum. Plant Methods. 2018;14(1):59.

123. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics. 2009;25(14):1754–60.

124. McCormick RF, Truong SK, Mullet JE. RIG: Recalibration and Interrelation of Genomic Sequence Data with the GATK. G3. 2015;5(4):655–65.

125. Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. Predicting the Functional Effect of Amino Acid Substitutions and Indels. PLoS One. 2012;7(10):e46688.

126. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32(5):1792–7.

127. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. Bioinformatics. 2009;25(9):1189–91.

128. Jones DT, Taylor WR, Thornton JM. The rapid generation of mutation data

matrices from protein sequences. Bioinformatics. 1992;8(3):275–82.

129. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol Biol Evol. 2016;33(7):1870–4.

130. Pfaffl MW. A new mathematical model for relative quantification in real-time RT–PCR. Nucleic Acids Res. Oxford, UK: Oxford University Press; 2001;29(9):e45–e45.

131. Zhu T, Nevo E, Sun D, Peng J. Phylogenetic analyses unravel the evolutionary history of NAC proteins in plants. Evolution (N Y). 2012;66(6):1833–48.

132. Souer E, van Houwelingen A, Kloos D, Mol J, Koes R. The No Apical Meristem Gene of Petunia Is Required for Pattern Formation in Embryos and Flowers and Is Expressed at Meristem and Primordia Boundaries. Cell. 1996;85(2):159–70.

133. Kadier Y, Zu YY, Dai QM, Song G, Lin SW, Sun QP, et al. Genome-wide identification, classification and expression analysis of NAC family of genes in sorghum [Sorghum bicolor (L.) Moench]. Plant Growth Regul. Springer Netherlands; 2017;83(2):1–12.

134. Zhao C, Craig JC, Petzold HE, Dickerman AW, Beers EP. The xylem and phloem transcriptomes from secondary tissues of the Arabidopsis root-hypocotyl. Plant Physiol. 2005;138(2):803–18.

135. Nakano Y, Yamaguchi M, Endo H, Rejab NA, Ohtani M. NAC-MYB-based transcriptional regulation of secondary cell wall biosynthesis in land plants. Front Plant Sci. 2015;6:288.

136. Peng X, Zhao Y, Li X, Wu M, Chai W, Sheng L, et al. Genomewide identification, classification and analysis of NAC type gene family in maize. J Genet. 2015;94(3):377–90.

137. Yin L-L, Xue H-W. The MADS29 Transcription Factor Regulates the Degradation of the Nucellus and the Nucellar Projection during Rice Seed Development. Plant Cell. 2012;24(3):1049–65.

138. Olvera-Carrillo Y, Van Bel M, Van Hautegem T, Fendrych M, Huysmans M, Simaskova M, et al. A conserved core of PCD indicator genes discriminates

developmentally and environmentally induced programmed cell death in plants. Plant Physiol. 2015;169(4):2684–99.

139. Klein RR, Klein PE, Mullet JE, Minx P, Rooney WL, Schertz KF. Fertility restorer locus Rf1 of sorghum (Sorghum bicolor L.) encodes a pentatricopeptide repeat protein not present in the colinear region of rice chromosome 12. Theor Appl Genet. 2005;111(6):994–1012.

140. Noh B, Murphy AS, Spalding EP. Multidrug Resistance-like Genes of Arabidopsis Required for Auxin Transport and Auxin-Mediated Development. Plant Cell. 2001;13:2441–54.

141. Lin Z, Li X, Shannon LM, Yeh CT, Wang ML, Bai G, et al. Parallel domestication of the Shattering1 genes in cereals. Nat Genet. 2012;44(6):720–4.

142. Escamez S, Tuominen H. Programmes of cell death and autolysis in tracheary elements: when a suicidal cell arranges its own corpse removal. J Exp Bot. 2014;65(5):1313–21.

143. Kebrom TH, Mckinley B, Mullet JE. Dynamics of gene expression during development and expansion of vegetative stem internodes of bioenergy sorghum. Biotechnol Biofuels. BioMed Central; 2017;10:1–16.

144. Steffens B, Geske T, Sauter M. Aerenchyma formation in the rice stem and its promotion by H2O2. New Phytol. 2011;190:369–78.

145. Kabbage M, Kessens R, Bartholomay LC, Williams B. The Life and Death of a Plant Cell. Annu Rev Plant Biol Vol 68. 2017;68:375–404.

146. Van Hautegem T, Waters AJ, Goodrich J, Nowack MK. Only in dying, life: programmed cell death during plant development. Trends Plant Sci. 2015;20(2):102–13.

147. Smith CW, Frederiksen RA. History of cultivar development in the United States: From "Memoirs of A. B. Maunder - Sorghum Breeder." In: Smith CW, editor. Sorghum: Origin, History Technology, and Production. John Wiley & Sons, Inc.; 2000. p. 191–223.

148. Waclawovsky AJ, Sato PM, Lembke CG, Moore PH, Souza GM. Sugarcane for

bioenergy production: an assessment of yield and regulation of sucrose content. Plant Biotechnol J. 2010;8(3):263–76.

149. Mace ES, Tai S, Gilding EK, Li Y, Prentis PJ, Bian L, et al. Whole-genome sequencing reveals untapped genetic potential in Africa's indigenous cereal crop sorghum. Nat Commun. 2013;4:337–42.

150. Boyles RE, Brenton ZW, Kresovich S. Genetic and Genomic Resources of Sorghum to Connect Genotype with Phenotype in Contrasting Environments. Plant J. 2018;19–39.

151. Burks PS, Felderhoff TJ, Viator HP, Rooney WL. The Influence of Hybrid Maturity and Planting Date on Sweet Sorghum Productivity during a Harvest Season. Agron J. 2013;105:263–7.

152. Teetor VH, Duclos D V., Wittenberg ET, Young KM, Chawhuaymak J, Riley MR, et al. Effects of planting date on sugar and ethanol yield of sweet sorghum grown in Arizona. Ind Crops Prod. 2011;34:1293–300.

153. Rooney WL, Blumenthal J, Bean B, Mullet JE. Designing sorghum as a dedicated bioenergy feedstock. Biofuels, Bioprod Biorefining. 2007;1(2):147–57.

154. Song YH, Ito S, Imaizumi T. Flowering time regulation: Photoperiod- and temperature-sensing in leaves. Trends Plant Sci. Elsevier Ltd; 2013;18(10):575–83.

155. Sanchez SE, Kay SA. The plant circadian clock: from a simple timekeeper to a complex developmental manager. Cold Spring Harb Perspect Biol. 2016;a027748.

156. Corbesier L, Vincent C, Jang S, Fornara F, Fan Q, Searle I, et al. FT Protein Movement Contributes to Long-Distance Signaling in Floral Induction of Arabidopsis. Science (80-). 2007;316:1030–3.

157. Abe M, Kobayashi Y, Yamamoto S, Daimon Y, Yamaguchi A, Ikeda Y, et al. FD, a bZIP protein mediating signals from the floral pathway integrator FT at the shoot apex. Science (80-). 2005;309:1052–6.

158. Schaffer R, Ramsay N, Samach A, Corden S, Putterill J, Carré IA, et al. The late elongated hypocotyl mutation of Arabidopsis disrupts circadian rhythms and the photoperiodic control of flowering. Cell. 1998;93(7):1219–29.

159. Millar AJ, Carre IA, Strayer CA, Chua NH, Kay SA. Circadian clock mutants in Arabidopsis identified by luciferase imaging. Science (80-). 1995;267(5201):1161 LP-1163.

160. Park DH, Somers DE, Kim YS, Choy YH, Lim HK, Soh MS, et al. Control of circadian rhythms and photoperiodic flowering by the Arabidopsis GIGANTEA gene. Science (80-). 1999;285(5433):1579–82.

161. Suárez-López P, Wheatley K, Robson F, Onouchi H, Valverde F, Coupland G. CONSTANS mediates between the circadian clock and the control of flowering in Arabidopsis. Nature. 2001;410(6832):1116.

162. Turnbull C. Long-distance regulation of flowering time. J Exp Bot. 2011;62(13):4399–413.

163. Dong Z, Danilevskaya O, Abadie T, Messina C, Coles N, Cooper M. A gene regulatory network model for Floral transition of the shoot apex in maize and its dynamic modeling. PLoS One. 2012;7(8).

164. Tamaki S, Matsuo S, Wong HL, Yokoi S, Shimamoto K. Hd3a Protein Is a Mobile Flowering Signal in Rice. Science (80-). 2007;316:1033–6.

165. Wolabu TW, Zhang F, Niu L, Kalve S, Bhatnagar-Mathur P, Muszynski MG, et al. Three FLOWERING LOCUS T-like genes function as potential florigens and mediate photoperiod response in sorghum. New Phytol. 2016;210:946–59.

166. Yang S, Murphy RL, Morishige DT, Klein PE, Rooney WL, Mullet JE. Sorghum Phytochrome B Inhibits Flowering in Long Days by Activating Expression of SbPRR37 and SbGHD7, Repressors of SbEHD1, SbCN8 and SbCN12. PLoS One. 2014;9(8):e105352.

167. Cho L-H, Yoon J, Pasriga R, An G. Homodimerization of Ehd1 is required to induce flowering in rice. Plant Physiol. 2016;170:2159–71.

168. Xue W, Xing Y, Weng X, Zhao Y, Tang W, Wang L, et al. Natural variation in Ghd7 is an important regulator of heading date and yield potential in rice. Nat Genet. 2008;40:761.

169. Yang S, Weers BD, Morishige DT, Mullet JE. CONSTANS is a photoperiod

regulated activator of flowering in sorghum. BMC Plant Biol. 2014;14:148.

170. Quinby JR. The Maturity Genes of Sorghum. Adv Agron. 1967;19:267–305.

171. Karper RE, Quinby JR. The history and evolution of milo in the United States. Agron J. 1946;38(5):441–53.

172. Lin Y, Schertz KF, Paterson AH. Comparative Analysis of QTLs Affecting Plant Height and Maturity Across the Poaceae,. Genet Soc Am. 1995;141:391–411.

173. Lander Eric S., Green Philip, Abrahamson Jeff, Barlow Aaron, Daly Mark J., Lincoln Stephen E., et al. MAPMAKER: An interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics. 1987;1(2):174–81.

174. Wang S, Basten CJ, Zeng ZB. Windows QTL Cartographer 2.5. Raleigh, NC, NC: Department of Statistics, North Carolina State University; 2012.

175. Manichaikul A, Moon JY, Sen Ś, Yandell BS, Broman KW. A model selection approach for the identification of quantitative trait loci in experimental crosses, allowing epistasis. Genetics. 2009;181(3):1077–86.

176. Morishige DT, Klein PE, Hilley JL, Sahraeian SME, Sharma A, Mullet JE. Digital genotyping of sorghum - a diverse plant species with a large repeat-rich genome. BMC Genomics. 2013;14:1–19.

177. Livak KJ, Schmittgen TD. Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2-ΔΔCt Method. Methods. 2001;25(4):402–8.

178. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol. 1987;4(4):406–25.

179. Quinby JR, Karper R. The inheritance of three genes that influence time of floral initiation and maturity date in Milo. J Am Soc Agron. 1945;(901):916–36.

180. Arends D, Prins P, Jansen RC, Broman KW. R/qtl: high-throughput multiple QTL mapping. Bioinformatics. 2010;26(23):2990–2.

181. Tuinstra MR, Ejeta G, Goldsbrough PB. Heterogeneous inbred family (HIF) analysis: A method for developing near-isogenic lines that differ at quantitative trait loci. Theor Appl Genet. 1997;95(5–6):1005–11.

182. Xu S, Zhong C, Zhang T, Ding J. Structure of human lysine methyltransferase Smyd2 reveals insights into the substrate divergence in Smyd proteins. J Mol Cell Biol. 2011;3(5):293–300.

183. Spellmon N, Holcomb J, Trescott L, Sirinupong N, Yang Z. Structure and function of SET and MYND domain-containing proteins. Int J Mol Sci. 2015;16(1):1406–28.

184. Min J, Zhang X, Cheng X, Grewal SIS, Xu RM. Structure of the SET domain histone lysine methyltransferase Clr4. Nat Struct Biol. 2002;9(11):828–32.

185. Gross CT, McGinnis W. DEAF-1, a novel protein that binds an essential region in a deformed response element. EMBO J. 1996;15(8):1961–70.

186. Tarumoto I. Thermo-sensitivity and photoperiod sensitivity genes controlling heading time and flower bud initiation in Sorghum, Sorghum bicolor Moench. Japan Agric Res Q. 2011;45(1):69–76.

187. Gan ES, Xu Y, Wong JY, Geraldine Goh J, Sun B, Wee WY, et al. Jumonji demethylases moderate precocious flowering at elevated temperature via regulation of FLC in Arabidopsis. Nat Commun. 2014;5:1–13.

188. Escamez S, Tuominen H. Programmes of cell death and autolysis in tracheary elements: When a suicidal cell arranges its own corpse removal. J Exp Bot. 2014;65(5):1313–21.

189. Postma JA, Lynch JP. Root cortical aerenchyma enhances growth of Zea mays L. on soils with suboptimal availability of nitrogen, phosphorus and potassium. Plant Physiol. Am Soc Plant Biol; 2011;156:1190–201.

APPENDIX

PRIMER SEQUENCES

Table A-1. Chapter III primers for Sobic.006G147400 DNA and cDNA sequencing primers

Type	Genotype	Forward primer (5' → 3')	Reverse primer (5' → 3')
DNA template	IS3620c, Tx2910, BTx3297, RTx2909, Rio, BTx623	CAGCAGCGGTTTCTTTTGCT	TCAGCTGAACAGTCAGAAACCTT
DNA template	Tx7000	CCTGGCATGACACTACAGCA	AGTCAGCTGAACAGTCAGAAAC
DNA template	All remaining genotypes	CAGGCTGCAAGAGCGAGATA	TCGATCAGTCCTCTCAGCCA
DNA Sequencing	-	CCTGCAGGCTTCCTCGTCCCTATAAA	GCTCGTTGGCCACCTTCTTGTAGAGG
DNA Sequencing	-	CCTCTACAAGAAGGTGGCCAACGAGC	CAGCCCTCTGCACCATGTGATGCAC
DNA Sequencing	-	GTGCATCACATGGTGCAGAGGGCTG	GGTGACGAGTCTTGTGCGTTGGGATC
DNA Sequencing	-	GATCCCAACGCACAAGACTCGTCACC	AAGACGAGCGTCTTCCTCATGCC
DNA Sequencing	-	CCGCATCATCATCAGGAGAACCTG	CAGGTTCTCCTGATGATGATGCGG
DNA Sequencing	-	GGCATGAGGAAGACGCTCGTCTT	CCTTTTGGATTGGAGAGGTGCACTC
cDNA template	BTx623, IS3620c	CAGGCTGCAAGAGCGAGATA	CCTTCACAATGCCCAGTCCA
cDNA sequencing	-	CAGGCTGCAAGAGCGAGATA	TCTTTCCGCTTCTGGAACAC
cDNA sequencing	-	GTGTGCCACTACCTCTACAAG	ACATCGCTCCGTTCATCAG
cDNA sequencing	-	CGAGTGGTACTTCTTCAGCTTC	CCTTCACAATGCCCAGTCCA

Table A-2. Chapter III qRT-PCR primers

Gene ID	gene name	Forward primer (5' → 3')	Reverse primer (5' → 3')
Sobic.006G147400	*SbNAC_D*	AGAGTGCACCTCTCCAATCC	GCAAATGAAAATGACACCTCCT
Sobic.007G172100	*SbXCP1*	GTGAAGAACTCGTGGGGACC	ATGCGATTCAGAGCTCGTCG
Sobic.001G526600	*SbUBC*	CATGCTGCACATTCGCATAG	AGAGACATGGTCCACAAGAAC

Table A-3. Chapter IV Ma_2 (Sobic.002G302700) sequencing and qPCR primer sequences

Type	F/R	Primer sequence
Template amplification	F	GCGTCATGCTATGTTTCAGCC
	R	CATCCTGCTGCATAAGCTCC
Sanger sequencing	F	ATGCATCAATTGTTTGGTACAGA
	R	TTGATTCACGTCTGCATGCTT
qPCR	F	TCGGGCACTTGGAATTTGAGA
	R	TGCTGTGCGCAACAAAATTAGA

Table A-4. Chapter IV Genes in the fine-mapped *Ma2* QTL region.

Gene Name	Description
Sobic.002G300800	
Sobic.002G300850	
Sobic.002G300900	
Sobic.002G300950	
Sobic.002G301000	similar to Os07g0479100 protein
Sobic.002G301100	similar to Putative Mannose-P-dolichol utilization defect 1 protein homolog
Sobic.002G301200	similar to Serine carboxypeptidase-like precursor
Sobic.002G301300	similar to Putative uncharacterized protein
Sobic.002G301400	similar to Os07g0479600 protein
Sobic.002G301500	PTHR11895//PTHR11895:SF68 - AMIDASE // SUBFAMILY NOT NAMED (1 of 3)
Sobic.002G301600	similar to OSIGBa0148J22.1 protein
Sobic.002G301700	similar to OSIGBa0148J22.1 protein
Sobic.002G301800	PF03514 - GRAS domain family (GRAS) (1 of 81)
Sobic.002G301900	similar to OSIGBa0115A19.4 protein
Sobic.002G302000	similar to Putative uncharacterized protein
Sobic.002G302100	similar to Putative uncharacterized protein
Sobic.002G302200	Predicted protein
Sobic.002G302300	similar to Putative uncharacterized protein
Sobic.002G302400	similar to Os06g0241000 protein
Sobic.002G302500	similar to H0525E10.7 protein
Sobic.002G302600	similar to Putative uncharacterized protein
Sobic.002G302700	PTHR12197:SF166 - ZINC ION BINDING PROTEIN (1 of 1)
Sobic.002G302800	similar to Putative uncharacterized protein
Sobic.002G302900	similar to Putative uncharacterized protein
Sobic.002G303000	similar to Os07g0481300 protein
Sobic.002G303100	weakly similar to Pollen signalling protein with adenylyl cyclase activity-like
Sobic.002G303200	similar to Putative uncharacterized protein OJ1136_F08.108
Sobic.002G303300	similar to Chromosome chr2 scaffold_105, whole genome shotgun sequence
Sobic.002G303400	
Sobic.002G303500	similar to Actin-depolymerizing factor 1
Sobic.002G303600	similar to Putative zinc finger (C3HC4-type RING finger) protein
Sobic.002G303700	similar to Os12g0576600 protein
Sobic.002G303800	similar to Putative uncharacterized protein OJ1753_E03.102
Sobic.002G303900	similar to Os07g0485000 protein
Sobic.002G303950	
Sobic.002G304000	similar to Putative uncharacterized protein
Sobic.002G304100	similar to Putative asparaginyl-tRNA synthetase, chloroplast/mitochondrial
Sobic.002G304300	KOG4510 - Permease of the drug/metabolite transporter (DMT) superfamily (1 of 1)
Sobic.002G304400	similar to Putative uncharacterized protein

Table A-4. continued

Gene	Description
Sobic.002G304500	
Sobic.002G304600	Predicted protein
Sobic.002G304700	K06316 - oligosaccharide translocation protein RFT1 (RFT1) (1 of 1)
Sobic.002G304801	PF08387 - FBD (FBD) (1 of 77)
Sobic.002G304900	similar to Ribosomal protein L15
Sobic.002G305000	similar to Auxin-regulated protein-like
Sobic.002G305100	similar to Putative uncharacterized protein
Sobic.002G305200	PF11321 - Protein of unknown function (DUF3123) (DUF3123) (1 of 6)
Sobic.002G305300	PF11321 - Protein of unknown function (DUF3123) (DUF3123) (1 of 6)
Sobic.002G305400	similar to Putative uncharacterized protein
Sobic.002G305500	
Sobic.002G305600	similar to Putative uncharacterized protein
Sobic.002G305700	similar to 2Fe-2S iron-sulfur cluster protein-like
Sobic.002G305750	
Sobic.002G305800	PF01357//PF03330 - Pollen allergen (Pollen_allerg_1) // Rare lipoprotein A (RlpA)-like double-psi beta-barrel (DPBB_1) (1 of 76)
Sobic.002G305900	similar to Putative uncharacterized protein P0627E10.27
Sobic.002G306000	similar to CaM-binding transcription factor
Sobic.002G306100	similar to Putative uncharacterized protein P0477A12.34
Sobic.002G306201	PTHR10516:SF178 - PEPTIDYL-PROLYL CIS-TRANS ISOMERASE FKBP20-2, CHLOROPLASTIC (1 of 1)
Sobic.002G306300	similar to Putative uncharacterized protein
Sobic.002G306400	similar to Chromosome chr15 scaffold_19, whole genome shotgun sequence
Sobic.002G306500	PTHR22761//PTHR22761:SF20 - SNF7 - RELATED // SUBFAMILY NOT NAMED (1 of 1)
Sobic.002G306600	similar to Putative uncharacterized protein
Sobic.002G306700	similar to Putative uncharacterized protein
Sobic.002G306800	similar to Putative uncharacterized protein
Sobic.002G306900	similar to Nucleoside diphosphate kinase 1
Sobic.002G307000	Predicted protein
Sobic.002G307100	KOG0583//KOG0594 - Serine/threonine protein kinase // Protein kinase PCTAIRE and related kinases (1 of 4)
Sobic.002G307200	KOG0583//KOG0594 - Serine/threonine protein kinase // Protein kinase PCTAIRE and related kinases (1 of 4)
Sobic.002G307300	KOG0583//KOG0594 - Serine/threonine protein kinase // Protein kinase PCTAIRE and related kinases (1 of 4)
Sobic.002G307400	3.4.19.12 - Ubiquitinyl hydrolase 1 / Ubiquitin thiolesterase (1 of 89)
Sobic.002G307500	3.4.19.12 - Ubiquitinyl hydrolase 1 / Ubiquitin thiolesterase (1 of 89)
Sobic.002G307600	similar to Putative uncharacterized protein

Table A-4. continued

Gene	Description
	weakly similar to Putative uncharacterized protein
Sobic.002G307700	OSJNBb0055I24.118
Sobic.002G307800	similar to Putative uncharacterized protein OSJNBb0055I24.117