TRANSCRIPTOME ANALYSIS DURING STEM DEVELOPMENT AND IN THE A_1 CYTOPLASMIC MALE STERILITY SYSTEM OF SORGHUM

A Dissertation

by

MAHNAZ KIANIFARIZ

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,	Patricia E. Klein
Co-Chair of Committee,	Robert R. Klein
Committee Members,	William L. Rooney
	Xiuren Zhang
Head of Department,	Dirk B. Hays

May 2017

Major Subject: Molecular and Environmental Plant Sciences

ABSTRACT

Sorghum (*Sorghum bicolor* (L) Moench) exhibits efficient use of water, nitrogen and energy resources and is grown throughout the world as a cereal, forage, syrup and more recently energy crop. A high-throughput RNA sequencing method (RNA-seq) was used to examine transcriptional dynamics during stem development of two genotypes of sorghum, i.e., BTx623 and R07020. A comparative transcriptome analysis of immature panicles was also conducted between a set of cytoplasmic male sterile (CMS) A_1-lines, iso-cytoplasmic maintainer B-lines, and F_1 sorghum hybrids.

By comparing stem nodal segments each at unique stages of cell wall deposition, more than 500 differentially expressed genes (DEGs) were identified that were annotated as being involved in processes related to cell wall biosynthesis. Categories of DEGs included genes involved in cellulose, hemicellulose, and lignin biosynthesis, transcription factors, glycosyl transferases and cell wall proteins. In non-elongating mature nodal segments actively laying down secondary cell walls, transcriptome profiles revealed inter-related biosynthetic pathways that were highly enriched in DEGs involved in phenylpropanoid metabolism and flavonoid biosynthesis. In immature nodal segments in the rapidly elongating region of the stem, a series of pathways enriched in DEGs were detected that are critical for the primary growth of stems including starch and sucrose metabolism and DNA replication. A co-expression network analysis identified a number of transcription factors with connectivity to lignin biosynthesis genes, which indicates a possible role for these transcription factors as

regulators of lignin biosynthesis in sorghum. A number of DEGs were found by comparing nuclear gene expression of CMS A_1-lines with associated B-lines and F_1 hybrids. In general, a larger proportion of DEGs were down-regulated in CMS A_1-lines when compared to pollen-fertile B-lines and F_1 hybrids. GO categories whose genes were down-regulated in CMS A_1-lines included metabolic processes, lipid biosynthetic/metabolic process, and oxidation/reduction. Examination of these DEGs (and their homologs) within these categories provided evidence that the down-regulation of these genes may relate to the production of viable pollen in fertile panicles. In pollen-sterile panicles of CMS A_1-lines, up-regulated genes included several stress-related genes such as heat shock and linoleic acid biosynthesis genes.

ACKNOWLEDGEMENTS

I would like to express my sincere appreciation to my committee chair, Prof. Patricia Klein, for giving me the opportunity to join her lab, as well as for her guidance and support during this journey. I sincerely appreciate my co-chair, Dr. Robert Klein, for his thoughtful reading and careful contributions, and for giving me valuable feedback and challenging me to be better every day. I am grateful for my dissertation committee members, Prof. William Rooney and Dr. Xiuren Zhang, who gave me insightful comments. I would like to thank Dr. Noushin Ghaffari and Jizhou Yang for bioinformatic assistance. I also extend my gratitude to our collaborator, Dr. Doreen Ware, at Cold Spring Harbor Laboratory. I would also like to acknowledge Dr. Stanislav Vitah and Dr. Leo Hoffmann who assisted me for stem histochemical and compositional analysis, respectively. I take this opportunity to express my sincere thanks to Prof. David Byrne, who played a crucial role in my academic career.

Many thanks go to the Sorghum Genomics lab members, past and present, including Natalie Patterson, Julie McCollum, Dr. Diana Dugas, Dr. Nikhil Patil, Jizhou Yang, Nolan Bentley, Bárbara Ferrara, Fabian Echeverria, Elmer Roldan, Austin Blevins, Rory Tucker, and John Rocha for their assistance and kind friendship that enriched my graduate study experience. I would like to thank all members of the Horticultural Sciences department as well as those within the MEPS program for their assistance during my graduate study. I am also so thankful to Nicholas Cenegy from the Texas A&M University Writing Center for all of his valuable assistance.

My time at TAMU was made all the more pleasant by having good friends. Thank you all for being a part of my life. Finally, I present my heartfelt thanks to my father, a source of strength and support, who has taught me to be the architect of my life and my mother, a deep ocean of love, who has always supported me. I am also very grateful to my brother, who is a source of encouragement, and my nephew Kiarash and my niece Yekta who always give me sweet happiness.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by a dissertation committee consisting of Prof. Patricia Klein of the Department of Horticultural Sciences, Dr. Robert Klein of the USDA-ARS, Prof. William Rooney of the Department of Soil and Crop Sciences, and Dr. Xiuren Zhang of the Department of Biochemistry & Biophysics.

For both projects, RNA-seq template preparation and paired-end sequencing on an Illumina HiSeq2500 was performed by Texas AgriLife Research Genomic and Bioinformatics Services. For the lignin histochemical staining of sorghum stem samples, sample dissection, microwave-assisted fixation, washing, and dehydration were conducted using protocols established by the Microscopy and Imaging Center at Texas A&M University by Dr. Stanislav Vitah. The cross-sections and staining were done at the Veterinary School Histopathology Lab by Sarah Jones and subsequently scanned using a slide scanner at the TAMU Veterinary School's GI lab by Kimberly Green. Compositional analysis of sorghum stems was performed by Dr. Leo Hoffmann at the TAMU Sorghum breeding lab.

Funding Sources

This work was made possible by funding from Texas AgriLife Research, the USDA-ARS, and the Department of Horticultural Sciences at Texas A&M University.

NOMENCLATURE

CMS	Cytoplasmic Male Sterility
DEGs	Differentially Expressed Genes
GO	Gene Ontology
KEGG	Kyoto Encyclopedia of Genes and Genomes
LN_2	Liquid Nitrogen
N/A	Not available
Rf	Fertility Restoration
RNA-seq	High-throughput mRNA sequencing
WGCNA	Weighted Gene Co-expression Network Analysis

TABLE OF CONTENTS

LIST OF FIGURES

Page

LIST OF TABLES

Page

CHAPTER I

INTRODUCTION

Increasing food and fuel demands, in parallel with increasing world wide population, have renewed the urgency for improving crop yield and developing sustainable biofuel resources. *Sorghum bicolor* (L) Moench, a C4 photosynthesis species from the Poaceae family, is an important crop species in the United States and around the world (Rooney, 2014). Sorghum has a number of advantageous characteristics, including high rates of carbon fixation, high water and nutrient use efficiency, high biomass productivity, and adaptation to diverse environments (Dalal, et al., 2012). Sorghum originated from semi-arid regions of Africa (Smith and Frederiksen, 2000), but it has been adapted to a wide variety of climates, including temperate and humid environments (Saballos, 2008). Sorghum is grown throughout the world as a cereal, forage, syrup and more recently energy crop (Rooney, 2014). Each of these has resulted in the development of lines with characteristics optimized for their end use (Saballos, 2008). Grain sorghum, as a cereal crop, has a high ratio of panicle-to-green biomass. The dwarf and low tillering characteristics of grain sorghum hybrids make them suitable for combine harvest, whereas forage type sorghums can display abundant tillering and produce green biomass as the main product and are usually harvested before the grain reaches maturity. Sweet sorghums are varieties that have a high concentration of soluble sugars (sucrose and/or glucose) in the stalk juice. In this type of sorghum, the fermentation process can proceed without pretreatment, but it needs to proceed quickly due to the instability of the sugar in the stalk

and/or juice (Saballos, 2008; Rooney, 2014). Biomass sorghum is a specific type of photoperiod-sensitive sorghum that, when grown in long day environments, accumulates large quantities of biomass. The extended vegetative growth allows the plant to capture a higher amount of solar energy and convert it to biomass (Miller, et al., 1968; Rooney, et al., 2007; Rooney, 2014). Biomass sorghum hybrids can accumulate more than twice as much biomass as grain sorghum (Mullet, et al., 2014). The development of these hybrids was facilitated by the identification of complementary alleles at specific maturity loci (Rooney and Aydin, 1999).

Most of the sorghums in the world collection are photoperiod sensitive, meaning that reproductive growth is initiated once day length is sufficiently reduced to meet the required short day photoperiod (Reddy, et al., 2006). Beginning in the late 1800s, photoperiod insensitive sorghum genotypes were developed for their use in the United States (Stephens, et al., 1967; Klein, et al., 2015). The Sorghum Conversion program provided further new and diverse germplasm by moving recessive dwarfing and maturity genes from a four-dwarf temperate zone variety into the genomes of exotic lines (Stephens, et al., 1967; Klein, et al., 2008).

Although sorghum is primarily self-pollinated, its improvement has long relied on hybridization within the species (Rooney, 2004; Kuhlman, et al., 2008; Hodnett, et al., 2010). The possibility of sorghum hybrid seed production began in the early 1950s when geneticists introduced a kafir nuclear genome into a milo cytoplasm, and cytoplasmic male sterility (CMS) was discovered (Stephens and Holland, 1954). In sorghum, fertility is restored in F_1 hybrids by pollinating CMS milo lines with sorghum cultivars of milo origin

that harbor nuclear-encoded restorer genes. This method was first used for hybrid seed production in 1956, and spread fast in the U.S. such that within five years, 90% of the U.S. grain sorghum cultivation area was planted in hybrids. The use of F_1 hybrids in grain sorghum breeding programs has allowed these programs to take advantage of heterosis (Rooney, 2014). Over the next 50 years, the development of other types of sorghum hybrids followed that of grain sorghum hybrids, including sweet sorghum (Broadhead, 1972; Broadhead, 1982; Rooney, 2014), forage sorghum (Pedersen, et al., 1982; Venuto and Kindiger, 2008), and biomass sorghum (Rooney and Aydin, 1999; Rooney, et al., 2007).

With the availability of a whole-genome sequence for sorghum (Paterson, et al., 2009), it has become possible to identify target genes in sorghum based on a combination of genomics, bioinformatics, and experimental data. The advent of next-generation high-throughput sequencing has significantly advanced biological studies and provided a more comprehensive view of biological development (Lister, et al., 2009; Marguerat and Bähler, 2010). High-throughput mRNA sequencing (RNA-seq) is a genome-wide profiling method that has been used to discover novel transcripts, splicing models, allele-specific expression, and unique transcript splice junctions (Malone and Oliver, 2011). Similar to the previous large-scale transcript profiling platforms that included microarrays, RNA-seq is progressively being used to examine transcriptional dynamics during various phases of plant growth and development (Wang, et al., 2009).

RNA-seq can advance our understanding of key biological processes that are involved in biomass production. Understanding cell wall biosynthesis gene regulatory networks may improve the production of biofuel from biomass. The production of bioethanol from

lignocellulosic biomass involves the process of saccharification, the conversion of cellulose within the cell wall to glucose. However, because cellulose forms a complex structure with hemicellulose and lignin, its conversion to sugar is an involved and costly process that makes the commercialization of bioethanol production from lignocellulosic biomass challenging (Alvira, et al., 2010). In addition, apart from the lack of information on cell wall biosynthesis, the mechanism of CMS and fertility restoration has not been fully elucidated in sorghum. A better understanding of the genetic control of cytoplasmic male sterility and fertility restoration in sorghum may have significant impacts on improving hybrid breeding programs. This can increase the genetic gain by improving the hybrids.

Specifically, this research was based on transcriptome analyses using RNA-seq and focused on the following objectives:

1. Characterizing the transcriptome in nodal segments and elucidating changes in gene expression profiles that arise in two sorghum genotypes during stem maturation.
2. Comparing the nuclear genome transcriptome of A_1 CMS and fertile lines to elucidate molecular mechanisms associated with A_1 CMS and the action of *Rf* genes in sorghum.

CHAPTER II

GENE EXPRESSION DURING STEM DEVELOPMENT OF TWO SORGHUM GENOTYPES

Introduction

A promising alternative to first generation biofuel produced from grain is the production of biofuel from the fermentation of lignocellulosic feedstocks (Demirbas, 2009; Kang, et al., 2014). Sorghum is a short-day tropical cereal that requires day lengths shorter than 12.5 hr for the induction of a floral meristem. In the long-day temperate growing regions, tropical sorghum often remains vegetative until late in the growing season, and through the selection of mutations in key flowering time (maturity) genes (Quinby and Karper, 1945), sorghum has been adapted for grain production in temperate climates. Armed with a detailed understanding of maturity genes and their control of floral initiation, sorghum geneticists have bred elite photoperiod-sensitive sorghum hybrids that will not flower (or flower very late in the growing season) in temperate climates and thus, will continue to accumulate vegetative biomass indeterminately for the duration of the growing season (Rooney and Aydin, 1999; Rooney, et al., 2007; Yang, et al., 2014b). Sorghum was first considered as a bioenergy crop during the fuel crisis in the late 1970s and early 1980s (Monk, et al., 1984). The indeterminate vegetative growth of photoperiod-sensitive sorghums in long days has been reported to yield double the vegetative production as photoperiod-insensitive grain sorghum (Rooney and Aydin, 1999; Rooney, et al., 2007; Olson, et al., 2012). By the end of the growing season, nearly

83% of the plant dry matter is in the stem of biomass sorghum (Olson, et al., 2012), which is desirable for cellulosic ethanol production compared with other cellulose sources such as leafy forage grasses (Prakasham, et al., 2014). In addition, sorghum's drought tolerance and relatively low input requirements (e.g., pesticides, supplemental irrigation) are important attributes for biomass crops that will likely be grown on low-input marginal lands (Mullet, et al., 2014).

As biomass sorghum is being bred specifically for the bioenergy market, the composition of the biomass is of concern (Stefaniak and Rooney, 2013). Plant dry matter is composed of a complex variety of components including cellulose, noncellulosic polysaccharides, lignin, ash, protein and extractive compounds such as chlorophyll, waxes, oils, terpenes and phenolics (Browning, 1963; Carroll and Somerville, 2009). Structural carbohydrates, cellulose, hemicelluloses, and lignin are the main source of lignocellulosic biomass that are deposited in plant cell walls (Jung and Ni, 1998; Stefaniak and Rooney, 2013). Cellulose, the most abundant biopolymer on earth, has a linear chain of β-1,4-linked D-glucan units while hemicellulose has a β-(1→4)-linked backbone with an equatorial configuration (Scheller and Ulvskov, 2010). Lignin, a hydrophobic phenolic polymer, saturates the cellulose and hemicellulose network that stabilizes the cell wall and also protects the secondary wall against biotic and abiotic injury (Albersheim, et al., 2010). Bioethanol production from lignocellulosic biomass involves the process of saccharification, i.e. the conversion of cellulose within the cell wall to glucose. However, since cellulose forms a complex structure with hemicellulose and lignin, it is not readily fermentable without pretreatment (Alvira, et al., 2010). Lignin,

the main obstacle in the conversion process, reduces the efficiency of pretreatment by binding enzymes and decreasing their ability to deconstruct cellulose and hemicellulose to simple sugars (Saritha and Arora, 2012). Nevertheless, lignin is a critical component that confers strength to stems thereby preventing lodging, which can be especially prevalent in high biomass sorghums. Therefore, the conversion of cellulose to sugar is a costly and complex process that makes the commercialization of bioethanol production from lignocellulosic biomass challenging (Limayem and Ricke, 2012; Furtado, et al., 2014).

In recent years, a number of studies have examined the genes that control cell wall biosynthesis in plants. In dicot species, high-throughput transcriptome profiling techniques, such as RNA-seq and microarrays have been used for the identification of expressed genes involved in cell wall formation in *Arabidopsis* (Imoto, et al., 2005; Yokoyama and Nishitani, 2006; Minic, et al., 2009; Cassan-Wang, et al., 2013), *Populus* (Andersson-Gunneråas, et al., 2006), *Acacia* (Wong, et al., 2011), and alfalfa (Yang, et al., 2011). Studies have also focused on the synthesis of cell wall components in grasses, such as maize (Guillaumie, et al., 2007; Bosch, et al., 2011; Courtial, et al., 2013), *Brachypodium* (Vogel, et al., 2010), barley (Christiansen, et al., 2011), and rice (Hirano, et al., 2013). Recently, McKinley, et al. (2016) characterized gene expression in the stems of the sweet sorghum cultivar Della starting at the transition from vegetative growth to floral initiation, and extended the characterization of stem gene expression through post grain maturity. The results clearly identified large gene families involved in stem growth, cell wall biosynthesis, and sucrose accumulation that are differentially expressed in stems

of sweet sorghums. The work of McKinley, et al. (2016) provides the opportunity to contrast the gene expression in stems of photoperiod-sensitive sorghums to that of sweet sorghums that have been bred for the dual purpose of stem sugar and grain production.

In the present study, RNA-seq technology was utilized to examine gene expression in nodal segments along the stem of two sorghum genotypes. Transcriptome profiling of nodal regions from the base to the apical region of the stem allowed an examination of the changes in gene expression that occur as secondary wall formation progresses in a developmental manner from the newly developed upper nodal segments to the more mature nodal segments at the stem base.

Materials and Methods

Sorghum genotypes and growth conditions

Two elite sorghum lines were examined in the study under a long-day photoperiod regime. Sorghum inbred BTx623 is a photoperiod-insensitive 3-dwarf cultivar with plant height of ~1.2 meters that flowers after 72 days under normal long-day environments (Burow, et al., 2011). The R07020 genotype is a photoperiod-sensitive line developed at Texas A&M University that is likely 1-dwarf with plant height of ~4.5 meters and flowers very late in a long-day environment (~120 days depending on the planting date). The two genotypes were grown under controlled greenhouse conditions with a temperature range of 24 ± 2°C (night) to 30 ± 2°C (day) with light (14hr/10hr light/dark regime) provided using sodium halide lights and natural sunlight. Plants were grown in 3-gallon pots containing Sunshine REPS soil mix (Sun Grow Horticulture Inc.,

Bellevue, WA). Into each 3-gallon pot of Sunshine REPS soil, the following were added and mixed in; 18 g of Osmocote (16% N, 3.5% P and 10% K), 15 g gypsum, 15 g dolomite and 5 g Micromix (6% Ca, 3% Mg, 12% S, 0.1% B, 1% Cu, 17% Fe, 2.5% Mn, 0.05% Mo and 1% Zn). The plants were grown with nine replicates for each genotype. For a subset of the plants, plant height, leaf and node number, and stem diameter at select internodes along the length of the stem were recorded each week for ~13 weeks to monitor plant growth.

Stem sectioning and RNA isolation

Sixty days after planting (i.e. during the phase of linear stem elongation), three individual plants of each genotype were harvested, leaves and leaf sheaths removed, and nodes were counted starting from the base of the plant. Nodes were labeled beginning with the node immediately above the brace roots (N1), and sequentially labeled to the top of the stem. Nodal segments were excised (node plus 5mm on either side) and immediately frozen in liquid nitrogen (LN_2) with subsequent storage at -80^{o}C. Three different stem nodal segments were excised from each stem as follows; nodal segment 6 (N6) representing a mature non-elongating region near the stem base, which is actively laying down secondary cell walls; nodal segment 10 (N10) for genotype R07020 or nodal segment 11 (N11) for genotype BTx623, each representing nodal segments in the middle section of the stem where rapid cell wall biosynthesis is occurring; and nodal segment 14 (N14) for genotype R07020 or nodal segment 16 (N16) for genotype BTx623, representing newly formed nodal segments near the apical meristem of the stem.

For RNA extraction, frozen nodal segments were ground under LN_2 into a fine powder using a mortar and pestle. RNA was extracted from 600 mg of frozen ground tissue using the Trizol™ reagent as detailed by the manufacturer (Invitrogen, Carlsbad, CA, USA) and subsequently treated with TURBO™ DNase (Ambion, Austin, TX, USA). RNA quality and quantity were assessed using the ND-1000 NanoDrop spectrophotometer (NanoDrop Technologies, Montchanin, DE, USA). Three equimolar RNA samples extracted from three technical reps for each biological rep were pooled after extraction for RNA-seq template preparation. The quality of each pooled RNA sample was assessed with an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA) by Texas AgriLife Research Genomic and Bioinformatics Services prior to RNA-seq template preparation.

RNA-seq library preparation, sequencing and quality control

RNA-seq template preparation and paired-end (PE) sequencing on an Illumina HiSeq2500 was completed at Texas AgriLife Research Genomic and Bioinformatics Services. Libraries were prepared using the Illumina TruSeq RNA kit (Illumina, San Diego, CA, USA) following the manufacturer's protocol. Each library was bar-coded, and nine libraries were pooled per lane on the Illumina flow cell. The experimental units were prepared individually to serve as a biological replicate for downstream data analysis. Three RNA-seq libraries were prepared for each of the nodal segments from sorghum genotypes BTx623 and R07020, and each library represented a biological replication in the RNA-seq analysis. The paired-end reads in the biological replicates ranged from 64 to 71 bp in length as the experiments were conducted over time and the

samples were run on different Illumina flow cells. The total read counts from the three reps were combined after trimming all reads to 64 bp to obtain higher sequencing depth for accurate detection of gene expression changes. The quality of RNA-seq reads was assessed using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/).

Gene expression analysis

Sequence reads were imported into the CLC Genomics Workbench version 8.5.1 (Qiagen, Valencia, CA, USA), trimmed to 64 bp, and mapped to the *Sorghum bicolor* BTx623 reference genome [Sbicolor_255 v.2.1, www.phytozome.jgi.doe.gov, (Paterson, et al., 2009)]. Based on total read counts for each annotated gene, differential gene and transcript expression analyses were conducted using the Empirical analysis of DGE tool, which implements the 'Exact Test' for two-group comparisons (Robinson, et al., 2010). Differentially expressed genes (DEGs) were defined as having a $\log_2$ fold change ≥ 1 or ≤ -1 with a false discovery rate (FDR) corrected *p*-value <0.05. Genes annotated as cell-wall related in other species (rice, maize, *Arabidopsis*) were identified based on various cell wall annotation databases and the literature, including MAIZEWALL (Guillaumie, et al., 2007), Cell Wall Genomics (https://cellwall.genomics.purdue.edu/), Bosch, et al. (2011), Hirano, et al. (2013) and Minic, et al. (2009). The orthologs of collected cell wall genes were identified in sorghum using the annotation information of the sorghum genome [Sbicolor_255 v.2.1, www.phytozome.jgi.doe.gov, (Paterson, et al., 2009)], BLAST analysis, and also using the Rice Genome Annotation Project Database (http://rice.plantbiology.msu.edu/annotation_pseudo_apk.shtml). The expression of orthologous cell-wall related genes was then examined in the nodal stem segments of

sorghum. For functional annotation, gene ontology was performed using AgriGO gene ontology analysis tools (Du, et al., 2010), and significantly enriched GO terms (in comparison to the genome background) were identified by REVIGO (Supek, et al., 2011). Pathway analysis was performed using the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (Kanehisa and Goto, 2000) and pathway enrichment analysis completed using the KOBAS server (version v.2) (Xie, et al., 2011).

Co-expression network analysis

The weighted gene co-expression network analysis (WGCNA) package in R (Jung, et al., 2012) was employed to construct the co-expression network and to identify highly correlated genes (Langfelder and Horvath, 2008). Network construction is theoretically straightforward: nodes represent genes and nodes are connected through edges if the corresponding genes are significantly co-expressed across samples (Zhang and Horvath, 2005). WGCNA was performed on data collected for 6,517 genes selected from those DEGs identified in a comparison of different nodal segments in genotype R07020. The expression of these genes was also examined in genotype BTx623. Each biological and technical replicate was considered as an individual dataset, totaling 18 samples. The power=14, merge at height=0.25, weight threshold=0.1 were used as selected parameters to construct the network. The network was visualized and analyzed in Cytoscape software (Shannon, et al., 2003). The lignin biosynthesis genes were selected as the hub genes. The resulting network was filtered by lignin biosynthesis genes, and by TFs to identify those TFs that are involved in secondary cell wall biosynthesis.

Lignin histochemical staining

To characterize cell wall deposition and lignin composition, nodal segments were cross-sectioned and histochemically analyzed. Nodal segments for histochemical analysis were collected in the same manner as those collected for RNA-seq analysis. Sample dissection, microwave-assisted fixation, washing, and dehydration were conducted using protocols established by the Microscopy and Imaging Center at Texas A&M University. Briefly, nodal segments preserved in 70% ethanol were cross-sectioned at 15 µm thickness with a sliding-type microtome. Sections were stained with Alcian Blue at pH 2.5 and 0.1% Safranin O. and subsequently scanned using the TAMU Veterinary School's GI lab slide scanner.

Compositional analysis

Nodal segments for compositional analysis were collected in the same manner as those collected for RNA-seq analysis. Samples from 60-day-old seedlings were harvested, and dried in an oven at 60°C for approximately five days. The dried samples were ground in a coffee grinder to a size that could pass a 2 mm sieve. Compositional analyses were conducted using near-infrared spectroscopy (NIR) with a FOSS XDS NIR-spectrometer (Foss North America, Eden Prairie, MN, USA) based on the method of Stefaniak, et al. (2012). Samples were collected from two biological replicates. The ground tissue from 10 technical replicates was pooled together to provide sufficient sample for NIR analysis. Each sample were scanned twice to ensure reproducibility of the NIR measurement. Near-infrared predictions for lignin (%), cellulose (%), xylan (%), glucan (%) and protein (%) were based on NIR calibration curves developed by the Texas A&M AgriLife Research

Sorghum Breeding Lab as described by Dykes, et al. (2014). Pairwise comparisons were made between different nodal segments for each genotype. Means were compared by independent student's T-test calculated with IBM SPSS Statistics (version v.23) (SPSS, Chicago, IL, USA). Statistical significance was considered at a *p*-value < 0. 05.

Results and Discussion

Vegetative growth of sorghum genotypes

R07020 is a photoperiod-sensitive genotype that does not flower under a 14hr/10hr light/dark photoperiod but rather requires a day length less than 12 hr 20 min to induce floral initiation (Rooney and Aydin, 1999). In contrast, BTx623 is photoperiod insensitive and thus will flower under long- or short-day photoperiods. In the present study, a series of comparisons were conducted to profile gene expression in stems of sorghum bred as a biofuel feedstock or for grain production. The vegetative growth characteristics of the two genotypes are shown in Figure 1. R07020 was ~250 cm tall after 90 days of growth and produced up to 21 leaves. In BTx623, stem elongation ceased after panicle emergence (after ~65 days) with plants growing to ~100 cm and producing up to 15 leaves. The growth differences were due to the selection of R07020 for vegetative production (dominant alleles at a majority of the *Dw* loci) and photoperiod-sensitivity (dominant alleles at the *Ma1 and Ma6* loci) rather than for dwarfism and photoperiod-insensitivity as selected for in BTx623 (Rooney and Aydin, 1999; Rooney, 2004).

A

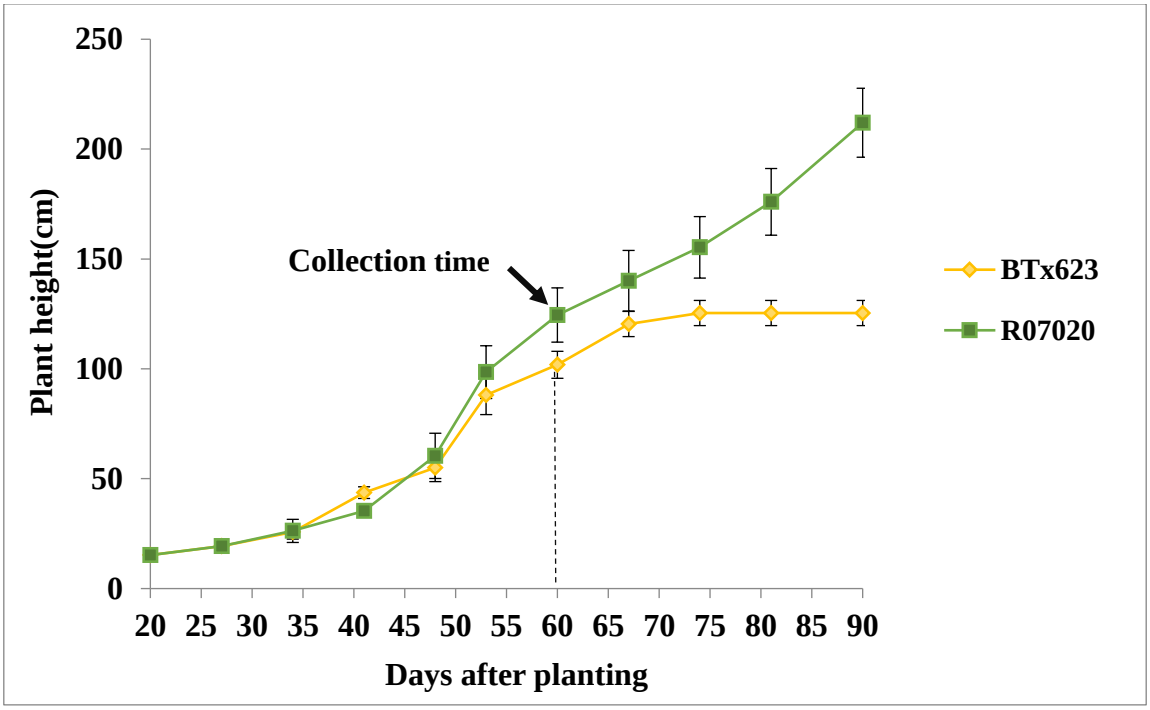
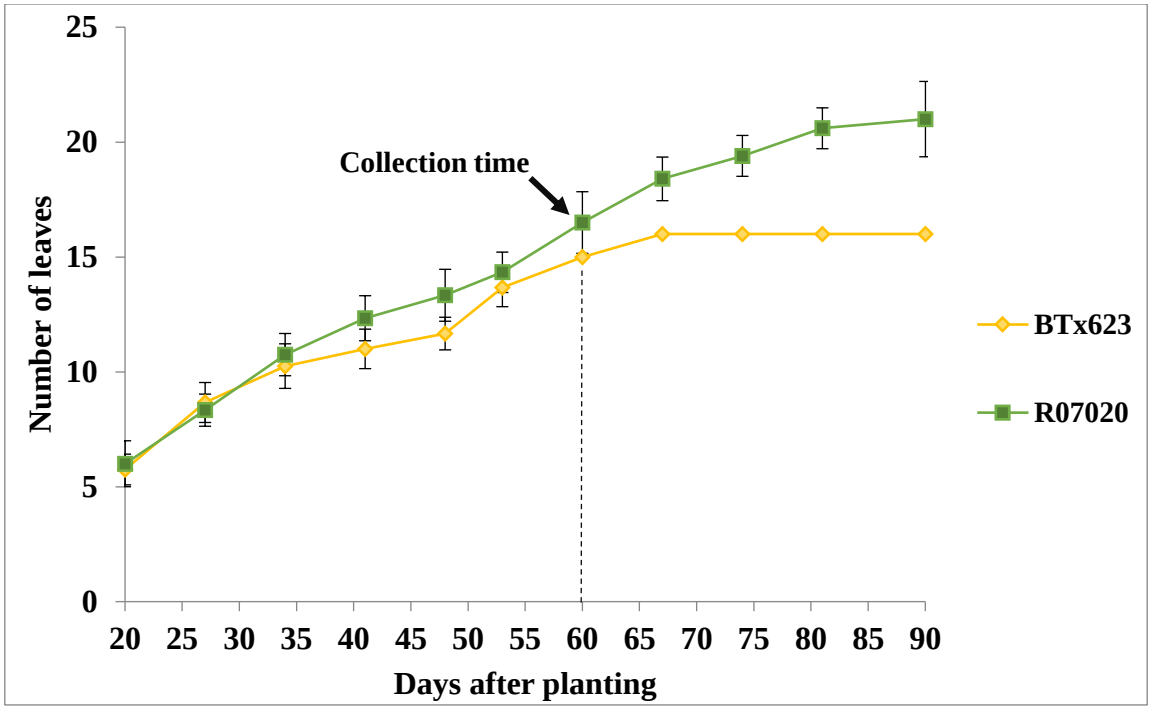


Figure 1: Vegetative growth characteristics of sorghum genotypes R07020 and BTx623. Plant height (A), leaf number (B), and diameter at three nodal segments along the length of the stem (C, D, E) were measured each week from 20 to 90 days after planting to monitor plant growth. Error bars indicate standard deviation.

B



C

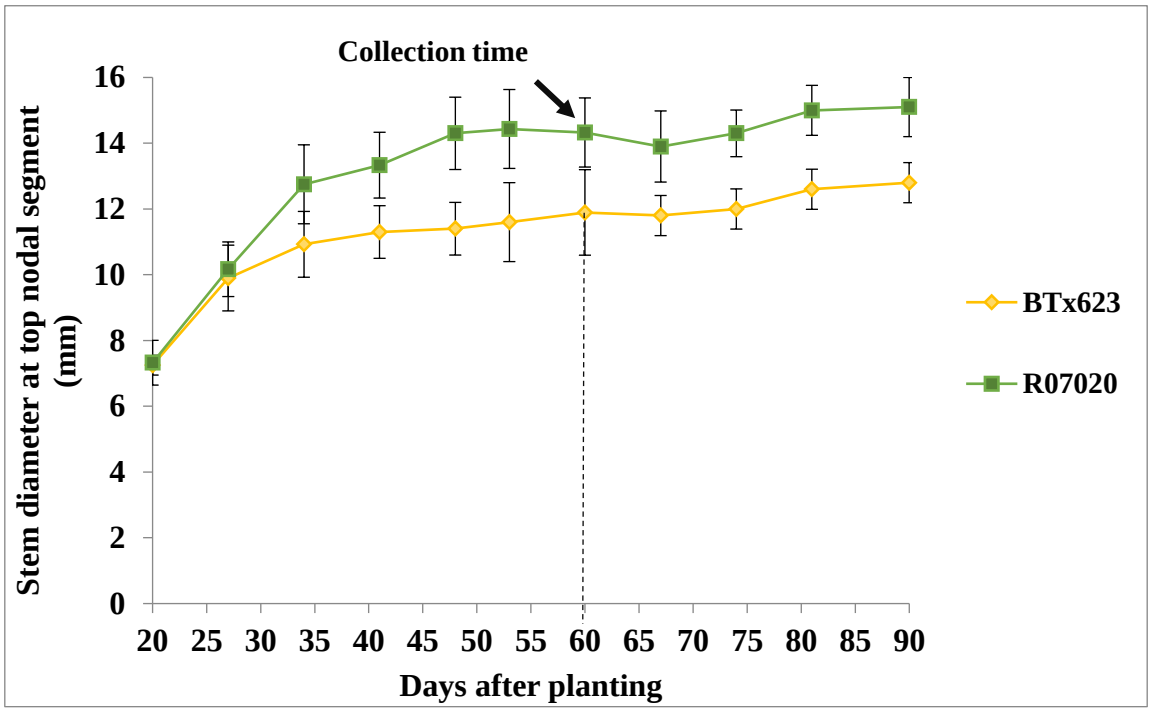
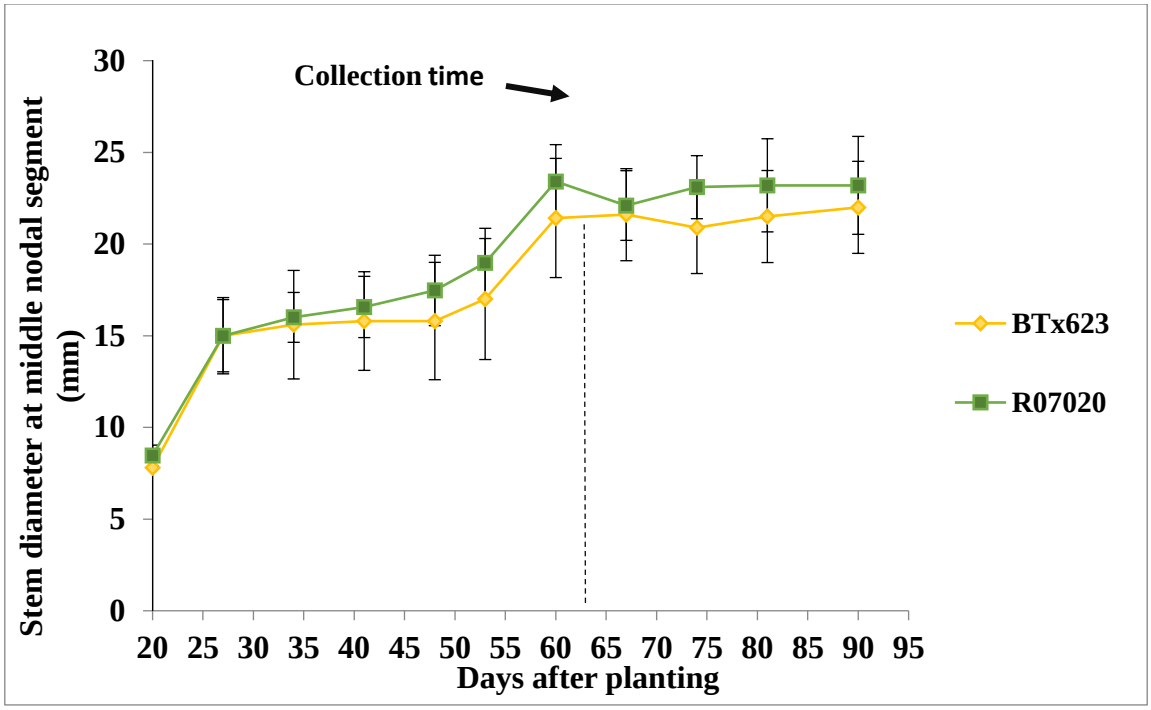


Figure 1: Continued

D



E

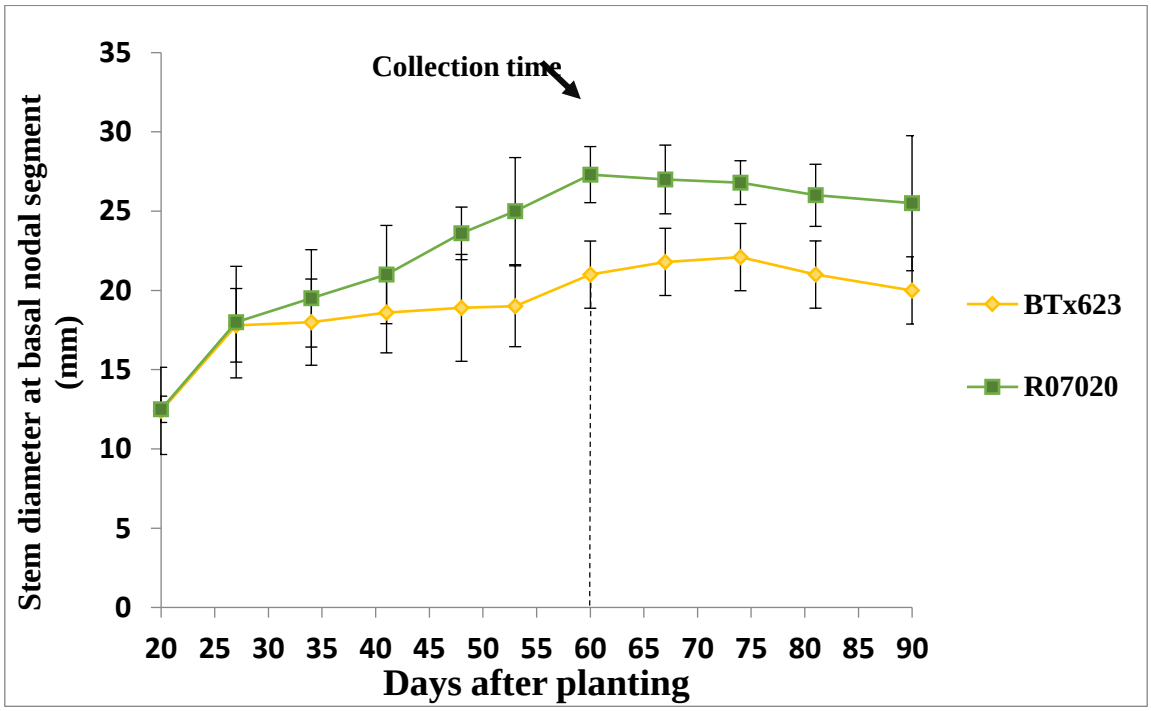


Figure 1: Continued

Histochemical analysis of sorghum stem nodal segments

To ascertain the level of lignin deposition and thus secondary wall formation, nodal sections from along the stem of 60-day-old R07020 and BTx623 plants were histochemically stained for lignin (Figure 2). Differences in lignin deposition were based on the developmental stage of the nodal segments and between the sorghum genotypes. Histochemical staining showed cells in mature basal nodal segments (N6, non-elongating region of the stem) contained a greater amount of lignin when compared with the upper immature nodal segments (N14 in R07020, N16 in BTx623) and nodal segments in the middle section of the stem (N10 in R07020, N11 in BTx623). In the upper immature nodal segments, lignin was mostly deposited in xylem in both R07020 and BTx623, while in the basal part of the stem, lignified sclerenchyma surrounded the entire vascular bundle. The lignified sclerenchyma cells were more pronounced in basal nodal segments of R07020 compared with BTx623.

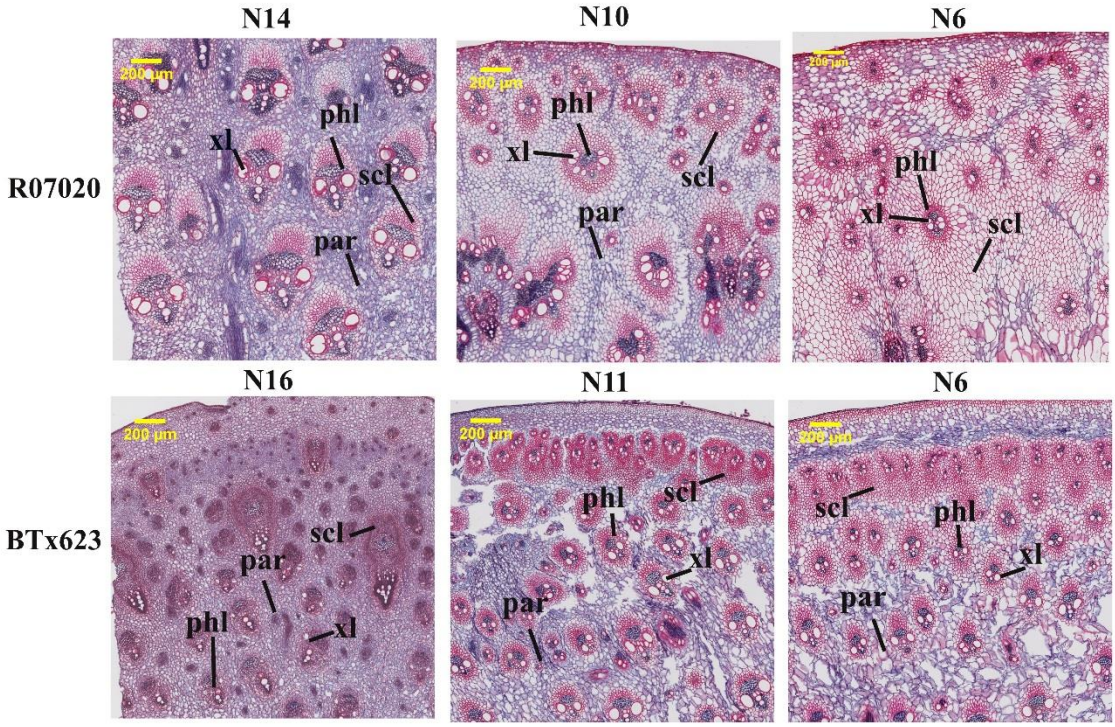


Figure 2: Cross sections of nodal segments of sorghum genotypes R07020 and BTx623. Fifteen µM cross-sections of nodal segments were stained with Alcian Blue and Safranin O. Red/pink coloration indicates the presence of lignin. xl, xylem; par, parenchyma; phl, phloem; scl, sclerenchyma.

Compositional analyses of sorghum stem nodal segments

The structural analysis of nodal segments along the developing stem of genotypes R07020 and BTx623 is shown in Table 1. Structural carbohydrates including cellulose, xylan, glucan and lignin showed significant differences in comparison of top and basal nodal segments. Cellulose content was higher in basal nodal segments in both genotypes In R07020, cellulose content ranged from 2.1% in the top nodal segment to 27% in the basal nodal segment. In BTx623, cellulose content ranged from 5.1% in the top nodal segment to 25.3% in the basal nodal segment. Xylan (one of the major hemicelluloses in secondary cell walls) content was also higher in basal nodal segments in both genotypes. Lignin content ranged from 6.04% in the top nodal segment to 11.16% in the basal nodal segment in R07020. Similarly, in genotype BTx623 lignin content ranged from 7.5% in the top nodal segment to 11.6% in the basal nodal segment. Protein content was also significantly different between the top and basal nodal segments in both R07020 and BTx623. Protein content in the top nodal segments was 16.86 and 16.34% in R07020 and BTx623, respectively, compared to 6.4 and 7.37% in the basal nodal segments. None of these components showed significant differences when comparing middle versus basal stem nodal segments, and this likely reflects that these segments are more similar in their developmental stage in comparison to basal versus upper nodal segments. In general, the trends for structural carbohydrates between nodal segments was very similar in both genotypes.

Table 1: Stem compositional analysis of sorghum genotypes R07020 and BTx623. The data are percentage means of two biological replicates taken at day 60 of growth. Different letters following means ± sd indicate statistically significant differences at *p*-value <0.05 in the designated nodal segments (Students *t*-test).

Genotype	Stem component	Nodal segment	Mean ± sd (%)
R07020	Cellulose	N14	2.10 ± 0.55 b
		N10	28.73 ± 1.9 a
		N6	26.95± 0.93 a
	Xylan	N14	7.20 ± 0.65 b
		N10	16.08 ± 0.48 a
		N6	15.45± 0.28 a
	Glucan	N14	19.3 ± 0.54 b
		N10	26.86 ± 1.7 a
		N6	28.70± 0.34 a
	Lignin	N14	6.04 ± 0.22 b
		N10	10.52 ± 0.02 a
		N6	11.16 ± 0.28 a
	Protein	N14	16.86 ± 0.75 a
		N10	7.39 ± 0.19 a
		N6	6.47 ± 1.27 b
BTx623	Cellulose	N16	5.19 ± 3.1 b
		N11	24.62 ± 0.7 a
		N6	25.32 ± 1.19 a
	Xylan	N16	8.20 ± 1.72 b
		N11	13.94 ± 0.39 a
		N6	14.38 ± 0.79 a
	Glucan	N16	20.38 ± 0.59 b
		N11	30.68 ± 0.79 a
		N6	31.17 ± 0.28 a
	Lignin	N16	7.50 ± 1.2 b
		N11	10.94 ± 0.26 a
		N6	11.60 ± 0.87 a
	Protein	N16	16.34 ± 0.15 a
		N11	7.70 ± 0.24 a
		N6	7.37 ± 0.4 b

RNA-seq transcriptome analysis

For each genotype, R07020 and BTx623, cDNA libraries from three nodal segments and three biological replications/nodal segment were prepared for a total of 18 cDNA libraries. Following construction of Illumina TruSeq RNA-seq libraries and sequencing on an Illumina HiSeq2500 the reads for each sample were mapped to the *Sorghum bicolor* genome (Sbicolor_255 v2.1) within the CLC Genomics Workbench and the details of the mapped reads for each stem nodal segment are shown in Table 2. The number of reads per library ranged from 53.3 M to 85.7 M in genotype R07020 and from 58.3 M to 116 M in genotype BTx623. Mapped reads for each RNA-seq library ranged from 47-106 M with an average of 92% of those reads mapping to gene exons. The reproducibility of data between biological replications was evaluated by Pearson Correlation Coefficient (PCC). The average PCC value for the three biological replicates was 0.92 and 0.90 for genotypes R07020 and BTx623, respectively (Table 3). DEGs were identified from the three biological replicates from each nodal segment using the Exact Test within the CLC Genomics Workbench. This test is similar to Fisher's Exact Test but accounts for the overdispersion caused by biological variability by replacing the Hypergeometirc distributions of Fisher's Exact Test by Negative binomial distributions (Robinson and Smyth, 2008).

The stems of monocots represent a developmental gradient with the more mature (non-elongating) regions residing at the base, and progressively younger tissues moving up the stem to the shoot apical meristem. This developmental system afforded the opportunity to examine differentially expressed genes in mature nodal segments that are

laying down lignified secondary cell walls and in more immature nodal segments where rapid primary cell wall biosynthesis is occurring. In this study, the characterization of DEGs in R07020 and BTx623 focused on long-day photoperiod growth conditions; conditions in which both genotypes are normally cultivated in temperate zones (Smith and Frederiksen, 2000).

In comparing the expression of genes in the basal nodal segment vs. the top nodal segment, a total of 5,886 genes in genotype R07020 were differentially expressed in nodal segment N6 vs. N14 (data not shown), and 4,640 genes in genotype BTx623 were differentially expressed in nodal segment N6 vs. N16 (data not shown). Of the differentially expressed genes in N6 vs. N14 in R07020, 3,281 (55%) were up-regulated in basal nodal segment N6 while the remaining 2,605 DEGs were up-regulated in N14. In BTx623, of the differentially expressed genes in nodal segment N6 vs. N16, 1,624 (35%) were up-regulated in N6 while the remaining 3,016 DEGs were up-regulated in N16. The higher number of up-regulated genes in the upper nodal segment of BTx623 may relate to the coincidence of panicle development and the transition to the reproductive phase at the apex of the stem in this genotype. At the stage of development when the samples were collected (i.e. 60 days after planting), rapid stem elongation is reduced and the peduncle begins to elongate (Vanderlip and Reeves, 1972). Fewer DEGs were identified when comparing the basal nodal segment to the middle nodal segment in both genotypes. Only 397 and 230 DEGs were identified between N6 and N10 in R07020 or between N6 and N11 in BTx623, respectively (data not shown). As basal and mid-stem nodal regions are both actively depositing secondary cell walls, fewer DEGs were expected. The DEGs that

were detected were probably related to the fact that nodal segment N6 is no longer elongating while N10 continues to elongate, which indicates that both primary and secondary cell wall biosynthesis were still occurring in N10. This is consistent with the hierarchical clustering of DEGs between basal and middle nodal segments indicating that there is an association of DEGs in N6 and N10 when contrasted to DEGs in N6 vs. N14 (Figure 3). The same clustering pattern was observed for basal and middle nodal segments in BTx623 (data not shown). Based on the results from hierarchical clustering and the fact that so few DEGs were detected when comparing basal and middle nodal segments all further analysis and comparisons were performed with top and basal nodal segments only.

Table 2: Summary of RNA-seq reads from stem nodal segments of sorghum genotypes R07020 and BTx623 mapped to the annotated genome of sorghum (Sbicolor_255 v.2.1). Unique RNA-seq reads mapping to exons, introns, and intergenic regions are shown as the percentage of total reads distributed to these annotated regions of the sorghum genome.

Genotype	Nodal Segment	Rep number	Total reads (M)	Mapped reads (M)	Exonic reads (%)	Intronic reads (%)	Intergenic reads (%)
R07020	14	1	79.4	69.68	92.46	4.94	2.6
		2	53.3	47.22	93.57	4.21	2.22
		3	75.00	68.61	90.38	6.42	3.2
Mean ± sd			69.27 ± 13.97	61.84 ± 12.66	92.13 ± 1.61	5.19 ± 1.12	2.67 ± 0.49
R07020	10	1	83.10	71.13	92.86	4.63	2.51
		2	67.86	60.59	93.56	4.17	2.27
		3	64.03	58.66	91.72	5.44	2.84
Mean ± sd			71.66 ± 10.08	63.46 ± 6.71	92.71 ± 0.92	4.74 ± 0.64	2.54 ± 0.28
R07020	6	1	79.71	71.21	93.90	3.91	2.19
		2	62.12	52.83	93.46	4.20	2.34
		3	85.72	75.22	92.58	4.75	2.67
Mean ± sd			75.85 ± 12.26	66.42 ± 11.93	93.31 ± 0.67	4.28 ± 0.42	2.4 ± 0.24
BTx623	16	1	81.32	71.62	92.20	5.14	2.66
		2	63.42	57.03	92.71	4.95	2.34
		3	61.25	56.48	91.91	5.39	2.71
Mean ± sd			68.66 ± 11.01	61.71 ± 8.58	92.27 ± 0.40	5.16 ± 0.22	2.57 ± 0.20

Table 2: Continued

Genotype	Nodal Segment	Rep number	Total reads (M)	Mapped reads (M)	Exonic reads (%)	Intronic reads (%)	Intergenic reads (%)
BTx623	11	1	82.08	73.62	91.62	5.42	2.96
		2	58.36	52.31	92.8	4.74	2.45
		3	62.83	48.48	90.39	5.68	3.93
Mean ± sd			67.76 ± 12.60	58.14 ± 13.54	91.6 ± 1.20	5.28 ± 0.48	3.11 ± 0.75
BTx623	6	1	80.05	72.07	92.38	4.93	2.69
		2	71.16	68.99	92.79	4.68	2.53
		3	116.08	106.24	91.23	5.73	3.03
Mean ± sd			90.43 ± 23.78	82.44 ± 20.67	92.13 ± 0.80	5.11 ± 0.54	2.75 ± 0.25

Table 3: Pearson's correlation of RNA-seq reads from stem nodal segments of R07020 and BTx623 genotypes from three independent biological experiments. Pair-wise Pearson's Correlation Coefficients were calculated from the gene expression values of all annotated genes of the sorghum genome (Sbicolor_255 v.2.1).

Genotype	Nodal segment	Rep1 vs. Rep2	Rep1 vs. Rep3	Rep2 vs. Rep3
R07020	N14	0.92	0.94	0.88
	N10	0.87	0.96	0.88
	N6	0.94	0.95	0.96
BTx623	N16	0.93	0.93	0.85
	N11	0.91	0.85	0.93
	N6	0.92	0.88	0.93

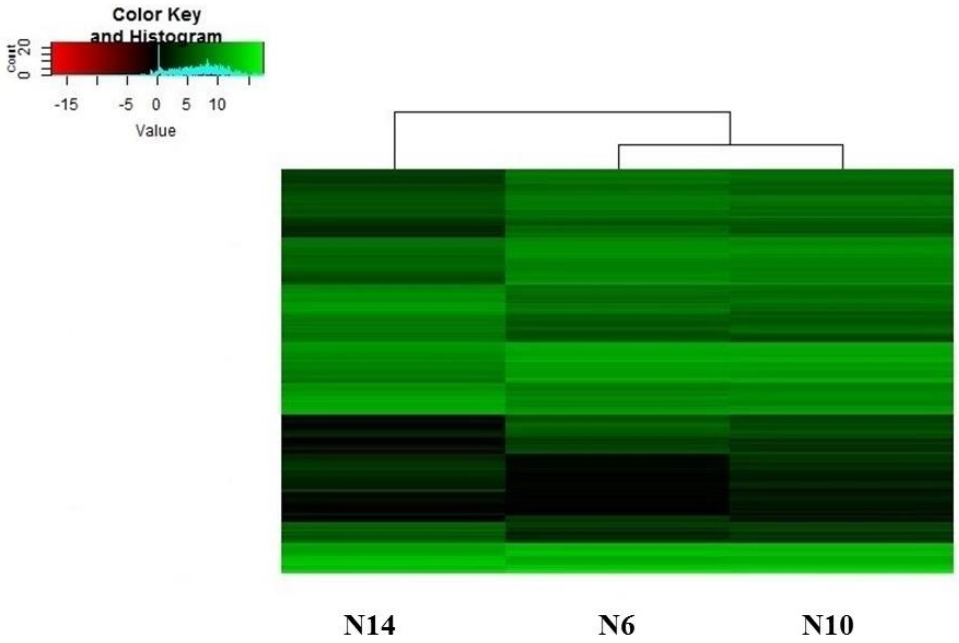


Figure 3: Hierarchical clustering of sorghum genotype R07020 nodal segments based on differentially expressed genes. The color key represents total counts normalized to $\log_2$ total counts. The hierarchical clusters were obtained based on Euclidian distance.

Gene ontology analysis

Based on the number of DEGs between the basal nodal segments and nodal segments near the apical meristem, gene ontology (GO) analysis was conducted. In R07020 and BTx623, a series of GO terms were enriched in basal nodes and nodes near the apical meristem, and these GO terms included numerous molecular functions and biological processes. The GO term carbohydrate metabolic process represented the highest number of enriched biological process GO terms in rapidly developing nodal segments near the stem apex in genotypes R07020 (N14) and BTx623 (N16) (Tables 4 and 5). Metabolism of carbohydrates includes the process where complex carbohydrates are broken down into simple mono- and disaccharides like glucose and sucrose, which are important components of metabolic pathways including primary cell wall synthesis that is occurring in rapidly elongating nodal regions. Bosch, et al. (2011) reported the preferential expression of genes involved in carbohydrate metabolism in the elongating internode of maize. In basal nodal segment N6 from both R07020 and BTx623, protein modification and post-translational protein modification were two prominent enriched GO terms in the biological process category (Tables 6 and 7). These modifications might be involved in transcriptional regulation of the monolignol biosynthesis pathway that provides source material for the biosynthesis of lignin (Jin, et al., 2000; Adams-Phillips, et al., 2010).

Table 4: Enriched GO terms associated with DEGs in nodal segment N14 from sorghum genotype R07020. The significant GO terms (FDR corrected *p*-value <0.05) were determined by contrasting expression of genes in nodal segment N14 to their expression in N6. DEGs were then grouped into molecular function and biological process categories. *P*-values indicate the statistical significance of differential expression observed between nodal segments N14 and N6 for genes associated with each GO term. Qnum are the number of probe sets that belong to the GO term from the query list (DEGs). B/Rnum values are the number of probe sets that belong to the GO term from the background based on a genome-wide set of genes.

GO term	GO type	GO name	Qnum	B/Rnum	FDR *p*-value
GO:0005975	Biological process	carbohydrate metabolic process	113	777	0.00
GO:0006259	Biological process	DNA metabolic process	43	239	0.01
GO:0007018	Biological process	microtubule-based movement	32	51	0.00
GO:0007017	Biological process	microtubule-based process	32	65	0.00
GO:0006260	Biological process	DNA replication	22	61	0.00
GO:0015979	Biological process	photosynthesis	20	81	0.02
GO:0043086	Biological process	negative regulation of catalytic activity	15	55	0.05
GO:0044092	Biological process	negative regulation of molecular function	15	55	0.05
GO:0015995	Biological process	chlorophyll biosynthetic process	9	12	0.00
GO:0015994	Biological process	chlorophyll metabolic process	9	15	0.01
GO:0016787	Molecular function	hydrolase activity	304	2358	0.00
GO:0030554	Molecular function	adenyl nucleotide binding	274	2531	0.03
GO:0001882	Molecular function	nucleoside binding	274	2532	0.03
GO:0001883	Molecular function	purine nucleoside binding	274	2531	0.03
GO:0032559	Molecular function	adenyl ribonucleotide binding	262	2392	0.02
GO:0005524	Molecular function	ATP binding	233	2074	0.02
GO:0016817	Molecular function	hydrolase activity, acting on acid anhydrides	95	627	0.00
GO:0016818	Molecular function	hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides	95	620	0.00
GO:0017111	Molecular function	nucleoside-triphosphatase activity	86	591	0.00

Table 4: Continued

GO term	GO type	GO name	Qnum	B/Rnum	FDR *p*-value
GO:0016462	Molecular function	pyrophosphatase activity	86	607	0.00
GO:0016798	Molecular function	hydrolase activity, acting on glycosyl bonds	75	402	0.00
GO:0016788	Molecular function	hydrolase activity, acting on ester bonds	74	558	0.05
GO:0004553	Molecular function	hydrolase activity, hydrolyzing O-glycosyl compounds	69	374	0.00
GO:0003774	Molecular function	motor activity	33	65	0.00
GO:0003777	Molecular function	microtubule motor activity	32	51	0.00
GO:0005507	Molecular function	copper ion binding	24	121	0.03
GO:0042802	Molecular function	identical protein binding	16	63	0.03
GO:0051002	Molecular function	ligase activity, forming nitrogen-metal bonds	9	12	0.00
GO:0051003	Molecular function	ligase activity, forming nitrogen-metal bonds, forming coordination complexes	9	12	0.00
GO:0016851	Molecular function	magnesium chelatase activity	9	12	0.00
GO:0005088	Molecular function	Ras guanyl-nucleotide exchange factor activity	7	11	0.02
GO:0005089	Molecular function	Rho guanyl-nucleotide exchange factor activity	7	11	0.02

Table 5: Enriched GO terms associated with DEGs in nodal segment N16 from sorghum genotype BTx623. The significant GO terms (FDR corrected *p*-value <0.05) were determined by contrasting expression of genes in nodal segment N16 to their expression in N6. Descriptors and other details are as described for Table 3.

GO term	GO type	GO name	Qnum	B/Rnum	FDR *p*-value
GO:0005975	Biological process	carbohydrate metabolic process	153	777	0.00
GO:0007017	Biological process	microtubule-based process	39	65	0.00
GO:0007018	Biological process	microtubule-based movement	36	51	0.00
GO:0006073	Biological process	cellular glucan metabolic process	24	86	0.01
GO:0044042	Biological process	glucan metabolic process	24	86	0.01
GO:0030312	Cellular component	external encapsulating structure	30	106	0.00
GO:0005618	Cellular component	cell wall	28	83	0.00
GO:0005576	Cellular component	extracellular region	16	58	0.04
GO:0048046	Cellular component	apoplast	12	32	0.03
GO:0003824	Molecular function	catalytic activity	903	7817	0.00
GO:0016787	Molecular function	hydrolase activity	360	2358	0.00
GO:0016798	Molecular function	hydrolase activity, acting on glycosyl bonds	110	402	0.00
GO:0004553	Molecular function	hydrolase activity, hydrolyzing O-glycosyl compounds	103	374	0.00
GO:0005507	Molecular function	copper ion binding	43	121	0.00
GO:0003774	Molecular function	motor activity	37	65	0.00
GO:0003777	Molecular function	microtubule motor activity	36	51	0.00

Table 6: Enriched GO terms associated with DEGs in nodal segment N6 from sorghum genotype R07020. The significant GO terms (FDR corrected *p*-value <0.05) were determined by contrasting expression of genes in nodal segment N6 to their expression in N14. Descriptors and other details are as described for Table 4.

GO term	GO type	GO name	Qnum	B/Rnum	FDR *p*-value
GO:0008152	Biological process	metabolic process	1160	8076	0.00
GO:0055114	Biological process	oxidation reduction	280	1485	0.00
GO:0006464	Biological process	protein modification process	274	1505	0.00
GO:0043412	Biological process	macromolecule modification	274	1553	0.00
GO:0043687	Biological process	post-translational protein modification	267	1402	0.00
GO:0006796	Biological process	phosphate metabolic process	262	1384	0.00
GO:0006793	Biological process	phosphorus metabolic process	262	1384	0.00
GO:0016310	Biological process	phosphorylation	258	1343	0.00
GO:0006468	Biological process	protein amino acid phosphorylation	253	1265	0.00
GO:0050794	Biological process	regulation of cellular process	221	1402	0.04
GO:0003824	Molecular function	catalytic activity	1145	7817	0.00
GO:0016740	Molecular function	transferase activity	476	2873	0.00
GO:0001883	Molecular function	purine nucleoside binding	390	2531	0.00
GO:0001882	Molecular function	nucleoside binding	390	2532	0.00
GO:0030554	Molecular function	adenyl nucleotide binding	390	2531	0.00
GO:0032559	Molecular function	adenyl ribonucleotide binding	369	2392	0.00
GO:0016491	Molecular function	oxidoreductase activity	312	1669	0.00
GO:0046872	Molecular function	metal ion binding	279	1712	0.00
GO:0043169	Molecular function	cation binding	279	1720	0.00
GO:0043167	Molecular function	ion binding	279	1724	0.00
GO:0016772	Molecular function	transferase activity, transferring phosphorus-containing groups	266	1633	0.00
GO:0016301	Molecular function	kinase activity	262	1420	0.00

Table 6: Continued

GO term	GO type	GO name	Qnum	B/Rnum	FDR *p*-value
GO:0016773	Molecular function	phosphotransferase activity, alcohol group as acceptor	260	1411	0.00
GO:0004672	Molecular function	protein kinase activity	256	1278	0.00
GO:0046914	Molecular function	transition metal ion binding	228	1315	0.00
GO:0016705	Molecular function	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen	160	529	0.00
GO:0005506	Molecular function	iron ion binding	159	588	0.00
GO:0009055	Molecular function	electron carrier activity	157	576	0.00
GO:0020037	Molecular function	heme binding	153	534	0.00
GO:0046906	Molecular function	tetrapyrrole binding	153	545	0.00
GO:0030528	Molecular function	transcription regulator activity	121	632	0.00
GO:0003700	Molecular function	transcription factor activity	118	573	0.00
GO:0016757	Molecular function	transferase activity, transferring glycosyl groups	115	529	0.00
GO:0016758	Molecular function	transferase activity, transferring hexosyl groups	98	427	0.00
GO:0043565	Molecular function	sequence-specific DNA binding	76	376	0.01
GO:0043531	Molecular function	ADP binding	66	326	0.02
GO:0016209	Molecular function	antioxidant activity	44	202	0.04
GO:0016684	Molecular function	oxidoreductase activity, acting on peroxide as acceptor	42	177	0.01
GO:0004601	Molecular function	peroxidase activity	42	177	0.01
GO:0016706	Molecular function	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, 2-oxoglutarate as one donor, and incorporation of one atom each of oxygen into both donors	40	122	0.00

Table 6: Continued

GO term	GO type	GO name	Qnum	B/Rnum	FDR *p*-value
GO:0031323	Biological process	regulation of cellular metabolic process	190	1051	0.00
GO:0019222	Biological process	regulation of metabolic process	190	1066	0.00
GO:0006350	Biological process	transcription	190	1120	0.00
GO:0006351	Biological process	transcription, DNA-dependent	190	1120	0.00
GO:0032774	Biological process	RNA biosynthetic process	190	1122	0.00
GO:0006355	Biological process	regulation of transcription, DNA-dependent	186	994	0.00
GO:0051252	Biological process	regulation of RNA metabolic process	186	994	0.00
GO:0045449	Biological process	regulation of transcription	186	996	0.00
GO:0019219	Biological process	regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	186	1004	0.00
GO:0051171	Biological process	regulation of nitrogen compound metabolic process	186	1004	0.00
GO:0009889	Biological process	regulation of biosynthetic process	186	1019	0.00
GO:0010556	Biological process	regulation of macromolecule biosynthetic process	186	1019	0.00
GO:0031326	Biological process	regulation of cellular biosynthetic process	186	1019	0.00
GO:0080090	Biological process	regulation of primary metabolic process	186	1023	0.00
GO:0010468	Biological process	regulation of gene expression	186	1028	0.00
GO:0060255	Biological process	regulation of macromolecule metabolic process	186	1035	0.00
GO:0055085	Biological process	transmembrane transport	116	646	0.02
GO:0050896	Biological process	response to stimulus	106	592	0.03
GO:0042221	Biological process	response to chemical stimulus	54	259	0.03
GO:0009607	Biological process	response to biotic stimulus	12	24	0.02

Table 7: Enriched GO terms associated with DEGs in nodal segment N6 from sorghum genotype BTx623. The significant GO terms (FDR corrected *p*-value <0.05) were determined by contrasting expression of genes in nodal segment N6 to their expression in N16. Descriptors and other details are as described for Table 4.

GO term	GO type	GO name	Qnum	B/Rnum	FDR *p*-value
GO:0008152	Biological process	metabolic process	622	8076	0.00
GO:0009987	Biological process	cellular process	499	6324	0.00
GO:0044238	Biological process	primary metabolic process	463	5708	0.00
GO:0044237	Biological process	cellular metabolic process	412	4912	0.00
GO:0043170	Biological process	macromolecule metabolic process	401	4419	0.00
GO:0044260	Biological process	cellular macromolecule metabolic process	368	3915	0.00
GO:0019538	Biological process	protein metabolic process	284	2674	0.00
GO:0044267	Biological process	cellular protein metabolic process	257	2214	0.00
GO:0006464	Biological process	protein modification process	248	1505	0.00
GO:0043412	Biological process	macromolecule modification	248	1553	0.00
GO:0043687	Biological process	post-translational protein modification	245	1402	0.00
GO:0006796	Biological process	phosphate metabolic process	241	1384	0.00
GO:0006793	Biological process	phosphorus metabolic process	241	1384	0.00
GO:0006468	Biological process	protein amino acid phosphorylation	237	1265	0.00
GO:0016310	Biological process	phosphorylation	237	1343	0.00
GO:0019219	Biological process	regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	95	1004	0.04
GO:0051171	Biological process	regulation of nitrogen compound metabolic process	95	1004	0.04

Pathway enrichment analysis of DEGs

A survey of metabolic pathways enriched with DEGs was obtained using KOBAS in conjunction with the KEGG PATHWAY database (Xie, et al., 2011). In both BTx623 and R07020, nearly 120 KEGG pathways were assigned to DEGs that were identified in mature nodal segments (N6) versus immature nodal segments near the apical meristem (N14, N16). A survey of DEG-enriched pathways in both BTx623 and R07020 revealed a series of highly-enriched inter-related biosynthetic pathways including phenylpropanoid metabolism and flavonoid biosynthesis, secondary metabolites, phenylalanine metabolism, and plant-pathogen interactions (Tables 8 and 9). There are a diverse group of compounds produced from the phenylpropanoid pathway that originates from the carbon skeleton of phenylalanine that are involved in structural/mechanical support and abiotic/biotic stress tolerance (Vogt, 2010). Lignin is one of the major products of the phenylpropanoid pathway (Anderson and Chapple, 2014), and enrichment of the phenylpropanoid pathway in N6 is consistent with active secondary cell wall synthesis and lignification that is occurring in this region of the stem. Other important secondary metabolites that are derived from the phenylpropanoid pathway include flavonoids, which play a key role in plant responses to both biotic and abiotic stress (Zhang, et al., 2016). Since flavonoids are synthesized by the phenylpropanoid pathway (Gray, et al., 2012), it is consistent that enrichment of DEGs in both phenylpropanoid metabolism and flavonoid biosynthesis was observed concomitantly.

Similar to the results obtained through the analysis of gene ontology, the pathway enrichment analysis of immature nodal segments near the apical meristem in R07020,

revealed a series of DEG-enriched pathways involved in the metabolic activities of actively elongating regions of the stem. Pathways that were enriched in nodal segment N14 of R07020 included starch and sucrose metabolism, DNA replication, photosynthesis-antenna proteins, and biosynthetic pathways for cutin, suberin, and wax (Table 10). The three major pathways enriched in nodal segment N16 in BTx623 were starch and sucrose metabolism, biosynthesis of cutin, suberin and wax, and phenylpropanoid biosynthesis (Table 11). In the stems of sweet sorghum cultivar Della prior to anthesis, McKinley, et al. (2016) reported elevated expression of sucrose metabolism genes that included sucrose synthase 4 (*SUS4*) and five cytosolic and vacuolar invertases from floral initiation until seven days before anthesis. These genes were down-regulated after anthesis which is consistent with the maturation of stems and the deposition of secondary cell walls (McKinley, et al., 2016). In the present study, the upregulation of biosynthetic pathway genes in immature nodal segments is consistent with the rapid deposition of polyesters and waxes in elongating stem segments that allow for the fast rate of stem elongation with a constant load and composition of the protective cuticle (Suh, et al., 2005). Cutin is embedded with waxes in the cuticle of plants, which usually functions as an amorphous protective layer covering leaves and stems of plants. Suberin also shapes essential physical and biological barriers for plants as a complex hydrophobic material, and is deposited on the walls around the bundle sheath cells in grasses (Albersheim, et al., 2010).

Table 8: Pathways enriched in DEGs based on the KOBAS web server in nodal segment N6 from sorghum genotype R07020. Input number indicates the total number of input genes mapped to the particular pathway while background indicates the number of baseline expressed genes mapped to the particular pathway based on the KEGG pathway database. Significant enriched pathways were selected based on an FDR corrected *p*-value <0.05.

Pathway	Database	Input number	Background number	FDR *p*-value
Biosynthesis of secondary metabolites	KEGG PATHWAY	147	1015	0.00
Phenylpropanoid biosynthesis	KEGG PATHWAY	58	240	0.00
Phenylalanine metabolism	KEGG PATHWAY	42	184	0.00
Flavonoid biosynthesis	KEGG PATHWAY	16	55	0.00
Cyanoamino acid metabolism	KEGG PATHWAY	15	59	0.01
Stilbenoid, diarylheptanoid and gingerol biosynthesis	KEGG PATHWAY	9	21	0.01

Table 9: Pathways enriched in DEGs based on KOBAS in nodal segment N6 from sorghum genotype BTx623. Descriptors and other details are as described for Table 8.

Pathway	Database	Input number	Background number	FDR *p*-value
Biosynthesis of secondary metabolites	KEGG PATHWAY	74	1015	0.00
Plant-pathogen interaction	KEGG PATHWAY	24	170	0.00
Phenylpropanoid biosynthesis	KEGG PATHWAY	24	240	0.00
Phenylalanine metabolism	KEGG PATHWAY	21	184	0.00

Table 10: Pathways enriched in DEGs based on KOBAS in nodal segment N14 from sorghum genotype R07020. Descriptors and other details are as described for Table 8.

Pathway	Database	Input number	Background number	FDR *p*-value
Starch and sucrose metabolism	KEGG PATHWAY	28	180	0.03
DNA replication	KEGG PATHWAY	16	73	0.02
Cutin, suberin and wax biosynthesis	KEGG PATHWAY	11	34	0.01
Photosynthesis - antenna proteins	KEGG PATHWAY	9	15	0.00

Table 11: Pathways enriched in DEGs based on KOBAS in nodal segment N16 from sorghum genotype BTx623. Descriptors and other details are as described for Table 8.

Pathway	Database	Input number	Background number	FDR *p*-value
Phenylpropanoid biosynthesis	KEGG PATHWAY	42	247	0.01
Starch and sucrose metabolism	KEGG PATHWAY	35	189	0.01
Cutin, suberin and wax biosynthesis	KEGG PATHWAY	13	40	0.01

DEGs involved in the phenylpropanoid pathway

An examination of DEGs in nodal segments N6 vs. N14 of genotype R07020 revealed 90 gene members of the phenylpropanoid biosynthesis pathway with multiple enzymatic steps of this pathway being differentially expressed (Table 12). The majority of these genes (66%) were up-regulated in nodal segment N6, which included 14 genes involved in both lignin monomers and flavonoid metabolism. Thirty-one genes in the phenylpropanoid pathway were also up-regulated in nodal segment N14. Among the DEGs in genotype BTx623, 66 genes mapped to the phenylpropanoid biosynthesis pathway with a majority (64%) up-regulated in the upper nodal segment N16 vs. nodal segment N6, while 24 genes in this pathway were up-regulated in N6 vs. N16 (Table 13). The two main enzymatic steps in lignin biosynthesis are the biosynthesis of monolignols and polymerization of monolignols (Boerjan, et al., 2003), and various genes involved in these processes were differentially expressed in basal nodal segments. *Sobic.004G220400*, encoding a *PAL* gene, which deaminates phenylalanine to produce cinnamic acid in the first step of the phenylpropanoid pathway, showed a 2.6-fold increase in expression in nodal segment N6 of genotype R07020, and another *PAL* gene, *Sobic.006G148900*, showed a 5.1-fold expression increase in the basal nodal segment of genotype BTx623. The 4-coumarate-CoA ligase (4CL) enzyme catalyzes the conversion of p-coumaric and caffeic acids to their thio ester forms, which are the precursors of monolignol (Saballos, et al., 2012). The *4CL* gene family members regulate the synthesis of monolignol precursors and therefore, control lignin content and composition (Saballos, et al., 2012). In R07020, three *4CL* gene family members, *Sobic.007G145600* (4.3-fold change),

Sobic.004G062500 (3.6-fold change), and *Sobic.007G089900* (2.3-fold change*)*, were up-regulated in basal segment N6 vs. N14. Three other *4CL* gene family members, including *Sobic.001G187000 (*11.3-fold change), *Sobic.004G272700* (3.3-fold change), and *Sobic.007G145600* (3.6-fold change) were up-regulated in N6 of BTx623 versus N16. Two members of the *F5H* gene family, *Sobic.005G088400* and *Sobic.001G196300*, that are involved in S-lignin biosynthesis, showed 26.5- and 2.9-fold increases in gene expression, respectively, in basal nodal segment N6 versus apical nodal segment N14 in R07020. Lignin that is rich in G units has more carbon-carbon bonds than lignin rich in S units, therefore, it is more resistant to chemical degradation (Lapierre, et al., 1999; Boerjan, et al., 2003). The down-regulation of *F5H* results in lignin composed of G units, whereas *F5H* overexpression can result in plants with lignins almost entirely composed of S units (Boerjan, et al., 2003). Therefore, the over-expression of *F5H* has been proposed as a strategy to improve the saccharification process of lignocellulosic biofuel production (Poovaiah, et al., 2014). The *COMT* (*Sobic.007G059100*, 75.1-fold change and *Sobic.007G058800*, 13.2-fold change) and *CCOMT* (*Sobic.002G242300*, 15.5-fold change and *Sobic.010G052200,* 2.5-fold change) genes are involved in producing sinapyl alcohol (Ralph, et al., 2001), and these genes showed increased expression in nodal segment N6 compared to N14 in R07020 (Table 12). In BTx623, *CCOMT* gene *Sobic.002G242300* also showed an ~8-fold increase in expression in nodal segment N6 compared to N16 (Table 13).

Mutations in the last steps of lignin biosynthesis including *COMT* and *CAD* reduce the lignin content in various grasses (Jones, et al., 2001; Yoon, et al., 2015). In the present

study, *CAD* gene *Sobic.006G014700* in genotype R07020 and *CAD* gene *Sobic.002G195600* in BTx623 showed >4-fold increases in expression in basal nodal segments, which are actively involved in lignin biosynthesis (Tables 12 and 13). These observations are consistent with reports of reduced lignin deposition associated with *COMT* mutants in switchgrass (Fu, et al., 2011) and sugarcane (Jung, et al., 2012). The *COMT* mutant *Bd5139* in *Brachypodium* (Ho-Yue-Kuang, et al., 2015) also altered stem and grain lignin and improved saccharification of biomass without a reduction in grain quality. In Della sweet sorghum, McKinley, et al. (2016) also found higher expression of similar genes including *4CL* (*Sobic.004G062500*), *CCOMT* (*Sobic.010G052200*), *CCR* (*Sobic.007G141200*), *CAD* (*Sobic.006G014700*) and *F5H* (*Sobic.001G196300*). In Della sweet sorghum, the peak expression of these lignin biosynthesis genes was coincident with the up-regulation of other secondary cell wall biosynthesis genes that occurred prior to anthesis.

In both genotypes examined here, genes encoding peroxidases and β-glucosidases were also differentially expressed, but the majority of peroxidase gene family members were up-regulated in the basal nodal segment of R07020 whereas in BTx623, these genes were up-regulated in the apical nodal segment N16 (data not shown). Peroxidases are involved in lignin polymerization that results from the oxidative coupling of monolignol to the growing polymer. β-glucosidase genes in general showed higher expression in the apical nodal segment of BTx623 whereas in R07020 a similar number of β-glucosidase genes were up-regulated in the basal and the apical nodal segments (data not shown). β-glucosidases play important roles in the formation of intermediates in cell wall

lignification and belong to the family 1 glycoside hydrolases that catalyze the hydrolysis of the β-glucosidic bond between two carbohydrate moieties or a carbohydrate and a glucose moiety (Escamilla-Treviño, et al., 2006). The observed differences in the two sorghum genotypes likely reflects the different developmental stages that existed after 60 days of growth in the two genotypes. In BTx623, floral initiation had already occurred and the panicle was in the boot stage. In this stage of reproductive development, rapid stem elongation has ceased and peduncle elongation has commenced (Vanderlip and Reeves, 1972). By comparison, R07020 is photoperiod sensitive and thus, vegetative growth (including stem elongation) continued unabated at the time of tissue sampling. Since secondary wall thickening coincides with the cessation of cell expansion in secondary wall-forming cell types, it is reasonable to assume that the majority of phenylpropanoid pathway genes in general showed higher expression in BTx623 in the apical stem section that has ceased rapid cell elongation.

Table 12: Differentially expressed key phenylpropanoid pathway genes in sorghum genotype R07020. DEGs represent expression differences between nodal segments N6 and N14. DEGs were defined as having a $\log_2$ fold change ≥ 1 or ≤ -1 with an FDR corrected *p*-value <0.05.

Gene	Description	Fold change[a]	FDR *p*-value	Gene ortholog	
				Arabidopsis	**Maize**
Sobic.004G062500	4-*coumarate--CoA ligase* (*4CL*)	3.60	0.00	*AT3G21240*	*GRMZM2G055320*
Sobic.007G089900	*4-coumarate--CoA ligase* (*4CL*)	2.35	0.02	*AT3G21240*	*GRMZM2G055320*
Sobic.007G145600	*4-coumarate--CoA ligase* (*4CL*)	4.31	0.00	*AT3G21240*	*GRMZM2G055320*
Sobic.007G059100	*caffeic acid 3-O-methyltransferase* (*COMT*)	75.14	0.00	*AT5G54160*	*AC196475.3_FG004*
Sobic.007G058800	*caffeic acid 3-O-methyltransferase* (*COMT*)	13.29	0.00	*AT5G54160*	*AC196475.3_FG004*
Sobic.002G242300	*caffeoyl-CoA O-methyltransferase* (*CCOMT*)	15.58	0.00	*AT4G34050*	*GRMZM2G127948*
Sobic.010G052200	*caffeoyl-CoA O-methyltransferase* (*CCOMT*)	2.48	0.00	*AT4G34050*	*GRMZM2G127948*
Sobic.002G146000	*cinnamoyl-CoA reductase* (Shen, et al.)	-2.39	0.00	*AT1G15950*	*GRMZM2G131205*
Sobic.007G141200	*cinnamoyl-CoA reductase* (Shen, et al.)	5.47	0.00	*AT1G15950*	*GRMZM2G131205*
Sobic.006G014700	*cinnamyl-alcohol dehydrogenase* (*CAD*)	5.72	0.00	*AT4G39330*	N/A
Sobic.010G178300	*coniferyl-aldehyde dehydroge*nase (*CALDH*)	-2.65	0.00	*AT3G24503*	*GRMZM2G071021*
Sobic.003G327800	c*oumaroylquinate*	-2.12	0.01	*AT2G40890*	*GRMZM2G140817*
Sobic.001G196300	*ferulate-5-hydroxylase* (*F5H*)	2.91	0.00	*AT4G36220*	*GRMZM2G100158*
Sobic.005G088400	*ferulate-5-hydroxylase* (*F5H*)	26.57	0.00	*AT4G36220*	*GRMZM2G100158*
Sobic.004G220400	*phenylalanine ammonia-lyase* (*PAL*)	2.61	0.00	*AT2G37040*	*GRMZM2G029048*
Sobic.004G220300	*phenylalanine/tyrosine ammonia-lyase* (*PAL/TAL*)	2.24	0.00	*AT3G53260*	*GRMZM2G029048*
Sobic.002G126600	*trans-cinnamate 4-monooxygenase* (*C4H*)	2.01	0.00	*AT2G30490*	*GRMZM2G010468*

[a]Positive values indicate greater expression in nodal segment N6 compared to N14. Negative values indicate greater expression in nodal segment N14 compared to N6.

Table 13: Differentially expressed key phenylpropanoid pathway genes in sorghum genotype BTx623. DEGs represent expression differences between nodal segments N6 and N16. DEGs were defined as having a $\log_2$ fold change ≥ 1 or ≤ -1 with an FDR corrected *p*-value <0.05.

Gene	Description and EC number	Fold change[a]	FDR *p*-value	Gene ortholog	
				Arabidopsis	**Maize**
Sobic.001G187000	*4-coumarate--CoA ligase* (*4CL*)	11.3	0.04	*AT4G05160*	*GRMZM2G122787*
Sobic.004G272700	*4-coumarate--CoA ligase* (*4CL*)	3.36	0.01	*AT1G65060*	*GRMZM2G054013*
Sobic.007G145600	*4-coumarate--CoA ligase* (*4CL*)	3.69	0.02	*AT3G21240*	*GRMZM2G055320*
Sobic.002G242300	*caffeoyl-CoA O-methyltransferase* (*CCOM*T)	8.14	0.00	*AT4G34050*	*GRMZM2G127948*
Sobic.002G146000	*cinnamoyl-CoA reductase* (Shen, et al.)	-8.34	0.00	*AT1G15950*	*GRMZM2G131205*
Sobic.002G195600	*cinnamyl-alcohol dehydrogenase* (*CAD*)	12.03	0.00	*AT4G39330*	N/A
Sobic.003G203600	*coniferyl-aldehyde dehydrogenase* (*CALDH*)	7.98	0.02	*AT3G24503*	*GRMZM2G071021*
Sobic.010G178300	*coniferyl-aldehyde dehydrogenase* (*CALDH*)	-26.91	0.00	*AT3G24503*	*GRMZM2G071021*
Sobic.006G148900	*phenylalanine ammonia-lyase* (*PAL*)	5.18	0.02	*AT2G37040*	*GRMZM2G029048*

[a]Positive values indicate greater expression in nodal segment N6 compared to N16. Negative values indicate greater expression in nodal segment N16 compared to N6.

Transcriptional regulation of cell wall biosynthesis

Previous studies have shown that regulation of cell wall biosynthesis can occur at the transcriptional level (Handakumbura and Hazen, 2007). In the present study, 329 and 176 DEGs were annotated as transcription factors (TFs) in R07020 and BTx623, respectively, including gene members of the *MYB* and *NAC* families (data not shown). Sorghum TFs of the *NAC* family that were differentially expressed included *Sobic.003G334600*, which was up-regulated 6.4- and 9-fold in basal nodal segments in genotypes R07020 and BTx623, respectively, and *Sobic.005G064600*, which was up-regulated 8.2- and 3-fold in basal nodal segments of R07020 and BTx623, respectively (Tables 14 and 15). *Sobic.003G334600* and *Sobic.005G064600* are orthologs of *NAC* TFs that have been reported to be up-regulated in non-elongating internodes of maize (Bosch, et al., 2011). Valdivia, et al. (2013) proposed that secondary wall-associated *NACs* (*SWNs*) function as the "master switches" for secondary cell wall biosynthesis. Christiansen, et al. (2011) also reported greater expression of three *NAC* genes in stem tissues of barley (*Hordeum vulgare* L.) that were actively laying down secondary cell walls. Several *MYB* transcription factors were up-regulated in the basal nodal segments of R07020 and BTx623 (Tables 14 and 15). For example, *MYB* TF (*Sobic.002G275500*) showed a 2-fold increase in the basal nodal segment N6 in R07020. Similarly, the maize ortholog (*GRMZM2G037650*) of this gene was reported to be up-regulated in non-elongating internodes of maize (Bosch et al, 2011). Additional TF families that were differentially expressed in nodal segments of BTx623 and R07020 included *AP2/EREBP*, *C2H2 zinc finger*, *WRKY*, *bHLH*, and *Aux/IAA* (Tables 14 and 15). Elevated expression of

transcription factors including members of the *AP2-EREBP*, *bHLH* and *WRKY* families was observed in basal nodal segments in both R07020 and BTx623 genotypes. The differential expression of *WRKY* TFs during the maturation of sorghum stems is consistent with their reported role in secondary cell wall formation in *Miscanthus* (Yu, et al., 2013). Among the other differentially expressed TFs were five members of the *AUX/IAA* family (*Sobic.001G161500*, *Sobic.003G291200*, *Sobic.004G336500*, *Sobic.007G009300*, *Sobic.010G065500*), which were up-regulated in nodal segment N6 of genotype R07020 (Table 14). In BTx623, three differentially expressed *AUX/IAA* TFs were observed, of which two were up-regulated in N16 as compared to N6 (Table 15). The *AUX/IAA* TF family is comprised of early auxin-response genes that have been reported to have a role in auxin-dependent developmental processes (Overvoorde, et al., 2005). The possible involvement of *AUX/IAA* TFs in cell wall biosynthesis has been reported in other species, such as maize (Bosch, et al., 2011) and sugarcane (Ferreira, et al., 2016).

Table 14: Differentially expressed key transcription factors in sorghum genotype R07020. DEGs represent expression differences between nodal segments N6 and N14. DEGs were defined as having a $\log_2$ fold change ≥ 1 or ≤ -1 with an FDR corrected *p*-value <0.05.

Feature ID	Annotation	Fold change[a]	FDR *p*-value	Gene ortholog	
				Arabidopsis	Maize
Sobic.001G075700	*AP2/EREBP*	-8.91	0	*AT4G37750*	*GRMZM2G028151*
Sobic.001G448000	*AP2/EREBP*	-12.97	0	*AT4G37750*	*GRMZM2G028151*
Sobic.001G481400	*AP2/EREBP*	6.13	0	*AT1G72360*	N/A
Sobic.002G022600	*AP2/EREBP*	-7.34	0	*AT4G37750*	*GRMZM2G028151*
Sobic.003G390600	*AP2/EREBP*	22.05	0	*AT5G17430*	*GRMZM2G141638*
Sobic.004G214300	*AP2/EREBP*	92.1	0	*AT1G51190*	N/A
Sobic.004G331300	*AP2/EREBP*	6.37	0	*AT5G25190*	N/A
Sobic.006G184700	*AP2/EREBP*	90.21	0	*AT2G36450*	N/A
Sobic.007G077100	*AP2/EREBP*	21.98	0	*AT3G20310*	*GRMZM2G068967*
Sobic.007G077200	*AP2/EREBP*	39.38	0	*AT5G44210*	N/A
Sobic.007G077300	*AP2/EREBP*	7.99	0	*AT1G12980*	*GRMZM2G120401*
Sobic.009G024900	*AP2/EREBP*	-4.73	0	*AT2G45490*	*GRMZM2G107645*
Sobic.009G122600	*AP2/EREBP*	4.96	0	*AT5G12330*	*GRMZM2G077752*
Sobic.009G142200	*AP2/EREBP*	24.94	0	*AT1G01720*	*GRMZM2G014653*
Sobic.009G155200	*AP2/EREBP*	14.6	0	*AT3G57670*	*GRMZM2G084014*
Sobic.009G238500	*AP2/EREBP*	5.31	0	*AT1G64380*	*GRMZM5G846057*
Sobic.009G247300	*AP2/EREBP*	18.23	0	*AT4G18170*	N/A
Sobic.010G052700	*AP2/EREBP*	5.29	0	*AT4G29080*	N/A
Sobic.010G220400	*AP2/EREBP*	-14.99	0	*AT1G30950*	*GRMZM2G109966*
Sobic.001G161500	*Aux/IAA*	3.95	0	*AT1G04240*	N/A
Sobic.003G291200	*Aux/IAA*	7.27	0	*AT3G16500*	*GRMZM2G057067*
Sobic.004G336500	*Aux/IAA*	12.97	0	*AT1G04550*	*GRMZM2G142768*
Sobic.005G098400	*Aux/IAA*	-7.16	0	*AT1G04550*	*GRMZM2G142768*
Sobic.007G009300	*Aux/IAA*	3.33	0	*AT1G51950*	*GRMZM2G057067*
Sobic.010G065500	*Aux/IAA*	27.31	0	*AT2G24260*	*GRMZM2G316758*
Sobic.001G068300	*bHLH*	4.2	0	*AT3G59060*	*GRMZM2G016756*
Sobic.001G107400	*bHLH*	10.64	0	*AT4G37850*	*GRMZM2G019806*
Sobic.001G430000	*bHLH*	-5.16	0	*AT1G68810*	*GRMZM2G043854*

Table 14: Continued

Feature ID	Annotation	Fold change[a]	FDR *p*-value	Gene ortholog	
				Arabidopsis	Maize
Sobic.001G518900	*bHLH*	-21.91	0	*AT2G34820*	*GRMZM2G137374*
Sobic.002G246600	*bHLH*	9.26	0	*AT2G42280*	*GRMZM2G142932*
Sobic.003G061400	*bHLH*	4.51	0	*AT2G40200*	*GRMZM2G088443*
Sobic.004G267400	*bHLH*	4.5	0	*AT4G34530*	*GRMZM2G101350*
Sobic.005G131900	*bHLH*	-6.84	0	*AT3G26744*	*GRMZM2G033356*
Sobic.006G175200	*bHLH*	21.32	0	*AT1G63650*	*GRMZM2G172795*
Sobic.006G175500	*bHLH*	181.32	0	*AT1G63650*	*GRMZM2G172795*
Sobic.007G155300	*bHLH*	57.554	0	*AT3G50330*	*GRMZM2G045431*
Sobic.007G224200	*bHLH*	5.23	0	*AT1G68920*	*GRMZM2G137541*
Sobic.008G188900	*bHLH*	2.39	0	*AT4G34610*	*GRMZM2G099319*
Sobic.010G070900	*bHLH*	-5.58	0	*AT2G42840*	*GRMZM2G092968*
Sobic.001G340900	*MYB*	-5.85	0	*AT1G22640*	*GRMZM2G089244*
Sobic.001G358900	*MYB*	-180.15	0	*AT1G57560*	*GRMZM2G017520*
Sobic.001G391500	*MYB*	11.37	0	*AT1G48000*	*GRMZM2G057027*
Sobic.001G397900	*MYB*	43.29	0	*AT2G47460*	*GRMZM2G022686*
Sobic.002G196000	*MYB*	-5.68	0	*AT5G16600*	*GRMZM2G106558*
Sobic.002G275500	*MYB*	2.35	0	*AT4G12350*	*GRMZM2G037650*
Sobic.002G423300	*MYB*	8.22	0	*AT5G49620*	*GRMZM2G070849*
Sobic.003G008700	*MYB*	-5.8	0	*AT4G32730*	*GRMZM2G470307*
Sobic.003G034500	*MYB*	34.56	0	*AT2G36890*	N/A
Sobic.003G087600	*MYB*	7.26	0	*AT1G68320*	*GRMZM2G096358*
Sobic.003G373000	*MYB*	5.13	0	*AT4G09460*	N/A
Sobic.006G030200	*MYB*	102.63	0	*AT5G57620*	*GRMZM2G011422*
Sobic.006G199800	*MYB*	16.28	0	*AT1G79180*	*GRMZM2G038722*
Sobic.007G177100	*MYB*	22.17	0	*AT4G34990*	*GRMZM2G000818*
Sobic.008G016800	*MYB*	8.15	0	*AT3G56930*	*GRMZM2G005834*
Sobic.008G117800	*MYB*	6.89	0	*AT3G62960*	N/A
Sobic.008G133700	*MYB*	3.2	0	*AT5G59780*	*GRMZM2G047626*
Sobic.009G075500	*MYB*	14.65	0	*AT4G18960*	*GRMZM2G359952*
Sobic.009G221400	*MYB*	10.58	0	*AT1G13260*	*GRMZM2G059939*
Sobic.003G334600	*NAC*	6.41	0	*AT5G08790*	N/A

Table 14: Continued

Feature ID	Annotation	Fold change[a]	FDR *p*-value	Gene ortholog	
				Arabidopsis	**Maize**
Sobic.005G064600	*NAC*	9.07	0	*AT1G77450*	N/A
Sobic.001G084000	*WRKY*	7.8	0	*AT1G69310*	*GRMZM2G018721*
Sobic.002G008600	*WRKY*	7.29	0	*AT1G69310*	*GRMZM2G018721*
Sobic.002G242500	*WRKY*	5.72	0	*AT5G56270*	*GRMZM2G031963*
Sobic.003G000600	*WRKY*	18.07	0	*AT1G62300*	*GRMZM2G366795*
Sobic.003G037400	*WRKY*	224.87	0	*AT5G64810*	*GRMZM2G137802*
Sobic.003G138400	*WRKY*	23.26	0	*AT4G04450*	*GRMZM2G448605*
Sobic.003G227300	*WRKY*	8.06	0	*AT4G18170*	N/A
Sobic.003G248400	*WRKY*	18.67	0	*AT4G18170*	N/A
Sobic.003G341100	*WRKY*	7.35	0	*AT2G38470*	*GRMZM2G036703*
Sobic.004G065900	*WRKY*	6.36	0	*AT1G80840*	*GRMZM2G120320*
Sobic.006G201000	*WRKY*	9.36	0	*AT1G30650*	*GRMZM2G024898*
Sobic.008G029400	*WRKY*	21.58	0	*AT2G46400*	*GRMZM2G025895*
Sobic.008G037900	*WRKY*	3.25	0	*AT1G08320*	*GRMZM2G060216*
Sobic.009G072200	*WRKY*	4.3	0	*AT5G56840*	*GRMZM2G049695*
Sobic.009G111200	*WRKY*	4.37	0	*AT3G16770*	N/A
Sobic.009G260600	*WRKY*	4.56	0	*AT3G44750*	N/A

[a]Positive values indicate greater expression in nodal segment N6 compared to N14. Negative values indicate greater expression in nodal segment N14 compared to N6.

Table 15: Differentially expressed key transcription factors in sorghum genotype BTx623. DEGs represent expression differences between nodal segments N6 and N16. DEGs were defined as having a $\log_2$ fold change ≥ 1 or ≤ -1 with an FDR corrected *p*-value <0.05.

Feature ID	Annotation	Fold change[a]	FDR *p*-value	Gene ortholog	
				Arabidopsis	Maize
Sobic.001G448000	*AP2/EREBP*	-8.45	0.00	*AT4G37750*	*GRMZM2G028151*
Sobic.001G481400	*AP2/EREBP*	13.71	0.00	*AT1G72360*	N/A
Sobic.004G214300	*AP2/EREBP*	274.13	0.00	*AT1G51190*	N/A
Sobic.006G184700	*AP2/EREBP*	52.21	0.00	*AT2G36450*	N/A
Sobic.009G024900	*AP2/EREBP*	-5.44	0.05	*AT2G45490*	*GRMZM2G107645*
Sobic.009G155200	*AP2/EREBP*	14.06	0.00	*AT3G57670*	*GRMZM2G084014*
Sobic.009G247300	*AP2/EREBP*	24.87	0.00	*AT4G18170*	N/A
Sobic.001G161500	*Aux/IAA*	-9.05	0.00	*AT1G04240*	N/A
Sobic.005G098400	*Aux/IAA*	-31.10	0.00	*AT1G04550*	*GRMZM2G142768*
Sobic.010G065500	*Aux/IAA*	7.64	0.01	*AT2G24260*	*GRMZM2G316758*
Sobic.001G068300	*bHLH*	4.20	0.00	*AT3G59060*	*GRMZM2G016756*
Sobic.001G430000	*bHLH*	-21.84	0.00	*AT1G68810*	*GRMZM2G043854*
Sobic.001G518900	*bHLH*	-74.00	0.00	*AT2G34820*	*GRMZM2G137374*
Sobic.002G246600	*bHLH*	19.48	0.00	*AT2G42280*	*GRMZM2G142932*
Sobic.004G267400	*bHLH*	4.61	0.00	*AT4G34530*	*GRMZM2G101350*
Sobic.007G224200	*bHLH*	2.98	0.03	*AT1G68920*	*GRMZM2G137541*
Sobic.010G070900	*bHLH*	-4.82	0.03	*AT2G42840*	*GRMZM2G092968*
Sobic.001G358900	*MYB*	-330.36	0.00	*AT1G57560*	*GRMZM2G017520*
Sobic.001G391500	*MYB*	4.44	0.01	*AT1G48000*	*GRMZM2G057027*
Sobic.003G008700	*MYB*	-13.74	0.00	*AT4G32730*	*GRMZM2G470307*
Sobic.003G034500	*MYB*	22.55	0.00	*AT2G36890*	N/A
Sobic.003G087600	*MYB*	5.71	0.00	*AT1G68320*	*GRMZM2G096358*
Sobic.007G177100	*MYB*	11.52	0.00	*AT4G34990*	*GRMZM2G000818*
Sobic.009G221400	*MYB*	19.43	0.00	*AT1G13260*	*GRMZM2G059939*
Sobic.003G334600	*NAC*	8.28	0.00	*AT5G08790*	N/A
Sobic.005G064600	*NAC*	3.07	0.05	*AT1G77450*	N/A
Sobic.001G084000	*WRKY*	5.22	0.00	*AT1G69310*	*GRMZM2G018721*
Sobic.002G008600	*WRKY*	3.36	0.02	*AT1G69310*	*GRMZM2G018721*
Sobic.002G242500	*WRKY*	2.25	0.03	*AT5G56270*	*GRMZM2G031963*

Table 15: Continued

Feature ID	Annotation	Fold change[a]	FDR *p*-value	Gene ortholog	
				Arabidopsis	**Maize**
Sobic.003G000600	*WRKY*	69.99	0.00	*AT1G62300*	*GRMZM2G366795*
Sobic.003G037400	*WRKY*	17.99	0.00	*AT5G64810*	*GRMZM2G137802*
Sobic.003G138400	*WRKY*	9.28	0.02	*AT4G04450*	*GRMZM2G448605*
Sobic.003G227300	*WRKY*	6.40	0.02	*AT4G18170*	N/A
Sobic.003G248400	*WRKY*	4.49	0.00	*AT4G18170*	N/A
Sobic.003G341100	*WRKY*	12.85	0.00	*AT2G38470*	*GRMZM2G036703*
Sobic.009G072200	*WRKY*	11.98	0.00	*AT5G56840*	*GRMZM2G049695*
Sobic.009G260600	*WRKY*	10.35	0.00	*AT3G44750*	N/A

[a]Positive values indicate higher expression in nodal segment N6 compared to N16. Negative values indicate higher expression in nodal segment N16 compared to N6.

Cellulose biosynthesis, hemicellulose metabolic process, and cell wall catabolism

The *cellulose synthase A* gene family which is found in all seed plant species (Joshi and Mansfield, 2007) is involved in cellulose biosynthesis at the plasma membrane (Handakumbura, et al., 2013). In rice (*Oryza sativa*), O*sCesA1*, O*sCesA3* and O*sCesA8* are involved in primary cell wall biosynthesis (Wang, et al., 2010), while *OsCesA4*, *OsCesA7* and *OsCesA9* are involved in secondary cell wall biosynthesis (Tanaka, et al., 2003). In R07020, the expression of genes *CesA7* (*Sobic.002G118700*), *CesA10* (*Sobic.001G224300*), and *CesA12* (*Sobic.002G205500*) was up-regulated 2-fold in basal nodal segment N6 in comparison to apical nodal segment N14 (Table 16). The involvement of *CesA10-12* in secondary cell wall biosynthesis has been reported in maize (Appenzeller, et al., 2004), while microRNA knockdown of *CesA7* in *Brachypodium distachyon* resulted in a reduction in secondary cell wall thickness (Handakumbura, et al., 2013).

The examination of differentially expressed gene families in immature stem nodal segments near the shoot apex revealed a series of *CesA* gene family members that were up-regulated. The *CesA5* gene (*Sobic.001G045700*) was up-regulated 2-fold in node N14 versus N6 of R07020 (Table 16) while in BTx623, genes *CesA8* (*Sobic.002G094600*, 3.6-fold change), *CesA10* (*Sobic.001G224300,* 2.5-fold change), and *CesA11* (*Sobic.003G296400*, 2.3-fold change) were up-regulated in node N16 compared to N6 (Table 17). Similar observations of the elevated expression of *CesA8* (*Sobic.002G094600*), *CesA10* (*Sobic.001G224300*) and *CesA12* (*Sobic.002G205500*)

until seven days before anthesis was reported in internode 10 in the sweet sorghum genotype Della (McKinley, et al., 2016).

Because of the similarity of the β1–4 linkages of cellulose to the linkages of hemicellulose backbones, it has been hypothesized that a group of cellulose synthase-like (*Csl*) genes are involved in glycan backbone biosynthesis in the Golgi apparatus (Richmond and Somerville, 2000). Among the DEGs related to the hemicellulose metabolic process, *CslF* (*Sobic.002G334100)* and *CslH* (*Sobic.006G080600)* were up-regulated more than 10-fold in nodal segments N14 and N16 of genotypes R07020 and BTx623, respectively (Tables 16 and 17). Within this family of genes, *CslF* and *CslH* are monocot-specific whereas C*slB*, and *CslG* are distinctive to dicots, and the remaining *Csl* genes are present in both monocots and dicots (Vogel, 2008).

Table 16: Differentially expressed genes in sorghum genotype R07020 involved in cellulose synthesis, hemicellulose metabolic process, and cell wall catabolism. DEGs represent expression differences between nodal segments N6 and N14. DEGs were defined as having a $\log_2$ fold change ≥ 1 or ≤ -1 with an FDR corrected *p*-value <0.05.

Feature ID	Annotation	Fold change[a]	FDR *p*-value	Gene ortholog	
				Arabidopsis	**Maize**
Sobic.001G045700	*CESA5*	-2.12	0.00	*AT5G05170*	*GRMZM2G111642*
Sobic.002G118700	*CESA7*	2.38	0.00	*AT5G64740*	*GRMZM2G025231*
Sobic.001G224300	*CESA10*	1.93	0.02	*AT5G44030*	*GRMZM2G445905*
Sobic.002G205500	*CESA12*	1.95	0.01	*AT5G17420*	*GRMZM2G002523*
Sobic.001G490000	*CSLA*	-2.79	0.00	*AT5G03760*	*GRMZM2G105631*
Sobic.010G197300	*CSLA*	2.21	0.00	*AT5G03760*	*GRMZM2G105631*
Sobic.001G075600	*CSLC*	-2.27	0.00	*AT4G07960*	*GRMZM2G028286*
Sobic.002G022700	*CSLC*	-3.47	0.00	*AT4G07960*	*GRMZM2G028286*
Sobic.003G308100	*CSLC*	-2.31	0.00	*AT4G07960*	*GRMZM2G028286*
Sobic.007G090600	*CSLC*	67.51	0.00	*AT2G24630*	*GRMZM2G135286*
Sobic.008G125700	*CSLD*	-5.52	0.00	*AT1G02730*	*GRMZM2G015886*
Sobic.003G442500	*CSLE*	83.32	0.00	*AT4G23990*	N/A
Sobic.004G255200	*CSLE*	1.93	0.04	*AT1G55850*	*GRMZM2G014558*
Sobic.002G237900	*CSLE-1*	16.64	0.00	*AT1G55850*	*GRMZM2G014558*
Sobic.002G238300	*CSLE-2*	4.75	0.00	*AT1G55850*	*GRMZM2G014558*
Sobic.002G334100	*CSLF-1*	-25.21	0.00	*AT1G02730*	*GRMZM2G015886*
Sobic.002G334200	*CSLF-2*	42.53	0.00	*AT1G02730*	*GRMZM2G015886*
Sobic.006G080600	*CSLH*	-10.23	0.00	*AT2G32540*	*GRMZM2G074546*

[a]Positive values indicate greater expression in nodal segment N6 compared to N14. Negative values indicate greater expression in nodal segment N14 compared to N6.

Table 17: Differentially expressed genes in sorghum genotype BTx623 involved in cellulose synthesis, hemicellulose metabolic process, and cell wall catabolism. DEGs represent expression differences between nodal segments N6 and N16. DEGs were defined as having a $\log_2$ fold change ≥ 1 or ≤ -1 with an FDR corrected *p*-value <0.05.

Feature ID	Annotation	Fold change[a]	FDR *p*-value	Gene ortholog	
				Arabidopsis	**Maize**
Sobic.002G094600	*CESA8*	-3.62	0.00	*AT4G39350*	*GRMZM2G082580*
Sobic.003G296400	*CESA11*	-2.30	0.04	*AT4G18780*	*GRMZM2G037413*
Sobic.002G385800	*CSLA*	-2.85	0.02	*AT5G03760*	*GRMZM2G105631*
Sobic.004G075900	*CSLA*	-19.94	0.00	*AT5G22740*	*GRMZM2G105631*
Sobic.004G238700	*CSLA*	-3.02	0.03	*AT5G03760*	*GRMZM2G105631*
Sobic.010G197300	*CSLA*	-16.12	0.00	*AT5G03760*	*GRMZM2G105631*
Sobic.003G308100	*CSLC*	-10.00	0.00	*AT4G07960*	*GRMZM2G028286*
Sobic.008G125700	*CSLD*	-19.43	0.00	*AT1G02730*	*GRMZM2G015886*
Sobic.004G255200	*CSLE*	3.80	0.00	*AT1G55850*	*GRMZM2G014558*
Sobic.002G334100	*CSLF*	-11.73	0.03	*AT1G02730*	*GRMZM2G015886*
Sobic.006G080600	*CSLH*	-11.06	0.00	*AT2G32540*	*GRMZM2G074546*

[a]Positive values indicate greater expression in nodal segment N6 compared to N16. Negative values indicate greater expression in nodal segment N16 compared to N6.

DEGs encoding glycosyl transferases

Cell wall polysaccharide synthesis requires refined biosynthetic processes with the involvement of glycosyltransferases. Glycosyltransferases (GTs) are enzymes that catalyze the transfer of a single sugar from nucleotide sugar donors to acceptors, such as glycan chains, peptides, or lipids (Scheible and Pauly, 2004; Breton, et al., 2012). In total, 42 glycosyltransferase family members have been reported in sorghum in the Carbohydrate-Active Enzymes database (CAZy; http://www.cazy.org) (Lombard, et al., 2014). Seventeen GT families showed more than a 2-fold change in expression between nodal segment N6 versus nodal segment N14 in R07020, and 21 GT families showed up-regulation in N6 versus N16 in BTx623 (Table 18 and Table 19). Differentially expressed GTs included those in the GT8, GT43, and GT47 families. Mutations in genes from a number of these GT families have resulted in an irregular xylem (IRX) phenotype (Wu, et al., 2009; Doering, et al., 2012). The expression of two genes in the GT47 family, *Sobic.004G159100* and *Sobic.009G220200*, were up-regulated 4.6- and 2.4-fold, respectively, in basal nodal segment N6 versus N14 in R07020 (Table 18), and *Sobic.003G410800* was up-regulated 2.6-fold in nodal segment N16 in BTx623 (Table 19). These genes are homologs of *IRX10* and *IRX10-L*, which were shown to be involved in xylan biosynthesis in secondary cell walls in *Arabidopsis* (Wu, et al., 2009; Doering, et al., 2012). Bosch, et al. (2011) also reported the up-regulation of one GT47 family member in non-elongating internodes in maize, and this maize gene (*GRMZM2G100143*) is an ortholog of *Sobic.004G159100,* which was up-regulated in N6 in R07020.

Table 18: Differentially expressed glycosyl transferases in sorghum genotype R07020. DEGs represent expression differences between nodal segments N6 and N14 of sorghum genotype. DEGs were defined as having a $\log_2$ fold change ≥ 1 or ≤ -1 with an FDR corrected *p*-value <0.05.

Feature ID	Annotation	Fold change[a]	FDR *p*-value	Gene ortholog	
				Arabidopsis	**Maize**
Sobic.001G030500	GT1	7.13	0.00	*AT3G11340*	*GRMZM2G173926*
Sobic.001G030900	GT1	-86.14	0.00	*AT3G55700*	*GRMZM5G892627*
Sobic.001G044400	GT1	3.70	0.03	*AT3G02100*	*GRMZM2G363545*
Sobic.001G084300	GT1	2.92	0.00	*AT3G02100*	*GRMZM2G363545*
Sobic.001G084400	GT1	8.49	0.00	*AT3G02100*	*GRMZM2G363545*
Sobic.001G084600	GT1	4.17	0.00	*AT3G02100*	*GRMZM2G363545*
Sobic.001G084700	GT1	126.09	0.00	*AT3G02100*	*GRMZM2G363545*
Sobic.001G236400	GT1	32.78	0.00	*AT2G36970*	*GRMZM2G337048*
Sobic.001G325600	GT1	7.09	0.00	*AT2G43820*	*GRMZM2G171548*
Sobic.001G377200	GT1	3.49	0.00	*AT3G16520*	*GRMZM2G449019*
Sobic.002G070800	GT1	-3.61	0.00	*AT5G49690*	*GRMZM2G174192*
Sobic.002G085200	GT1	12.15	0.00	*AT3G55700*	*GRMZM5G892627*
Sobic.002G085300	GT1	-2.57	0.01	*AT3G11340*	*GRMZM2G173926*
Sobic.002G085400	GT1	3.71	0.01	*AT5G05860*	*GRMZM2G117878*
Sobic.002G130200	GT1	15.73	0.00	*AT3G16520*	*GRMZM2G449019*
Sobic.002G173900	GT1	31.55	0.00	*AT1G01420*	*GRMZM2G376272*
Sobic.002G311900	GT1	57.15	0.00	*AT3G21750*	*GRMZM2G113794*
Sobic.002G337400	GT1	-7.97	0.00	*AT3G16520*	*GRMZM2G449019*
Sobic.002G366500	GT1	3.33	0.00	*AT2G15480*	*GRMZM5G888620*
Sobic.003G042900	GT1	12.34	0.01	*AT4G15550*	*GRMZM5G896260*
Sobic.003G232900	GT1	15.10	0.00	*AT2G15480*	*GRMZM5G888620*
Sobic.003G233000	GT1	3.77	0.00	*AT2G15480*	*GRMZM5G888620*
Sobic.003G233100	GT1	6.63	0.00	*AT2G15480*	*GRMZM5G888620*
Sobic.003G287500	GT1	3.81	0.00	*AT3G16520*	*GRMZM2G449019*
Sobic.003G287600	GT1	3.55	0.00	*AT3G16520*	*GRMZM2G449019*
Sobic.003G287700	GT1	53.79	0.00	*AT3G16520*	*GRMZM2G449019*
Sobic.003G288000	GT1	5.31	0.00	*AT3G16520*	*GRMZM2G449019*
Sobic.003G288100	GT1	2.70	0.01	*AT3G16520*	*GRMZM2G449019*
Sobic.003G288200	GT1	17.35	0.00	*AT3G16520*	*GRMZM2G449019*

Table 18: Continued

Feature ID	Annotation	Fold change[a]	FDR *p*-value	Gene ortholog	
				Arabidopsis	**Maize**
Sobic.003G328300	GT1	-2.54	0.01	*AT2G30140*	*GRMZM2G440902*
Sobic.004G039500	GT1	2.43	0.01	*AT4G01070*	*GRMZM2G426415*
Sobic.004G071200	GT1	6.62	0.00	*AT4G15480*	*GRMZM5G818068*
Sobic.004G087100	GT1	53.03	0.00	*AT2G36780*	*GRMZM2G161625*
Sobic.004G087200	GT1	110.64	0.00	*AT2G36750*	*GRMZM2G325023*
Sobic.004G087300	GT1	7.55	0.04	*AT2G36790*	*GRMZM2G035755*
Sobic.004G191100	GT1	10.38	0.00	*AT1G22400*	*GRMZM2G476049*
Sobic.004G229800	GT1	58.88	0.00	*AT1G22360*	*GRMZM5G870067*
Sobic.005G032000	GT1	3.72	0.01	*AT4G14090*	N/A
Sobic.005G032200	GT1	3.46	0.00	*AT1G05560*	*GRMZM2G475884*
Sobic.006G097800	GT1	-46.62	0.00	*AT4G15550*	*GRMZM5G896260*
Sobic.006G123700	GT1	-4.63	0.01	*AT5G12890*	*GRMZM2G159404*
Sobic.006G124100	GT1	3.82	0.00	*AT1G07250*	*GRMZM2G448483*
Sobic.006G139900	GT1	3.12	0.01	*AT5G12890*	*GRMZM2G159404*
Sobic.006G174400	GT1	41.31	0.00	*AT2G15490*	*GRMZM2G168474*
Sobic.006G174600	GT1	11.60	0.03	*AT2G15490*	*GRMZM2G168474*
Sobic.007G135000	GT1	10.59	0.00	*AT1G22360*	*GRMZM5G870067*
Sobic.009G064300	GT1	2.42	0.05	*AT2G36780*	*GRMZM2G161625*
Sobic.009G187000	GT1	4.21	0.00	*AT2G36800*	*GRMZM2G479038*
Sobic.009G205600	GT1	2.01	0.04	*AT3G16520*	*GRMZM2G449019*
Sobic.009G205700	GT1	23.43	0.00	*AT3G16520*	*GRMZM2G449019*
Sobic.009G224100	GT1	-29.87	0.00	*AT4G15480*	*GRMZM5G818068*
Sobic.009G230400	GT1	-2.43	0.01	*AT3G55700*	*GRMZM5G892627*
Sobic.010G088800	GT1	42.89	0.01	*AT1G22340*	*GRMZM2G148316*
Sobic.010G091100	GT1	3.95	0.03	*AT2G36750*	*GRMZM2G325023*
Sobic.010G116900	GT1	2.31	0.04	*AT1G73880*	*GRMZM2G426242*
Sobic.010G178900	GT1	-11.92	0.00	*AT1G05560*	*GRMZM2G475884*
Sobic.010G179000	GT1	-7.08	0.00	*AT4G15550*	*GRMZM5G896260*
Sobic.010G179100	GT1	-2.64	0.02	*AT4G15550*	*GRMZM5G896260*
Sobic.010G179400	GT1	11.74	0.00	*AT1G01420*	*GRMZM2G376272*
Sobic.010G179500	GT1	4.04	0.00	*AT4G15550*	*GRMZM5G896260*
Sobic.010G210800	GT1	63.88	0.00	*AT1G22370*	N/A
Sobic.001G075600	GT2	-2.27	0.00	*AT4G07960*	*GRMZM2G476964*
Sobic.001G490000	GT2	-2.80	0.00	*AT5G03760*	*GRMZM2G443715*
Sobic.002G022700	GT2	-3.47	0.00	*AT4G07960*	*GRMZM2G476964*

Table 18: Continued

Feature ID	Annotation	Fold change[a]	FDR *p*-value	Gene ortholog	
				Arabidopsis	**Maize**
Sobic.007G090600	GT2	67.51	0.00	*AT4G31590*	*GRMZM2G365475*
Sobic.010G197300	GT2	2.22	0.01	*AT5G03760*	*GRMZM2G443715*
Sobic.001G378300	GT4	2.63	0.00	*AT4G02280*	*GRMZM5G842830*
Sobic.003G076900	GT4	-2.54	0.01	*AT5G01220*	*GRMZM2G477503*
Sobic.003G240500	GT4	98.98	0.00	*AT4G19460*	*GRMZM2G022311*
Sobic.004G357600	GT4	-9.18	0.00	*AT1G73370*	*GRMZM2G410704*
Sobic.005G089600	GT4	2.58	0.01	*AT4G10120*	*GRMZM2G008507*
Sobic.010G276700	GT4	-8.35	0.00	*AT1G73370*	*GRMZM2G410704*
Sobic.001G131900	GT8	-2.20	0.00	*AT3G61130*	*GRMZM2G565856*
Sobic.001G391300	GT8	10.98	0.00	*AT2G47180*	*GRMZM5G872256*
Sobic.001G479800	GT8	5.10	0.00	*AT4G33330*	*GRMZM2G109431*
Sobic.002G241100	GT8	-2.18	0.00	*AT5G47780*	*GRMZM2G391000*
Sobic.002G398400	GT8	150.42	0.00	*AT3G06260*	*GRMZM5G810254*
Sobic.002G423600	GT8	3.70	0.02	*AT1G56600*	N/A
Sobic.004G237800	GT8	7.46	0.00	*AT3G02350*	N/A
Sobic.008G022500	GT8	-4.48	0.00	*AT2G38650*	*GRMZM2G065309*
Sobic.009G144200	GT8	-4.08	0.00	*AT3G18660*	*GRMZM2G441987*
Sobic.010G101400	GT8	-2.47	0.03	*AT1G70090*	*GRMZM2G469828*
Sobic.007G111500	GT9	-2.44	0.00	*AT1G15980*	*GRMZM5G838196*
Sobic.002G188500	GT20	2.13	0.00	*AT1G23870*	*GRMZM2G527891*
Sobic.002G116000	GT5	-3.02	0.00	*AT1G32900*	*GRMZM2G444178*
Sobic.010G022600	GT5	60.30	0.00	*AT1G32900*	*GRMZM2G444178*
Sobic.002G194700	GT20	3.01	0.00	*AT1G68020*	*GRMZM5G872163*
Sobic.007G126100	GT20	5.36	0.00	*AT1G23870*	*GRMZM2G527891*
Sobic.003G394100	GT21	-2.62	0.00	*AT2G19880*	*GRMZM2G163464*
Sobic.001G050800	GT31	-3.59	0.00	*AT5G62620*	*GRMZM2G176774*
Sobic.002G068800	GT31	-3.04	0.00	*AT5G62620*	*GRMZM2G176774*
Sobic.002G210300	GT31	31.86	0.00	*AT5G57500*	*GRMZM2G181358*
Sobic.002G210400	GT31	51.82	0.00	*AT5G57500*	*GRMZM2G181358*
Sobic.002G222800	GT31	5.63	0.00	*AT1G77810*	*GRMZM2G431006*
Sobic.004G322000	GT31	288.90	0.00	*AT5G57500*	*GRMZM2G181358*
Sobic.004G322100	GT31	136.28	0.00	*AT5G57500*	*GRMZM2G181358*
Sobic.009G223400	GT31	4.07	0.01	*AT5G57500*	*GRMZM2G181358*
Sobic.010G230300	GT31	-2.10	0.01	*AT1G05170*	*GRMZM2G140278*

Table 18: Continued

Feature ID	Annotation	Fold change[a]	FDR *p*-value	Gene ortholog	
				Arabidopsis	**Maize**
Sobic.005G144500	GT34	2.57	0.01	*AT2G22900*	*GRMZM2G435120*
Sobic.001G083900	GT35	-3.56	0.00	*AT3G29320*	*GRMZM2G074158*
Sobic.003G358600	GT35	-2.81	0.00	*AT3G46970*	*GRMZM2G085577*
Sobic.004G125100	GT37	33.55	0.00	*AT2G03220*	*GRMZM5G870571*
Sobic.004G308200	GT37	7.13	0.00	*AT2G03220*	*GRMZM5G870571*
Sobic.004G308400	GT37	2.07	0.03	*AT2G03220*	*GRMZM5G870571*
Sobic.004G308500	GT37	5.11	0.01	*AT2G03220*	*GRMZM5G870571*
Sobic.004G308600	GT37	-11.86	0.00	*AT2G03220*	*GRMZM5G870571*
Sobic.006G097900	GT37	29.02	0.01	*AT2G03220*	*GRMZM5G870571*
Sobic.010G082000	GT37	137.60	0.00	*AT2G03220*	*GRMZM5G870571*
Sobic.001G506600	GT47	108.81	0.00	*AT2G20370*	*GRMZM5G865819*
Sobic.004G159100	GT47	4.66	0.00	*AT5G61840*	*GRMZM2G448834*
Sobic.006G186100	GT47	18.18	0.00	*AT5G62220*	*GRMZM2G305146*
Sobic.008G077900	GT47	-7.56	0.00	*AT2G29040*	*GRMZM2G317731*
Sobic.009G220200	GT47	2.39	0.00	*AT5G61840*	*GRMZM2G448834*
Sobic.001G521500	GT48	3.54	0.00	*AT5G13000*	*GRMZM5G840560*
Sobic.003G179600	GT48	-3.33	0.00	*AT1G06490*	*GRMZM2G465764*
Sobic.003G180100	GT48	-4.18	0.00	*AT1G06490*	*GRMZM2G465764*
Sobic.010G275800	GT48	-3.75	0.00	*AT5G13000*	*GRMZM5G840560*
Sobic.002G401600	GT61	4.47	0.00	*AT3G57380*	N/A
Sobic.003G087700	GT61	7.71	0.00	*AT3G18170*	*GRMZM6G474234*
Sobic.003G094700	GT61	318.29	0.00	*AT3G18170*	*GRMZM6G474234*
Sobic.003G095200	GT61	-3.12	0.00	*AT3G18170*	*GRMZM6G474234*
Sobic.003G095700	GT61	52.67	0.00	*AT3G18170*	*GRMZM6G474234*
Sobic.003G108500	GT61	31.82	0.00	*AT2G41640*	*GRMZM5G869788*
Sobic.003G431100	GT61	127.66	0.00	*AT2G41640*	*GRMZM5G869788*
Sobic.010G030900	GT61	3.23	0.00	*AT3G18170*	*GRMZM6G474234*
Sobic.010G143900	GT61	8.78	0.00	*AT3G18170*	*GRMZM6G474234*
Sobic.010G144400	GT61	2.74	0.00	*AT3G18170*	*GRMZM6G474234*
Sobic.010G152300	GT61	31.12	0.04	*AT3G18170*	*GRMZM6G474234*
Sobic.010G152500	GT61	35.98	0.02	*AT3G18170*	*GRMZM6G474234*
Sobic.010G256400	GT61	-2.82	0.05	*AT3G18170*	*GRMZM6G474234*
Sobic.001G012700	GT77	7.00	0.00	*AT4G19970*	*GRMZM2G421126*
Sobic.001G012800	GT77	4.82	0.00	*AT2G02061*	N/A

Table 18: Continued

Feature ID	Annotation	Fold change[a]	FDR *p*-value	Gene ortholog	
				Arabidopsis	**Maize**
Sobic.001G519700	GT77	5.14	0.05	*AT1G14590*	*GRMZM2G481052*
Sobic.002G110600	GT77	3.80	0.00	*AT1G14590*	*GRMZM2G481052*
Sobic.002G257100	GT77	11.74	0.00	*AT4G19970*	*GRMZM2G421126*
Sobic.003G404600	GT77	7.16	0.00	*AT1G14590*	*GRMZM2G481052*
Sobic.003G405200	GT77	14.06	0.03	*AT1G14590*	*GRMZM2G481052*
Sobic.007G194600	GT77	-4.33	0.00	*AT2G02061*	N/A
Sobic.004G303000	GT90	4.10	0.00	*AT5G23850*	*GRMZM2G505380*
Sobic.006G159800	GT90	26.07	0.00	*AT5G23850*	*GRMZM2G505380*
Sobic.007G001100	GT90	21.96	0.00	*AT5G23850*	*GRMZM2G505380*

[a]Positive values indicate greater expression in nodal segment N6 compared to N14. Negative values indicate greater expression in nodal segment N14 compared to N6.

Table 19: Differentially expressed glycosyl transferases in sorghum genotype BTx623. DEGs represent expression differences between nodal segments N6 and N16 of sorghum genotype BTx623. DEGs were defined as having a $\log_2$ fold change ≥ 1 or ≤ -1 with an FDR corrected *p*-value <0.05.

Feature ID	Annotation	Fold change[a]	FDR *p*-value	Gene ortholog	
				Arabidopsis	**Maize**
Sobic.001G030600	GT1	8.14	0.02	*AT3G11340*	*GRMZM2G173926*
Sobic.001G030900	GT1	-1154.92	0.00	*AT3G55700*	*GRMZM5G892627*
Sobic.001G377200	GT1	3.62	0.01	*AT3G16520*	*GRMZM2G449019*
Sobic.002G173900	GT1	55.04	0.00	*AT1G01420*	*GRMZM2G376272*
Sobic.002G305600	GT1	37.50	0.05	*AT1G22360*	*GRMZM5G870067*
Sobic.002G311300	GT1	-23.45	0.00	*AT3G16520*	*GRMZM2G449019*
Sobic.002G311600	GT1	-3.56	0.04	*AT1G07250*	*GRMZM2G448483*
Sobic.002G337400	GT1	-13.08	0.00	*AT3G16520*	*GRMZM2G449019*
Sobic.002G337500	GT1	-2.36	0.04	*AT3G16520*	*GRMZM2G449019*
Sobic.003G287700	GT1	14.63	0.00	*AT3G16520*	*GRMZM2G449019*
Sobic.003G288000	GT1	2.35	0.02	*AT3G16520*	*GRMZM2G449019*
Sobic.003G288100	GT1	3.43	0.01	*AT3G16520*	*GRMZM2G449019*
Sobic.004G071200	GT1	3.68	0.04	*AT4G15480*	*GRMZM5G818068*
Sobic.004G087100	GT1	16.07	0.00	*AT2G36780*	*GRMZM2G161625*
Sobic.004G087200	GT1	134.18	0.00	*AT2G36750*	*GRMZM2G325023*
Sobic.004G191100	GT1	4.25	0.02	*AT1G22400*	*GRMZM2G476049*
Sobic.004G229800	GT1	27.30	0.00	*AT1G22360*	*GRMZM5G870067*
Sobic.006G048000	GT1	-5.53	0.04	*AT4G01070*	*GRMZM2G426415*
Sobic.006G097800	GT1	-9.28	0.03	*AT4G15550*	*GRMZM5G896260*
Sobic.006G123700	GT1	-6.59	0.01	*AT5G12890*	*GRMZM2G159404*
Sobic.006G174600	GT1	22.28	0.00	*AT2G15490*	*GRMZM2G168474*
Sobic.007G027200	GT1	-4.96	0.00	*AT1G01420*	*GRMZM2G376272*
Sobic.007G135000	GT1	12.24	0.00	*AT1G22360*	*GRMZM5G870067*
Sobic.009G176600	GT1	326.53	0.00	*AT1G22380*	*GRMZM2G470524*
Sobic.009G205600	GT1	2.35	0.04	*AT3G16520*	*GRMZM2G449019*
Sobic.009G205700	GT1	5.49	0.02	*AT3G16520*	*GRMZM2G449019*
Sobic.009G211800	GT1	-4.13	0.01	*AT5G12890*	*GRMZM2G159404*
Sobic.010G091100	GT1	5.02	0.02	*AT2G36750*	*GRMZM2G325023*
Sobic.010G179000	GT1	-30.98	0.00	*AT4G15550*	*GRMZM5G896260*

Table 19: Continued

Feature ID	Annotation	Fold change[a]	FDR *p*-value	Gene ortholog	
				Arabidopsis	**Maize**
Sobic.010G179100	GT1	-5.82	0.00	*AT4G15550*	*GRMZM5G896260*
Sobic.010G210800	GT1	51.13	0.02	*AT1G22370*	N/A
Sobic.001G490000	GT2	-8.57	0.00	*AT5G03760*	*GRMZM2G443715*
Sobic.002G385800	GT2	-2.85	0.03	*AT5G03760*	*GRMZM2G443715*
Sobic.004G075900	GT2	-19.95	0.00	*AT5G22740*	*GRMZM2G105631*
Sobic.004G238700	GT2	-3.03	0.03	*AT5G03760*	*GRMZM2G443715*
Sobic.010G197300	GT2	-16.13	0.00	*AT5G03760*	*GRMZM2G443715*
Sobic.003G076900	GT4	-3.02	0.05	*AT5G01220*	*GRMZM2G477503*
Sobic.003G403300	GT4	-5.90	0.00	*AT1G04920*	*GRMZM5G875238*
Sobic.004G357600	GT4	-101.06	0.00	*AT1G73370*	*GRMZM2G410704*
Sobic.010G276700	GT4	-172.21	0.00	*AT1G73370*	*GRMZM2G410704*
Sobic.001G131900	GT8	-3.56	0.00	*AT3G61130*	*GRMZM2G565856*
Sobic.001G302200	GT8	4.49	0.00	*AT2G35710*	*GRMZM2G462261*
Sobic.001G384200	GT8	-2.11	0.05	*AT2G38650*	*GRMZM2G065309*
Sobic.001G400100	GT8	-3.64	0.00	*AT3G62660*	*GRMZM5G881123*
Sobic.002G420100	GT8	-3.36	0.00	*AT2G38650*	*GRMZM2G065309*
Sobic.004G177000	GT8	-6.35	0.00	*AT3G18660*	*GRMZM2G441987*
Sobic.008G022500	GT8	-7.74	0.00	*AT2G38650*	*GRMZM2G065309*
Sobic.009G144200	GT8	-4.15	0.00	*AT3G18660*	*GRMZM2G441987*
Sobic.010G092400	GT8	-3.90	0.00	*AT3G02350*	N/A
Sobic.007G111500	GT9	-3.40	0.01	*AT1G15980*	*GRMZM5G838196*
Sobic.002G188500	GT20	2.64	0.02	*AT1G23870*	*GRMZM2G527891*
Sobic.002G194700	GT20	2.62	0.01	*AT1G68020*	*GRMZM5G872163*
Sobic.007G126100	GT20	6.23	0.00	*AT1G23870*	*GRMZM2G527891*
Sobic.001G130600	GT31	-52.74	0.00	*AT1G74800*	*GRMZM2G399158*
Sobic.001G183600	GT31	-2.36	0.03	*AT1G05170*	*GRMZM2G140278*
Sobic.002G210300	GT31	4.56	0.03	*AT5G57500*	*GRMZM2G181358*
Sobic.002G210400	GT31	17.94	0.00	*AT5G57500*	*GRMZM2G181358*
Sobic.009G223400	GT31	6.67	0.00	*AT5G57500*	*GRMZM2G181358*
Sobic.010G230300	GT31	-5.65	0.00	*AT1G05170*	*GRMZM2G140278*
Sobic.001G401600	GT34	-4.21	0.00	*AT4G02500*	*GRMZM2G477256*
Sobic.004G256400	GT34	-43.50	0.00	*AT2G22900*	*GRMZM2G435120*
Sobic.005G144500	GT34	-40.45	0.00	*AT2G22900*	*GRMZM2G435120*
Sobic.001G083900	GT35	-3.43	0.00	*AT3G29320*	*GRMZM2G074158*

Table 19: Continued

Feature ID	Annotation	Fold change[a]	FDR *p*-value	Gene ortholog	
				Arabidopsis	**Maize**
Sobic.004G125100	GT37	15.40	0.03	*AT2G03220*	*GRMZM5G870571*
Sobic.004G308600	GT37	-4.98	0.00	*AT2G03220*	*GRMZM5G870571*
Sobic.001G409100	GT43	-2.77	0.01	*AT2G37090*	*GRMZM2G118959*
Sobic.002G430700	GT43	-5.59	0.00	*AT2G37090*	*GRMZM2G118959*
Sobic.003G129400	GT43	3.72	0.02	*AT1G27600*	*GRMZM2G179504*
Sobic.001G229100	GT47	-9.10	0.01	*AT2G20370*	*GRMZM5G865819*
Sobic.003G410600	GT47	-2.98	0.01	*AT5G61840*	*GRMZM2G448834*
Sobic.003G410800	GT47	-2.60	0.01	*AT1G27440*	*GRMZM5G898668*
Sobic.009G162700	GT47	-3.76	0.00	*AT5G61840*	*GRMZM2G448834*
Sobic.001G521500	GT48	4.31	0.00	*AT5G13000*	*GRMZM5G840560*
Sobic.003G179600	GT48	-3.83	0.04	*AT1G06490*	*GRMZM2G465764*
Sobic.003G180100	GT48	-8.47	0.00	*AT1G06490*	*GRMZM2G465764*
Sobic.010G275800	GT48	-7.75	0.00	*AT5G13000*	*GRMZM5G840560*
Sobic.003G095200	GT61	-9.35	0.00	*AT3G18170*	*GRMZM6G474234*
Sobic.003G095500	GT61	-5.57	0.00	*AT3G18170*	*GRMZM6G474234*
Sobic.003G431100	GT61	27.97	0.00	*AT2G41640*	*GRMZM5G869788*
Sobic.004G134100	GT61	-2.29	0.02	*AT3G18170*	*GRMZM6G474234*
Sobic.009G126100	GT61	-2.41	0.04	*AT2G41640*	*GRMZM5G869788*
Sobic.010G143900	GT61	5.40	0.00	*AT3G18170*	*GRMZM6G474234*
Sobic.010G256400	GT61	-19.21	0.00	*AT3G18170*	*GRMZM6G474234*
Sobic.010G256600	GT61	-2.34	0.01	*AT3G18180*	*GRMZM2G700386*
Sobic.003G293400	GT66	-2.36	0.03	*AT5G19690*	*GRMZM2G154165*
Sobic.002G368600	GT75	-7.02	0.00	*AT3G02230*	*GRMZM2G481027*
Sobic.003G304600	GT75	-23.90	0.00	*AT3G02230*	*GRMZM2G481027*
Sobic.001G012700	GT77	5.78	0.01	*AT4G19970*	*GRMZM2G421126*
Sobic.001G327700	GT77	-2.92	0.02	*AT1G19360*	*GRMZM2G172726*
Sobic.002G257100	GT77	6.96	0.00	*AT4G19970*	*GRMZM2G421126*
Sobic.007G194600	GT77	-120.31	0.00	*AT2G02061*	N/A
Sobic.006G159800	GT90	12.25	0.00	*AT5G23850*	*GRMZM2G505380*
Sobic.007G001100	GT90	18.96	0.00	*AT5G23850*	*GRMZM2G505380*
Sobic.010G196400	GT92	-3.58	0.02	*AT2G33570*	*GRMZM5G890953*

[a]Positive values indicate greater expression in nodal segment N6 compared to N16. Negative values indicate greater expression in nodal segment N16 compared to N6.

DEGs encoding cell wall proteins

Plant cell walls contain a vast array of proteins, glycoproteins, and proteoglycans. The structural proteins that include arabinogalactan proteins (AGPs), glycine-rich proteins (GRPs), proline-rich proteins (PRPs), leucine-rich repeat proteins (LRPs) and hydroxyproline-rich glycoproteins (HRGPs), constitute the majority of the cell wall proteins (Albersheim, et al., 2010). Differential expression of many LRPs, PRPs, and HRGPs along with other protein family genes was observed in basal versus top nodal segments in R07020 and BTx623 with a greater number of genes encoding structural proteins being up-regulated in apical segments (Tables 20 and 21). In both sorghum genotypes R07020 and BTx623, five proline-rich proteins showed higher expression in upper stem nodal segments in comparison with basal nodal segments. In addition, a majority of leucine-rich repeat (LRR) proteins were up-regulated in the upper nodal segments of both R07020 and BTx623. Many LRR proteins have been reported to be involved in cell wall related processes (Cassab, 1998; Bosch, et al., 2011). The LRR protein *Sobic.004G215000* was up-regulated in the upper nodal segments of BTx623 and R07020 with 397- and 6.2-fold increases, respectively. The ortholog of this gene (*GRMZM2G042181*) also showed higher expression in the elongating internodes versus basal internodes in maize although the increased expression was 5.6-fold greater compared to a nearly 400-fold increase observed in BTx623. Fasciclin-like arabinogalactan-protein (*Sobic.002G351000*), a homolog of *AtFLA11* in *Arabidopsis*, showed a greater than 7-fold increase in expression in the mature basal nodal segment of both R07020 and BTx623 (Tables 20 and 21). A correlation between *AtFLA11* and/or *AtFLA12* transcript abundance

in *Arabidopsis* and the onset of expression of secondary cell wall cellulose synthases in *Arabidopsis* stems has been reported (Brown, et al., 2005; Persson, et al., 2005). Gene-function analyses revealed that stems of *Arabidopsis* T-DNA knockout double mutant (*AtFLA11* and *AtFLA12*) had altered stem biomechanics with reduced tensile strength as well as altered cell-wall architecture (MacMillan, et al., 2010).

Co-expression network analysis

Co-expression network mining is a method to elucidate functional relationships between genes and gene families, and has been used to predict functional associations of TF expression to cell wall biosynthesis genes in rice (Hirano, et al., 2013) and sugarcane (Ferreira, et al., 2016). Utilizing lignin annotated genes as hubs and filtering for TFs, co-expression network mining identified 276 genes (nodes). The resulting network was further divided into two sub-networks which included 224 and 51 genes, respectively (data not shown). Eight lignin biosynthesis genes, including *4CL* (*Sobic.004G062500, Sobic.007G089900*), *CCR* (*Sobic.007G141200, Sobic.002G146000*), *COMT* (*Sobic.007G059100, Sobic.007G058800*), *CAD* (*Sobic.010G178300*) and *PAL* (*Sobic.004G220300*) were identified in the first sub-network (Figure 4). As the key concept of network analysis is node connectivity, Zhang and Horvath (2005) concluded that TFs that have more connections with a greater number of hub genes are more likely to play important roles in cell wall biosynthesis regulation. In addition to *NAC* and *MYB* TF family members, a large number of TFs showed high neigborhood connectivity and thus, may represent TFs that regulate lignin biosynthesis in sorghum. These TFs included members of the *WRKY*, *AP2/EREBP*, *bHLH*, *C2H2*, *HB*, *bZIP*, *B3* and *Aux/IAA* families

(Table 22). Using a co-expression network, Ferreira, et al. (2016) identified TFs in sugarcane related to cell wall biosynthesis that included members of the *AP2/EREBP*, *WRKY*, *AUX/IAA*, *bHLH*, *Homeobox* and *bZIP* families. As transcriptional co-regulation is one of the major coordinators in metabolic pathways (Allocco, et al., 2004), the present study provides further insight into the relationship between lignin biosynthesis genes and transcription factors.

Conclusion

This study extends the knowledge of the cascade of gene expression that occurs in monocot stems during the complex developmental transition from the shoot apex to the stem base. It also provides an examination of gene expression in sorghum genotypes whose stems were specifically bred as a biomass feedstock versus stems of a traditional grain sorghum inbred in a temperate production environment. This study provides basic knowledge of gene expression during plant development, and provides an atlas of genes for future studies that focus on specific tissues or cell types that will be required for a more comprehensive understanding of genes involved in cell wall biosynthesis.

Table 20: Differentially expressed genes in sorghum genotype R07020 involved in the synthesis of cell wall proteins. DEGs represent expression differences between nodal segments N6 and N14. DEGs were defined as having a $\log_2$ fold change ≥ 1 or ≤ -1 with an FDR corrected *p*-value <0.05.

Feature ID	Annotation	Fold change[a]	FDR *p*-value	Gene ortholog	
				Arabidopsis	**Maize**
Sobic.004G114800	similar to Hydroxyproline-rich glycoprotein-like	-2.52	0.00	*AT4G01050*	*GRMZM2G095082*
Sobic.004G157700	similar to Hydroxyproline-rich glycoprotein-like	5.02	0.00	*AT2G33490*	*GRMZM2G467640*
Sobic.001G389600	similar to Leucine Rich Repeat family protein, expressed	-5.43	0.00	*AT5G10020*	*GRMZM2G081857*
Sobic.001G403200	similar to Leucine Rich Repeat family protein, expressed	-6.71	0.00	*AT3G51740*	*GRMZM2G066274*
Sobic.005G005700	similar to Leucine Rich Repeat family protein, expressed	-6.26	0.00	*AT3G56100*	
Sobic.005G199700	similar to Leucine Rich Repeat family protein, expressed	-3.20	0.00	*AT1G03440*	*GRMZM2G099981*
Sobic.008G006800	similar to Leucine Rich Repeat family protein, expressed	-9.74	0.00	*AT3G51740*	*GRMZM2G066274*
Sobic.008G060000	similar to Leucine Rich Repeat family protein, expressed	2.17	0.03	*AT1G34420*	*GRMZM2G059214*
Sobic.008G074500	similar to Leucine Rich Repeat family protein, expressed	10.16	0.00	*AT1G79620*	*GRMZM2G023715*
Sobic.008G146900	similar to Leucine Rich Repeat family protein, expressed	2.89	0.00	*AT4G33300*	*GRMZM2G044724*
Sobic.008G156600	similar to Leucine Rich Repeat family protein, expressed	-3.86	0.00	*AT3G54650*	*GRMZM2G100121*
Sobic.008G186400	similar to Leucine Rich Repeat family protein, expressed	2.27	0.00	*AT5G49660*	*GRMZM2G080503*
Sobic.005G091300	similar to Leucine Rich Repeat, putative, expressed	2.81	0.00	*AT1G72180*	*GRMZM2G167253*
Sobic.008G190000	similar to Leucine-rich repeat family protein, putative, expressed	-4.27	0.00	*AT5G62710*	*GRMZM2G147857*
Sobic.001G061400	similar to Leucine-rich repeat transmembrane protein kinase, putative, expressed	-14.61	0.00	*AT3G03770*	*GRMZM2G120657*
Sobic.001G381100	similar to Leucine-rich repeat transmembrane protein kinase, putative, expressed	-5.74	0.00	*AT3G56370*	*GRMZM2G119850*
Sobic.004G215000	similar to Leucine-rich repeat-like protein	-6.23	0.00	*AT4G06744*	*GRMZM2G042181*
Sobic.009G254400	similar to Leucine-rich repeat family protein-like	-3.63	0.00	*AT3G17640*	*GRMZM2G099981*
Sobic.006G206700	similar to Leucine-rich repeat protein 1	6.29	0.00	*AT2G19330*	*GRMZM2G072518*
Sobic.007G027900	similar to Proline-rich protein family-like	-4.06	0.00	*AT3G25690*	*GRMZM2G062738*
Sobic.005G127800	weakly similar to Leucine Rich Repeat family protein	4.51	0.00	*AT2G34930*	*GRMZM2G372058*
Sobic.010G130700	weakly similar to Predicted leucine rich repeat protein	10.64	0.00	*AT2G34930*	*GRMZM2G372058*
Sobic.001G508100	similar to Pre-mRNA processing protein PRP39, putative, expressed	2.41	0.00	*AT5G43950*	*GRMZM2G093139*

Table 20: Continued

Feature ID	Annotation	Fold change[a]	FDR *p*-value	Gene ortholog	
				Arabidopsis	**Maize**
Sobic.003G223800	similar to Proline transport protein 2-like	39.25	0.00	*AT5G41800*	*GRMZM2G057733*
Sobic.001G265900	similar to Proline-rich protein precursor	-12.29	0.00	*AT4G38770*	*GRMZM2G003909*
Sobic.001G266100	similar to Proline-rich protein precursor	-16.33	0.00	*AT2G21140*	*GRMZM2G003909*
Sobic.001G266300	similar to Proline-rich protein, putative, expressed	-20.99	0.00	*AT2G21140*	*GRMZM2G003909*
Sobic.001G438500	similar to Proline-rich protein, putative, expressed	-8.85	0.00	*AT2G21140*	*GRMZM2G003909*
Sobic.004G286700	similar to Putative anter-specific proline-rich protein APG	-4.02	0.00	*AT5G55050*	*GRMZM2G340084*
Sobic.002G351000	similar to Putative fasciclin-like arabinogalactan-protein	7.89	0.03	*AT5G03170*	*GRMZM2G177242*
Sobic.004G034200	similar to Putative leucine-rich repeat transmembrane protein kinase 2	-4.28	0.00	*AT4G03390*	*GRMZM2G335638*
Sobic.004G073800	similar to Putative leucine-rich repeat transmembrane protein kinase	-2.35	0.00	*AT4G22130*	*GRMZM2G103070*
Sobic.003G042200	similar to Putative leucine-rich repeat/extensin 1	-3.51	0.00	*AT4G13340*	*GRMZM2G149201*
Sobic.001G135600	similar to Putative proline-rich protein	-6.68	0.00	*AT5G03820*	N/A

[a]Positive values indicate greater expression in nodal segment N6 compared to N14. Negative values indicate greater expression in nodal segment N14 compared to N6.

Table 21: Differentially expressed genes in sorghum genotype BTx623 involved in the synthesis of cell wall proteins. DEGs represent expression differences between nodal segments N6 and N16. DEGs were defined as having a $\log_2$ fold change ≥ 1 or ≤ -1 with an FDR corrected *p*-value <0.05.

Feature ID	Annotation	Fold change[a]	FDR *p*-value		Gene ortholog	
					Arabidopsis	Maize
Sobic.001G061400	similar to Leucine-rich repeat transmembrane protein kinase, putative, expressed	-11.00	0.00		*AT3G03770*	*GRMZM2G120657*
Sobic.001G265900	similar to Proline-rich protein precursor	-125.63	0.00		*AT4G38770*	*GRMZM2G003909*
Sobic.001G266100	similar to Proline-rich protein precursor	-213.06	0.00		*AT2G21140*	*GRMZM2G003909*
Sobic.001G266300	similar to Proline-rich protein, putative, expressed	-98.97	0.00		*AT2G21140*	*GRMZM2G003909*
Sobic.001G381100	similar to Leucine-rich repeat transmembrane protein kinase, putative, expressed	-14.82	0.00		*AT3G56370*	*GRMZM2G119850*
Sobic.001G389600	similar to Leucine Rich Repeat family protein, expressed	-19.45	0.00		*AT5G10020*	*GRMZM2G081857*
Sobic.001G403200	similar to Leucine Rich Repeat family protein, expressed	-19.86	0.00		*AT3G51740*	*GRMZM2G066274*
Sobic.001G438500	similar to Proline-rich protein, putative, expressed	-103.63	0.00		*AT2G21140*	*GRMZM2G003909*
Sobic.001G508100	similar to Pre-mRNA processing protein PRP39, putative, expressed	2.19	0.02		*AT5G43950*	*GRMZM2G093139*
Sobic.002G351000	similar to Putative fasciclin-like arabinogalactan-protein	7.71	0.04		*AT5G03170*	*GRMZM2G177242*
Sobic.003G042200	similar to Putative leucine-rich repeat/extensin 1	-12.99	0.00		*AT4G13340*	*GRMZM2G149201*
Sobic.003G223800	similar to Proline transport protein 2-like	3.66	0.02		*AT5G41800*	*GRMZM2G057733*
Sobic.004G034200	similar to Putative leucine-rich repeat transmembrane protein kinase 2	-25.45	0.00		*AT4G03390*	*GRMZM2G335638*
Sobic.004G073800	similar to Putative leucine-rich repeat transmembrane protein kinase	-9.46	0.00		*AT4G22130*	*GRMZM2G103070*
Sobic.004G157700	similar to Hydroxyproline-rich glycoprotein-like	4.50	0.00		*AT2G33490*	*GRMZM2G467640*
Sobic.004G215000	similar to Leucine-rich repeat-like protein	-397.29	0.00		*AT4G06744*	*GRMZM2G042181*
Sobic.004G286700	similar to Putative anter-specific proline-rich protein APG	-4.71	0.00		*AT5G55050*	*GRMZM2G340084*
Sobic.005G005700	similar to Leucine Rich Repeat family protein, expressed	-14.95	0.00		*AT3G56100*	N/A
Sobic.005G091300	similar to Leucine Rich Repeat, putative	3.34	0.00		*AT1G72180*	*GRMZM2G167253*

Table 21: Continued

Feature ID	Annotation	Fold change[a]	FDR *p*-value	Gene ortholog	
				Arabidopsis	**Maize**
Sobic.005G127800	weakly similar to Leucine Rich Repeat family protein	50.39	0.00	*AT2G34930*	GRMZM2G372058
Sobic.007G027900	similar to Proline-rich protein family-like	-5.44	0.01	*AT3G25690*	GRMZM2G062738
Sobic.008G006800	similar to Leucine Rich Repeat family protein, expressed	-29.59	0.00	*AT3G51740*	GRMZM2G066274
Sobic.008G060000	similar to Leucine Rich Repeat family protein, expressed	7.73	0.00	*AT1G34420*	GRMZM2G059214
Sobic.008G074500	similar to Leucine Rich Repeat family protein, expressed	5.46	0.00	*AT1G79620*	GRMZM2G023715
Sobic.008G156600	similar to Leucine Rich Repeat family protein, expressed	-10.85	0.00	*AT3G54650*	GRMZM2G100121
Sobic.008G190000	similar to Leucine-rich repeat family protein, putative, expressed	-8.54	0.00	*AT5G62710*	GRMZM2G147857
Sobic.009G254400	similar to Leucine-rich repeat family protein-like	-51.23	0.00	*AT3G17640*	GRMZM2G099981
Sobic.010G130700	weakly similar to Predicted leucine rich repeat protein	27.4	0.00	*AT2G34930*	GRMZM2G372058
Sobic.001G125100	similar to Prp18 domain, putative	5.10	0.04	*AT1G03140*	GRMZM2G034651
Sobic.001G261400	similar to Putative leucine-rich repeat transmembrane protein kinase	-7.68	0.00	*AT1G53730*	GRMZM2G153393
Sobic.001G330500	similar to Leucine rich repeat containing protein	-13.03	0.00	*AT5G06860*	GRMZM2G025105
Sobic.002G376100	similar to Leucine-rich repeat transmembrane protein kinase 1-like protein	-3.89	0.00	*AT1G53730*	GRMZM2G153393
Sobic.003G206100	similar to Proline-rich protein-like	-43.94	0.02	N/A	N/A
Sobic.003G250700	similar to Arabinogalactan protein-like	-5.72	0.00	*AT5G03170*	GRMZM2G177242
Sobic.003G420300	similar to Hydroxyproline-rich glycoprotein-like	-2.48	0.00	*AT3G56590*	GRMZM2G136830
Sobic.004G060900	similar to Putative leucine-rich repeat transmembrane protein kinase	-2.13	0.03	*AT1G11130*	N/A
Sobic.004G101200	similar to Proline-and threonine-rich protein	-14.79	0.00	*AT2G42840*	GRMZM2G092968
Sobic.005G187700	similar to Leucine Rich Repeat family protein	9.50	0.00	*AT3G47570*	GRMZM2G048801
Sobic.008G075700	similar to Leucine Rich Repeat family protein, expressed	3.39	0.05	*AT1G45616*	GRMZM2G174128
Sobic.010G199700	similar to Putative leucine-rich repeat transmembrane protein kinase	-7.95	0.00	*AT4G22130*	GRMZM2G103070

[a]Positive values indicate greater expression in nodal segment N6 compared to N16. Negative values indicate greater expression in nodal segment N16 compared to N6.

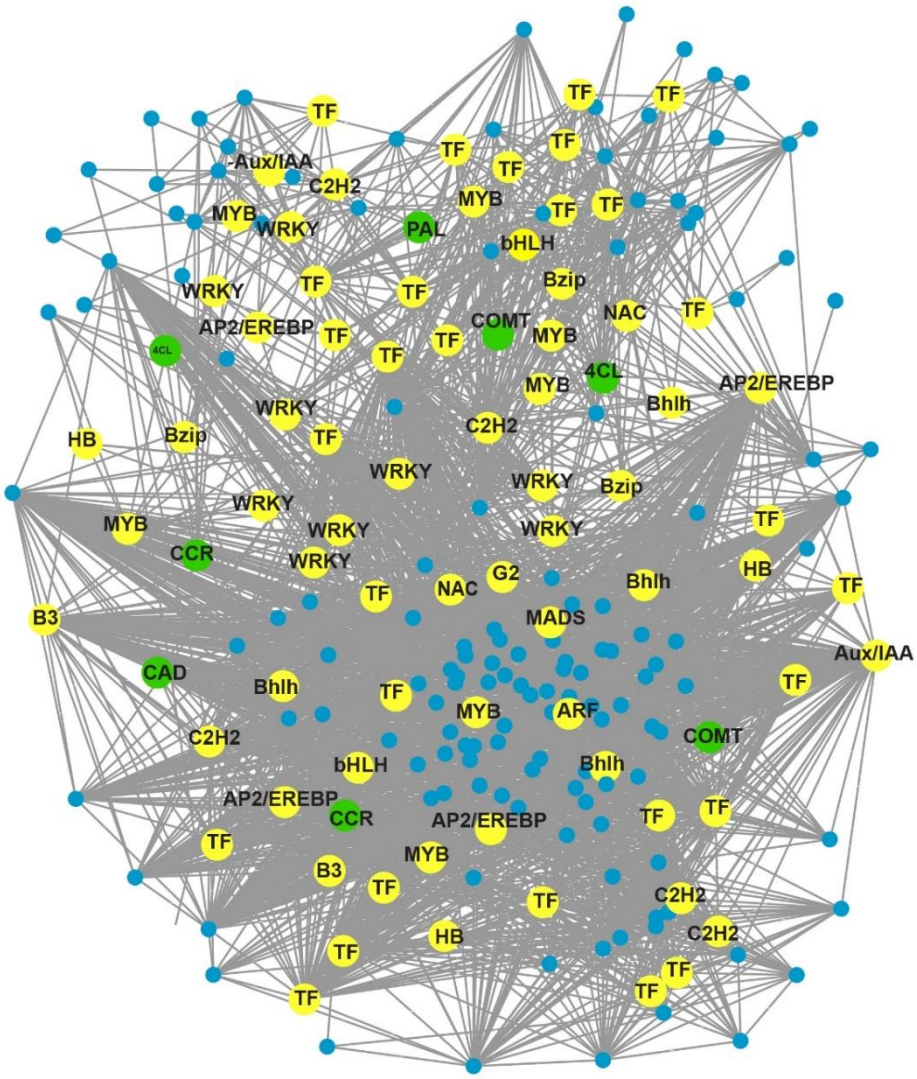


Figure 4: Co-expression network of sorghum using lignin biosynthesis genes as hubs. This sub-network comprises 224 genes (nodes); green nodes represent lignin biosynthesis genes and yellow nodes represent transcription factors with a high level of connectivity with hub genes. All other TFs are depicted as small blue nodes.

Table 22: List of TFs with a high level of connectivity with hub genes. The lignin biosynthesis genes were selected as the hub genes.

Feature ID	Annotation	Neighborhood connectivity[a]	Number of directed edges[b]	Gene ortholog	
				Arabidopsis	**Maize**
Sobic.001G107400	*bHLH*	110.50	10.00	*AT2G22750*	*GRMZM2G036554*
Sobic.005G131900	*bHLH*	96.44	72.00	*AT3G26744*	*GRMZM2G173534*
Sobic.003G128900	*MYB*	95.83	36.00	*AT1G69560*	*GRMZM2G460869*
Sobic.001G518900	*bHLH*	95.33	72.00	*AT2G34820*	N/A
Sobic.010G169200	*HB*	95.00	74.00	*AT4G32980*	*GRMZM2G333565*
Sobic.003G145100	*MYB*	94.61	69.00	*AT2G47460*	*GRMZM2G129872*
Sobic.002G022600	AP2/EREBP	92.54	84.00	*AT4G37750*	*GRMZM2G146688*
Sobic.006G241000	*HB*	92.26	82.00	*AT1G46480*	N/A
Sobic.001G075700	*AP2/EREBP*	91.47	88.00	*AT4G37750*	*GRMZM2G146688*
Sobic.001G379100	*bHLH*	90.40	91.00	*AT2G40435*	*GRMZM2G043493*
Sobic.001G479600	*B3*	90.40	91.00	*AT3G19184*	*GRMZM2G111123*
Sobic.009G124200	*AP2/EREBP*	90.28	90.00	*AT2G41710*	*GRMZM2G013657*
Sobic.001G448000	*AP2/EREBP*	90.16	94.00	*AT4G37750*	*GRMZM2G146688*
Sobic.009G083800	*MYB*	90.10	94.00	*AT4G32730*	*GRMZM2G531738*
Sobic.002G142000	*B3*	89.82	95.00	*AT3G19184*	*GRMZM2G111123*
Sobic.003G008700	*MYB*	89.82	95.00	*AT4G32730*	*GRMZM2G531738*
Sobic.006G255300	*AUX/IAA*	88.85	94.00	*AT1G19850*	*GRMZM2G086949*
Sobic.004G123200	*MYB*	88.54	97.00	*AT5G11510*	*GRMZM2G073836*
Sobic.006G262100	*AUX/IAA*	88.29	83.00	*AT1G30330*	*GRMZM2G475882*
Sobic.006G261200	*MYB*	88.27	98.00	*AT1G31310*	N/A
Sobic.010G211200	*C2H2*	86.08	95.00	*AT2G19380*	*GRMZM2G381638*
Sobic.003G251700	*ARF*	85.69	65.00	*AT2G33860*	*GRMZM5G874163*
Sobic.008G169400	*AUX/IAA*	84.15	102.00	*AT1G30330*	*GRMZM2G475882*
Sobic.001G430000	*bHLH*	81.17	112.00	*AT1G68810*	*GRMZM2G526668*
Sobic.006G133700	*bZIP*	81.00	113.00	*AT1G06070*	*AC200057.4_FG007*
Sobic.005G098400	*AUX/IAA*	79.57	108.00	*AT1G04550*	*GRMZM2G148188*
Sobic.001G455300	*C2H2*	68.08	53.00	*AT2G41940*	*GRMZM2G703281*

Table 22: Continued

Feature ID	Annotation	Neighborhood connectivity[a]	Number of directed edges[b]	Gene ortholog	
				Arabidopsis	Maize
Sobic.008G137500	*MYB*	64.12	42.00	*AT2G37630*	*GRMZM2G403620*
Sobic.002G290900	*NAC*	56.47	36.00	*AT3G17730*	*GRMZM5G885329*
Sobic.002G221400	*C2H2*	54.32	25.00	*AT2G01940*	*GRMZM2G465595*
Sobic.002G416400	*bHLH*	47.73	30.00	*AT2G40435*	*GRMZM2G043493*
Sobic.004G295500	*AP2/EREBP*	46.79	28.00	*AT2G44940*	*GRMZM5G889719*
Sobic.008G112200	*MYB*	45.90	29.00	*AT5G12870*	*GRMZM2G052606*
Sobic.001G488900	*HB*	38.43	21.00	*AT1G69780*	*GRMZM5G803812*
Sobic.002G275500	*MYB*	31.14	7.00	*AT4G22680*	N/A
Sobic.010G224200	*MYB*	30.15	20.00	*AT1G14600*	*GRMZM2G454449*
Sobic.003G076600	*MYB*	29.50	18.00	*AT5G58900*	*GRMZM2G311822*
Sobic.003G327000	*bZIP*	29.44	9.00	*AT2G16770*	*GRMZM2G175870*
Sobic.010G052600	*AP2/EREBP*	29.33	18.00	*AT4G36900*	N/A
Sobic.003G334600	*NAC*	27.91	11.00	*AT5G08790*	*GRMZM2G347043*
Sobic.005G224800	*MYB*	27.47	19.00	*AT5G59780*	*GRMZM2G130149*
Sobic.003G353700	*MYB*	27.36	22.00	*AT3G46640*	*GRMZM2G074908*
Sobic.005G064600	*NAC*	26.80	5.00	*AT5G63790*	*GRMZM2G126936*
Sobic.006G256200	*MYB*	26.71	28.00	*AT5G06800*	*GRMZM2G124495*
Sobic.003G000600	*WRKY*	26.09	23.00	*AT1G62300*	*GRMZM5G871347*
Sobic.005G038700	*bZIP*	25.83	12.00	*AT1G08320*	*GRMZM2G366264*
Sobic.003G363600	*bZIP*	25.47	19.00	*AT1G08320*	*GRMZM2G366264*
Sobic.001G084000	*WRKY*	24.88	16.00	*AT1G69310*	*GRMZM2G143204*
Sobic.001G068300	*bHLH*	24.16	32.00	*AT3G59060*	*GRMZM2G065374*
Sobic.001G358900	*MYB*	24.00	6.00	*AT1G09540*	*GRMZM2G171781*
Sobic.009G068900	*WRKY*	23.90	29.00	*AT5G26170*	*GRMZM2G163054*
Sobic.007G177100	*MYB*	23.84	19.00	*AT4G34990*	N/A
Sobic.001G384300	*MYB*	22.78	37.00	*AT4G28610*	*GRMZM2G379167*
Sobic.003G417500	*C2H2*	22.11	38.00	*AT1G03840*	*GRMZM2G151309*
Sobic.003G087600	*MYB*	21.38	8.00	*AT1G68320*	*GRMZM2G162709*
Sobic.002G424000	*bZIP*	20.45	38.00	*AT5G06950*	*GRMZM2G056099*
Sobic.008G037900	*bZIP*	19.33	6.00	*AT1G08320*	*GRMZM2G366264*
Sobic.002G344700	*bZIP*	19.30	10.00	*AT2G41870*	*GRMZM2G099239*

Table 22: Continued

Feature ID	Annotation	Neighborhood connectivity[a]	Number of directed edges[b]	Gene ortholog	
				Arabidopsis	**Maize**
Sobic.006G218300	*NAC*	18.88	8.00	*AT1G56010*	*GRMZM2G167018*
Sobic.009G072200	*MYB*	17.18	17.00	*AT5G56840*	*GRMZM2G071977*
Sobic.002G246600	*bHLH*	16.75	4.00	*AT2G42280*	*GRMZM5G879527*
Sobic.009G024600	*AP2/EREBP*	16.63	8.00	*AT2G28550*	*GRMZM5G862109*
Sobic.004G065900	*WRKY*	15.50	4.00	*AT1G80840*	*GRMZM2G125653*
Sobic.007G009300	*AUX/IAA*	15.09	11.00	*AT1G51950*	N/A
Sobic.009G016600	*MYB*	15.04	25.00	*AT3G46130*	*GRMZM2G305856*
Sobic.006G184700	*AP2/EREBP*	14.33	6.00	*AT5G52020*	*GRMZM2G434203*
Sobic.003G341100	*WRKY*	13.28	25.00	*AT2G38470*	*GRMZM2G449681*
Sobic.001G391500	*MYB*	13.18	11.00	*AT5G49620*	*GRMZM2G160840*
Sobic.009G221400	*MYB*	12.50	4.00	*AT1G25560*	N/A
Sobic.003G227300	*WRKY*	12.00	4.00	*AT1G29860*	*GRMZM5G812272*
Sobic.006G183200	*HB*	10.80	5.00	*AT3G61150*	*GRMZM2G060444*
Sobic.010G192400	*bZIP*	10.00	5.00	*AT5G06839*	*GRMZM2G380897*
Sobic.003G037400	*WRKY*	9.50	6.00	*AT5G64810*	*GRMZM5G863420*
Sobic.008G131400	*MYB*	9.33	6.00	*AT5G49620*	*GRMZM2G160840*
Sobic.003G138400	*WRKY*	9.17	6.00	*AT4G22070*	*GRMZM5G851490*

[a]The average connectivity of all neighbors of each node.
[b]The number of directed edges that are connected to each node.

CHAPTER III

TRANSCRIPTOME ANALYSIS OF A_1 CYTOPLASMIC MALE STERILE LINES, ISO-CYTOPLASMIC B-LINES, AND F_1 SORGHUM HYBRIDS

Introduction

In sorghum (*Sorghum bicolor (L.)* Moench), cytoplasmic male sterility (CMS) was first described by Stephens and Holland (1954), who observed that the interaction of Milo, or A_1 cytoplasm, and nuclear genes of Kafir origin produced plants with male sterility and normal female fertility. This system was quickly adopted by sorghum breeders, and by the 1960's almost all commercial varieties grown in developed countries were F_1 hybrids produced using the A_1 CMS system. Initial research into the genetics of fertility restoration in A_1 cytoplasm indicated that it is controlled by two-to-three major nuclear-encoded restoration of fertility (*Rf*) genes, with interference from modifiers or partial fertility (*Pf*) genes (Erichsen and Ross, 1963; Maunder and Pickett, 1963; Miller and Pickett, 1964). Fertility restoration is also influenced by environmental conditions; cool conditions around flowering favor sterility and high temperatures favor pollen fertility (Downes and Marshall, 1971; Brooking, 1976; Brooking, 1979). The combination of moderately complex genetic control and environmental variation in pollen sterility and fertility restoration make the development of new parental lines both laborious and costly.

Due to their economic importance, CMS-*Rf* systems have been extensively examined in many crop species (He, et al., 1995; Schnable and Wise, 1998; Hanson and Bentolila, 2004; Gabay-Laughnan and Newton, 2005; Chase, 2007; Ivanov and Dymshits, 2007; Li,

et al., 2007; Carlsson, et al., 2008; Carlsson and Glimelius, 2011; Park, et al., 2013; An, et al., 2014; Liu, et al., 2016a; Xie, et al., 2016). An interaction between the mitochondria and the nucleus results in an erroneous development of stamens and flowers resulting in inhibited production of viable pollen. Numerous investigations of the underlying basis of CMS have shown that in each case, a novel mitochondrial genomic lesion, consisting of a unique, expressed chimeric open reading frame (ORF), appears to confer the phenotype (Rhoads, et al., 1995; Tang, et al., 1996; Tang, et al., 1999; Hanson and Bentolila, 2004; Allen, et al., 2007; Yang, et al., 2010; Matsunaga, et al., 2011; Jing, et al., 2012; Kumar, et al., 2012). It has been hypothesized that the mitochondrial CMS-causing chimeric ORF affects the nuclear genome through retrograde signaling resulting in programmed cell death (PCD) and pollen abortion (Rhoads and Subbaiah, 2007; Diamond and McCabe, 2011; Eckardt, 2011; Schwarzlaender, et al., 2012).

In sorghum, there are a series of CMS-inducing cytoplasms (Reddy, et al., 2010), but nearly all commercial seed production relies on the A_1 CMS-*Rf* system. Despite its crucial economic importance, the CMS-inducing mitochondrial gene for A_1 cytoplasm is unknown. By comparison, considerably more is known about the *Rf* genes in sorghum, beginning with the cloning of the *Rf1* gene in sorghum by Klein, et al. (2005). *Rf1* was found to encode a pentatricopeptide repeat (PPR) protein, and subsequently, two other major fertility restoration loci in sorghum, *Rf2* and *Rf5*, were mapped with *Rf*-like PPR candidate genes identified (Jordan, et al., 2010; Jordan, et al., 2011). In recent years, numerous PPR proteins have been identified as fertility restorers in CMS systems from various plant species (Chen and Liu, 2014; Horn, et al., 2014; Tang, et al., 2014; Huang,

et al., 2015; Liu, et al., 2016b). PPR proteins are defined by the presence of canonical 35-amino acid degenerate motifs, which are found in tandem arrays of up to 30 repeats (Barkan and Small, 2014). In *Arabidopsis* about 450 PPR proteins have been identified and most of them are predicted to be targeted to mitochondria and/or chloroplasts. A number of functional studies revealed that PPR proteins exhibit sequence-specific RNA binding activity, supporting their general importance in organelle RNA metabolism (Barkan, et al., 2012; Barkan and Small, 2014). Mitochondrial gene regulation is largely controlled by post-transcriptional processes that include frequent RNA editing, trans-splicing of mitochondrial transcripts, 5' and 3' end processing, altered RNA stability, and translational control (Schmitz-Linneweber and Small, 2008; Farajollahi and Maas, 2010; Hammani and Giege, 2014). As transcripts of CMS-causing ORFs have been altered in restored lines of various plant species, it is postulated that PPR-*Rf* genes restore fertility by altering the expression of CMS-associated mitochondrial transcripts, either by editing/processing CMS-associated RNA or by altering the translation of these transcripts (Tang, et al., 1996; Tang, et al., 1999; Bentolila, et al., 2002; Brown, et al., 2003; Desloire, et al., 2003; Kazama and Toriyama, 2003; Koizuka, et al., 2003; Komori, et al., 2004; Wang, et al., 2006; Kazama, et al., 2008; Uyttewaal, et al., 2008; Barr and Fishman, 2010; Fujii, et al., 2011; Chen and Liu, 2014).

With the advent of cost-effective whole genome sequencing, comparative sequencing of related mitochondrial genomes and transcriptomes has become the preferred first steps in the process of identifying CMS-causing genes and the action of *Rf* genes in ameliorating pollen sterility. The use of RNA-seq technology to examine gene expression in crop

CMS systems has been reported in different species such as radish (Park, et al., 2013; Xie, et al., 2016), cotton (Yang, et al., 2014a), kenaf (Chen, et al., 2014), soybean (An, et al., 2014), and onion (Liu, et al., 2016a). Utilizing transcriptome analysis of CMS and maintainer lines in soybean, Li, et al. (2015a) identified a group of key nuclear-encoded genes related to male sterility, that included genes involved in carbohydrate and energy metabolism, transcription factors, regulation of pollen development, elimination of reactive oxygen species (ROS), cellular signal transduction, and PCD. Nuclear genome expression profiles of fertile and sterile floral buds of *pol* CMS in *Brassica napus* revealed differentially expressed genes that were mainly involved in metabolic and protein synthesis pathways while a set of unigenes controlling anther development were dramatically down-regulated in sterile buds (An, et al., 2014). In onions, Liu, et al. (2016a) reported the differential expression of several mitochondrial protein complex subunits including that of F-type ATPase, NADH dehydrogenase, and cytochrome c oxidase.

In the present study, RNA-seq technology was utilized to characterize gene expression in the nuclear genome of an A_1 cytoplasmic male sterile (CMS) system of sorghum. Gene expression was compared in immature panicles of A-lines, their iso-cytoplasmic maintainer B-lines, and F_1 sorghum hybrids to better understand the metabolic pathways that are differentially regulated in CMS sorghum.

Materials and Methods

Plant material and growth conditions

Sorghum plants from a series of A_1 CMS lines (A_1-line), maintainer lines (B-line), and F_1 (restored) hybrids (Table 23) were grown under controlled greenhouse conditions with a temperature range of 24/30 ± 2°C (light/dark) with a 14 hr light period provided using sodium halide lights and natural sunlight. Plants were grown in 3-gallon pots containing Sunshine REPS soil mix (Sun Gro Horticulture Inc., Bellevue, WA) with the following added and mixed in; 18 g of Osmocote (16% N, 3.5% P and 10% K), 15 g gypsum, 15 g dolomite and 5 g Micromix (6% Ca, 3% Mg, 12% S, 0.1% B, 1% Cu, 17% Fe, 2.5% Mn, 0.05% Mo and 1% Zn). The greenhouse experiment was conducted as a complete randomized design with four replicates for each line. In a separate field-based study, a set of sorghum genotypes, which included CMS line ATx623 and F_1 hybrids (ATx623×RTx2783, ATx623×RTx436) were grown in the Texas A&M AgriLife Research farms at the Brazos River Bottom, Burleson County, TX, and immature panicles were harvested for RNA isolation as detailed below.

Table 23: List of sorghum genotypes examined in the present study.

Lines	Name	Additional descriptors	Growing location
A_1-line	AQL33	CMS line-highly male sterile	Greenhouse
A_1-line	A992422	CMS line-male sterility, reverts to fertility under high temperatures	Greenhouse
B-line	BQL33	Iso-cytoplasmic maintainer line	Greenhouse
B-line	B992422	Iso-cytoplasmic maintainer line	Greenhouse
F_1 hybrid	AQL33×QL12		Greenhouse
F_1 hybrid	AQL33×QL36		Greenhouse
F_1 hybrid	AQL33×R931945-2-2		Greenhouse
F_1 hybrid	A992422×QL12		Greenhouse
F_1 hybrid	A992422×QL36		Greenhouse
F_1 hybrid	A992422×R931945-2-2		Greenhouse
A_1-line	ATx623	CMS line-highly male sterile	BRB[a] Field
F_1 hybrid	ATx623×RTx2783		BRB Field
F_1 hybrid	ATx623×RTx436		BRB Field

[a]BRB, Brazos River Bottom

RNA isolation, template library preparation and sequencing

Immature panicles from the early boot stage (cream-colored florets, wispy panicles) were harvested, florets stripped and immediately frozen in liquid nitrogen (LN_2) with subsequent storage at –80^{o}C. For RNA extraction, frozen florets were ground under LN_2 into a fine powder using a mortar and pestle. RNA was extracted from frozen ground tissue using the miRNeasy Mini Kit (Qiagen, Valencia, CA, USA) and subsequently treated with TURBO™ DNase (Ambion, Austin, TX, USA). Prior to RNA-seq template preparation, RNA quality was assessed using the ND-1000 NanoDrop spectrophotometer (NanoDrop Technologies, Montchanin, DE, USA), and RNA integrity was assessed with the Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA) by the Texas AgriLife Research Genomic and Bioinformatics Services. RNA-seq template libraries were prepared by the Texas AgriLife Research Genomic and Bioinformatics Services using the ScriptSeq^{TM} RNA-seq ⁻library preparation kit (Epicenter Technologies, Chicago, IL, USA) following the manufacturer's protocol and 125 bp strand-specific paired-end (PE) sequencing was performed on an Illumina HiSeq2500. Prior to sequencing, each library was bar-coded, and fourteen libraries pooled per lane on the Illumina flow cell. Samples were run on four lanes and reads from all lanes were combined for higher read depth.

Gene expression analysis

Sequence reads were imported into the CLC Genomics Workbench Version 8.5.1 (Qiagen, Valencia, CA, USA). The quality of the reads was examined using FastQC and the reads mapped to the BTx623 sorghum reference genome (Sbicolor_313 v.3.1) using

the CLC Genomics Workbench. Differential gene and transcript expression analyses were conducted based on reads per kilobase per million (RPKM) value for each annotated gene using the proportions-based Kal's test (Kal, et al., 1999). Differentially expressed genes (DEGs) were found from the comparisons of A_1-lines with B-lines and F_1 (restored) hybrids. DEGs were defined as having the value of $\log_2$ fold change ≥ 1 or ≤ -1 with an FDR corrected *p*-value <0.05. A list of PPR gene family members (see Table 25) that have been characterized as putative *Rf*-like genes was obtained for this analyses from Professor Ian Small, The University of Western Australia (personal communication). Expression of a putative *Rf*-like or *Pf*-like (partial fertility restorer) gene was compared between A_1-lines and F_1 hybrids and between A/B iso-cytoplasmic line pairs to identify any with differential expression.

For functional annotation, gene ontology was performed using AgriGO gene ontology analysis tools (Du, et al., 2010), and significantly enriched GO terms (in comparison to the genome background) were identified by REVIGO (Supek, et al., 2011). Pathway analysis was performed using the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (Kanehisa and Goto, 2000) and pathway enrichment analysis completed by KOBAS v. 2.0 (Xie, et al., 2011).

Results and Discussion

RNA-seq data quality

The number of paired-end reads that were collected for each RNA-seq library ranged from 79 M to 105 M, and the number of mapped reads ranged from 62 M to 95 M with an

average of 92% of those reads mapped to gene exons (Table 24). The high percentage of reads mapping to exonic regions is expected for high-quality RNA-seq data.

Table 24: Details of mapped reads of RNA-seq data from different A_1-lines, B-lines, and F_1 hybrids. Unique RNA-seq reads mapping to exons, introns, and intergenic regions are shown as the percentage of total reads distributed to these annotated regions of the sorghum genome (Sbicolor_313 v.3.1).

Genotype	Total reads (M)	Number of mapped reads (M)	Exonic reads (%)	Intronic reads (%)	Intergenic reads (%)
AQL33	84.18	64.7	93.82	0.99	5.19
BQL33	89.74	69.36	93.47	1.02	5.51
AQL33×QL12	87.98	66.95	93.98	0.95	5.07
AQL33×QL36	94.45	73.44	93.93	1.25	4.82
AQL33×R931945-2-2	86.10	65.12	92.37	0.93	6.71
A992422	81.60	63.09	93.64	0.97	5.39
B992422	105.70	82.01	93.95	1.02	5.03
A992422×QL12	95.22	73.19	93.64	1.15	5.21
A992422×QL36	83.89	64.57	93.84	0.96	5.20
A992422×R931945-2-2	80.45	62.52	93.36	1.00	5.64
ATx623	90.52	70.09	93.53	1.06	5.41
ATx623×RTx2783	85.13	66.02	93.99	1.19	4.83
ATx623×RTx436	85.60	66.32	93.48	1.09	5.43

Differentially expressed Rf and Rf-Like genes

The CMS phenotype is restored through the action of nuclear-encoded genes that are members of the large family of pentatricopeptide repeat (PPR) proteins (Delannoy, et al., 2007; Hu, et al., 2012). The expression of *Rf1* and *Rf-like* PPR genes in sorghum lines ranged from 0 to 6.5 RPKM (Table 25), but none of the annotated *Rf*-like genes were differentially expressed in the present study. The lack of differential expression of *Rf1* and *Rf-like* PPR genes between CMS sterile A_1-lines and fertile F_1 hybrids likely relates to the fact that regulation of the action of the *Rf* genes may not be at the transcriptional level. The regulation of these genes in sorghum may be related to the loss-of-function mutations that have been discovered in recessive alleles of cloned *rf* genes (Hu, et al., 2014; Kitazaki, et al., 2015; Gaborieau, et al., 2016). However, as additional genes involved in fertility restoration are annotated that include modifier genes (partial restorers), it is likely that control of these *pf* genes resides at the level of gene transcription, and that environmental cues (e.g., high temperature immediately preceding anther exertion) are critical to expression of *pf* genes.

Table 25: Expression of putative *Rf-like* and *Pf-like* (partial fertility restorer) genes in different lines of sorghum.

Feature ID	Annotation	A_1-lines			B-lines		F1 hybrids								Gene ortholog	
		A992422	AQL33	ATx623	B992422	BQL33	A992422×R931945-2-2	A992422×QL12	A992422×QL36	AQL33×R931945-2-2	AQL33×QL12	AQL33×QL36	ATx623×RTx436	ATx623×RTX2783	*Arabidopsis*	Maize
Sobic.002G054100	N/A	1[a]	1	2	1	1	2	2	2	2	2	2	2	2	*AT1G63130*	*GRMZM2G021303*
Sobic.002G054200	similar to Protein Rf1, mitochondrial precursor	3	3	2	3	3	2	3	4	3	3	4	2	4	*AT1G62910*	*GRMZM2G450166*
Sobic.002G055300	N/A	2	2	2	2	2	2	2	2	2	2	2	2	2	*AT5G64320*	*AC186379.3_FG001*
Sobic.002G057050	N/A	2	2	2	2	2	1	2	2	2	1	1	2	2	*AT5G55840*	*GRMZM2G393935*
Sobic.002G059700	N/A	0	0	0	0	0	0	0	0	0	0	0	0	0	*AT1G62910*	*GRMZM2G450166*
Sobic.002G227400	N/A	0	0	0	0	0	0	0	0	0	0	0	0	0	*AT1G62670*	*GRMZM2G416201*
Sobic.003G138550	N/A	0	0	0	0	0	0	0	0	0	0	0	0	0	*AT1G63070*	N/A
Sobic.005G011000	similar to Rf1 protein, mitochondrial, putative, expressed	1	1	1	2	1	1	1	2	1	1	1	1	1	*AT1G63130*	*GRMZM2G021303*
Sobic.005G024600	N/A	1	1	1	2	1	1	1	2	1	1	1	1	1	*AT1G63130*	*GRMZM2G021303*
Sobic.005G026500	N/A	0	0	0	0	0	0	0	1	0	1	0	1	1	*AT1G62670*	*GRMZM2G416201*

Table 25: Continued

Feature ID	Annotation	A_1-lines			B-lines		F1 hybrids								Gene ortholog	
		A992422	AQL33	ATx623	B992422	BQL33	A992422×R931945-2-2	A992422×QL12	A992422×QL36	AQL33×R931945-2-2	AQL33×QL12	AQL33×QL36	ATx623×RTx436	ATx623×RTX2783	*Arabidopsis*	Maize
Sobic.005G026900	**similar to PPR protein**	**1**	**1**	**1**	**2**	**1**	**1**	**1**	**1**	**1**	**1**	**0**	**1**	**1**	*AT1G63130*	*GRMZM2G021303*
Sobic.005G027600	N/A	**2**	**2**	**1**	**2**	**1**	**1**	**2**	**3**	**2**	**3**	**2**	**2**	**2**	*AT1G63130*	*GRMZM2G021303*
Sobic.005G027680	N/A	**1**	**1**	**1**	**1**	**0**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	*AT1G62910*	*GRMZM2G450166*
Sobic.005G027760	N/A	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	*AT1G63130*	*GRMZM2G021303*
Sobic.005G027840	N/A	**4**	**5**	**3**	**6**	**3**	**2**	**5**	**5**	**4**	**5**	**4**	**4**	**5**	*AT1G62910*	*GRMZM2G450166*
Sobic.005G028200	N/A	**2**	**2**	**2**	**3**	**1**	**1**	**2**	**3**	**2**	**3**	**1**	**2**	**2**	*AT1G63130*	*GRMZM2G021303*
Sobic.005G028300	**similar to Rf1 protein, mitochondrial, putative, expressed**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	*AT1G62670*	*GRMZM2G416201*
Sobic.005G030850	N/A	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	*AT5G65560*	*GRMZM2G030594*
Sobic.005G168000	N/A	**4**	**3**	**2**	**7**	**2**	**3**	**4**	**5**	**4**	**3**	**3**	**3**	**3**	*AT1G63130*	*GRMZM2G021303*
Sobic.007G087900	**similar to Protein Rf1, mitochondrial precursor**	**2**	**1**	**1**	**2**	**1**	**1**	**2**	**2**	**2**	**1**	**1**	**1**	**1**	*AT1G63130*	*GRMZM2G021303*
Sobic.008G099900	N/A	**1**	**0**	**0**	**0**	**0**	**0**	**1**	**0**	**0**	**0**	**0**	**0**	**0**	*AT3G29230*	*GRMZM2G077320*

Table 25: Continued

Feature ID	Annotation	A_1-lines			B-lines		F1 hybrids								Gene ortholog	
		A992422	AQL33	ATx623	B992422	BQL33	A992422×R931945-2-2	A992422×QL12	A992422×QL36	AQL33×R931945-2-2	AQL33×QL12	AQL33×QL36	ATx623×RTx436	ATx623×RTX2783	*Arabidopsis*	Maize
Sobic.008G100400	**similar to Pentatricopeptide, putative**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	*AT3G23330*	*GRMZM2G115957*
Sobic.008G147400	N/A	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	*AT4G35130*	*GRMZM2G009754*
Sobic.008G149200	**similar to Putative uncharacterized protein**	**1**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**1**	**0**	**0**	**0**	**1**	*AT1G68930*	*AC194339.3_FG004*
Sobic.008G163400	**similar to Pentatricopeptide, putative, expressed**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	*AT4G02750*	*GRMZM2G004888*
Sobic.009G253101	N/A	**0**	**0**	**0**	**1**	**0**	**0**	**0**	**1**	**1**	**1**	**0**	**1**	**1**	*AT1G63130*	GRMZM2G021303
Sobic.010G113900	N/A	**3**	**3**	**2**	**3**	**2**	**2**	**2**	**4**	**3**	**3**	**2**	**5**	**3**	*AT1G62670*	GRMZM2G416201

[a]Each number shows the RPKM value for each gene

Transcriptome analysis and gene ontology of iso-cytoplasmic A_1 and B paired lines

In an initial examination of the A_1 CMS system of sorghum, the expression of nuclear-encoded genes in A_1-lines and the corresponding iso-cytoplasmic B-lines was compared to identify genes differentially expressed during the early phase of panicle maturation. A limited number of differentially expressed genes (DEGs) were detected between A_1/B paired lines, which in part reflects the fact that each A_1/B pair shares the same nuclear genome and differ only in their cytoplasm (i.e., A_1 sterile vs. fertile cytoplasm). In a comparison between AQL33 and its maintainer line BQL33, 51 DEGs were detected with eight genes up-regulated and 43 genes down-regulated in AQL33. A limited number of DEGs were also detected in a comparison of gene expression in immature panicles of A_1/B paired lines A992422 and maintainer B992422 with 12 up-regulated and 10 down-regulated genes in A992422 (Table 26).

The functions of these DEGs were further studied based on annotation information ([Sbicolor_313 v.3.1, www.phytozome.jgi.doe.gov, (Paterson, et al., 2009)] and gene ontology (GO) analysis. Based on the limited number of DEGs between A_1/B paired lines, a gene ontology (GO) analysis did not reveal significant enrichment in any GO categories (Table 27). Despite this fact, examination of those genes differentially expressed between A_1/B paired lines revealed a series of genes involved in different biological processes (data not shown). Down-regulated genes in AQL33 and A992422 were involved in biological processes such as metabolic and cellular process, regulation of transcription, regulation of gene expression, and regulation of metabolic and biosynthetic process. Differentially expressed genes that were up-regulated in AQL33

and A992422 lines were involved in biological processes such as response to stress and metabolic process. A closer examination and discussion of DEGs between iso-cytoplasmic A_1 and B line pairs will be addressed in conjunction with those nuclear genes differentially expressed between A_1 and F_1 hybrids.

Transcriptome analysis and gene ontology of A_1 lines versus F_1 hybrids

Pollen-fertile F_1 hybrids in sorghum are produced by crossing A_1 CMS lines to restorer (R) lines that harbor *Rf* genes capable of restoring pollen fertility. Examination of nuclear gene expression between immature panicles of A_1 CMS females and their corresponding F_1 hybrids revealed a number of DEGs, and the quantity of DEGs was generally greater than that observed between iso-cytoplasmic A_1/B paired lines (Table 26). Unlike iso-cytoplasmic A_1/B paired lines that share the same nuclear genome, A_1 CMS lines and F_1 hybrids differ in their nuclear genomes, which likely accounts for the greater number of observed differentially expressed genes. Nevertheless, the number of DEGs detected in the present comparisons was limited (less than 300 DEGs), which is a small proportion of the over 30,000 annotated nuclear genes for sorghum [Sbicolor_313 v.3.1, www.phytozome.jgi.doe.gov, (Paterson, et al., 2009)]. As noted above, the relatively low number of DEGs may in part relate to the fact that a single stage of panicle development was examined in this study, and later (or earlier) stages during panicle maturation (e.g., anthesis) may have revealed additional changes in gene expression. In addition, while A_1 CMS panicles are pollen sterile, other facets of floral development progress normally including development of female reproductive

structures and thus, expression of various nuclear genes involved in floral development are likely unaffected.

The number of DEGs reported in comparison of CMS and fertile lines in other species varies widely. For example, Li, et al. (2015a) reported 365 DEGs (339 down-regulated and 26 up-regulated in CMS line) between NJCMS1A and its maintainer NJCMS1B line during the flowering period (prior to pollen abortion) of soybean. By comparison, Mei, et al. (2016) reported 3843 DEGs (2487 up-regulated and 1356 down-regulated genes in CMS line) in a comparison of radish HYBP-A CMS floral buds with HYBP-B maintainer line.

Differentially expressed genes between three A_1-lines and related F_1 hybrids were significantly enriched in a number of different GO categories and multiple biological processes (Table 27). In general, a larger proportion of DEGs were down-regulated in CMS A_1-lines when compared to pollen-fertile F_1 hybrids. A series of GO categories whose genes were down-regulated in CMS A_1-lines included nucleic acid/DNA binding, general category of metabolic processes, catalytic activities, lipid biosynthetic/metabolic process, oxidation/reduction, and organelle-related. Examination of these DEGs (and their homologs in other plant species) within these categories provided evidence that the down-regulation of these genes may relate to the eventual abortion of pollen that will occur within these panicles and to the large number of metabolic processes intimately linked to pollen production (see Appendix A). For instance, the down-regulation of the lipid biosynthesis process in CMS A_1-lines could be due to the importance of lipids or lipid-derived structures in pollen grains (Appendix A). Pollen grains contain several lipid

structures that play a key role in their development as male gametophytes (Piffanelli, et al., 1998), and the role of lipid synthesis for normal pollen development has also been hypothesized through mutational analysis (Ariizumi, et al., 2004) and chemical application studies (Li, et al., 2015b). Mutations in the novel plant protein NEF1 are known to affect lipid metabolism, and pollen cell wall formation which leads to reduced fertility in *Arabidopsis* (Ariizumi, et al., 2004). Li, et al. (2015b) found that the chemical hybridization agent monosulfuran ester sodium, induces male sterility in *Brassica napus* L. by blocking lipid and carbohydrate metabolism.

A number of genes with oxidoreductase activity were down-regulated in CMS A_1- lines. This is in agreement with Liu, et al. (2013) and Du, et al. (2016) who reported the up-regulation of oxidoreductase activity genes in fertile lines of Chili pepper and rapeseed, respectively. The down-regulated oxidoreductase genes in CMS A_1-lines included two dihydroflavonol-4-reductase genes (*Sobic.007G206000, Sobic.002G251200*), oxidoreductase from short chain dehydrogenase/reductase family gene (*Sobic.004G203900*), polyphenol oxidase (*Sobic.007G068700*), and 3-oxoacyl-reductase gene (*Sobic.008G087300*) (Appendix A). The involvement of *Arabidopsis* orthologs of some of these genes in anther and pollen development has been reported through isolation and characterization of numerous male sterile or partially sterile mutants (Morant, et al., 2007; Beaudoin, et al., 2009; Dobritsa, et al., 2009a; Dobritsa, et al., 2009b; Tang, et al., 2009; Dobritsa, et al., 2010; Kim, et al., 2010; Li and Zhang, 2010; Yang, et al., 2014c; Hu, et al., 2016). An ortholog of *Sobic.002G251200* and *Sobic.007G206000* in *Arabidopsis* i.e. *DRL1* (*AT4G35420*), which is an anther specific gene, has been reported

to be essential for male fertility in *Arabidopsis* (Tang, et al., 2009). The insertion mutants (*drl-1* and *drl-2*) of the *DRL1* gene caused impaired pollen development that lead to different levels of male sterility in *Arabidopsis* (Tang, et al., 2009). In support of these findings, the sorghum ortholog of *Arabidopsis DRL1* (*Sobic.007G206000)* was up-regulated in F_1 hybrids including AQL33×QL36 (2.1-fold change), A992422×QL12 (15-fold change), A992422×QL36 (3.6-fold change), A992422×R931945-2-2 (6-fold change), and ATx623×RTx2783 (2.3-fold change) in comparison with their related A_1-lines (Appendix A). The other ortholog of *DRL1*, *Sobic.002G251200*, was up-regulated in F_1 hybrids including AQL33×QL36 (2-fold change), A992422×QL12 (71.5-fold change), A992422×QL36 (12.7-fold change), and A992422×R931945-2-2 (18.3-fold change) (Appendix A).

Genes involved in exine (the outer pollen wall) formation were up-regulated in F_1 hybrids. These genes included *Sobic.001G428300* (annotated as similar to male fertility protein) and two putative chalcone synthase genes, i.e. *Sobic.002G115700* and *Sobic.001G215000*. The role of *Arabidopsis* orthologs of these genes in pollen exine formation has been reported (Dobritsa, et al., 2009a; Dobritsa, et al., 2010; Kim, et al., 2010). Dobritsa, et al. (2009a) investigated the role of *LAP3* (*AT3G59530*) in pollen development in *Arabidopsis* using insertion mutation and metabolite profile assays of anther tissues containing developing pollen grains. They found that the *lap3-2* defect leads to a broad range of metabolic changes in anthers including marked changes in levels of a straight-chain hydrocarbon nonacosane and naringenin chalcone, obligate compounds in the flavonoid biosynthesis pathway. Flavonoids are important for pollen

germination and fertility in several plant species (Van Der Meer, et al., 1992), and the disruption of key flavonoid biosynthesis enzymes, such as chalcone synthase, can disrupt pollen exine development (Zang, et al., 2009; Dobritsa, et al., 2010; Li and Zhang, 2010). Two known chalcone synthase genes in *Arabidopsis*, *LAP5* (*AT4G34850*) and *LAP6* (*AT1G02050*), encode anther-specific proteins that are essential for pollen exine development (Dobritsa, et al., 2010). Based on *in vitro* assays, it has been suggested that *LAP5* and *LAP6* are multifunctional enzymes and may play a role in both the synthesis of pollen fatty acids and phenolics found in exine (Kim, et al., 2010). In support of these findings, the sorghum orthologs of *Arabidopsis LAP3* (*Sobic.001G428300*), *LAP5* (*Sobic.002G115700*), and *LAP6* (*Sobic.001G215000*) were all up-regulated in F_1 hybrids. *Sobic.001G428300* (*LAP3* ortholog) was up-regulated in F_1 hybrids including AQL33×QL36 (2.3-fold change), A992422×QL12 (48-fold change), A992422×QL36 (10-fold change), and A992422×R931945-2-2 (19-fold change) (Appendix A). *Sobic.002G115700* (*LAP5* ortholog) was up-regulated in F_1 hybrids including AQL33×QL36 (2.1-fold change), A992422×QL12 (20-fold change), A992422×QL36 (4.3-fold change), and A992422×R931945-2-2 (7.5-fold change) (Appendix A). Similarly, the *LAP6* ortholog in sorghum, i.e. *Sobic.001G215000*, was up-regulated in F_1 hybrids including A992422×QL12 (17-fold change), A992422×QL36 (3.6-fold change), and A992422×R931945-2-2 (4.6-fold change) (Appendix A). The two genes, *Sobic.001G491400* and *Sobic.007G029900*, encoding cytochrome P450 enzymes, which were categorized in the metabolic process and oxidoreductase activity GO categories were up-regulated in F_1 hybrids. The orthologs of these genes in *Arabidopsis*, i.e. *CYP703B1*

(*AT1G69500*) and *CYP703A2* (*AT1G01280*), are involved in a major component of exine synthesis (sporopollenin) (Dobritsa, et al., 2009b). The mutants of these genes in *Arabidopsis* have been reported to cause moderate to severe effects in exine deposition and pollen grain development (Morant, et al., 2007; Dobritsa, et al., 2009b). Li and Zhang (2010) reported the involvement of the rice ortholog of *Sobic.001G491400* (*LOC_Os03g07250*) in male reproductive development. They identified this gene as a common fatty acid in the ω-hydroxylation pathway for synthesizing anther cuticle and pollen exine. *Sobic.001G491400* was up-regulated in F_1 hybrids including AQL33×QL36 (2.1-fold change), A992422×QL12 (20-fold change), A992422×QL36 (3.3-fold change), and A992422×R931945-2-2 (4.5-fold change). *Sobic.007G029900* was up-regulated in AQL33×QL36 (3.3-fold change), A992422×QL12 (39-fold change), and A992422×R931945-2-2 (6.4-fold change) (Appendix A). Two other genes involved in pollen development, *Sobic.006G079500* and *Sobic.003G004900*, were up-regulated in F_1 hybrids. The ortholog of these genes in *Arabidopsis*, i.e. *ACOS5* (*AT1G62940*), has been reported to be required for fatty acyl-CoA synthase for pollen development and sporopollenin biosynthesis (de Azevedo Souza, et al., 2009). de Azevedo Souza, et al. (2009) characterized a mutant of the *Arabidopsis ACOS5* gene, *acos5-1*, and found that *acos5-1* homozygotes are completely male sterile. The *Sobic.006G079500* and *Sobic.003G004900* genes were up-regulated in F_1 hybrids (AQL33×QL12, AQL33×QL36, A992422×QL12, A992422×QL36, A992422×R931945-2-2) (Appendix A). *Sobic.002G384400*, which was categorized in the metabolic process and oxidation reduction GO categories, was up-regulated in F_1 hybrid AQL33×QL36. The ortholog of

this gene in *Arabidopsis,* i.e. *Cp51* (*AT5G54770*) that encodes a cysteine protease is involved in anther development. *Cp51*–RNAi transgenic plants displayed a steady male sterility phenotype (Yang, et al., 2014c).

While a greater proportion of nuclear genes were down-regulated in CMS A_1-lines, there were a limited number of genes up-regulated in these lines. Three phenylalanine ammonia lyase genes (*Sobic.004G220400, Sobic.006G148800, Sobic.006G148900*) were up-regulated in two A_1-lines. *Sobic.004G220400* showed a 2.18-fold increase in AQL33. Both *Sobic.006G148800* and *Sobic.006G148900* showed 2-fold increases in A992422 (Appendix A). Although expression increased in these genes, the magnitude of this change was limited. This is in contrast to the large fold-change expression observed for many down-regulated genes in A_1 CMS lines. Phenylalanine ammonia lyase (PAL) is involved in the first step of the phenylpropanoid pathway and is therefore involved in the biosynthesis of polyphenol compounds such as flavonoids and phenylpropanoids. The phenylpropanoid natural products may play important roles as signal molecules, both in plant development and plant defense (Dixon, et al., 2002; Naoumkina, et al., 2010). The up-regulation of the ortholog of these genes in Arabidopsis has been reported in response to stress (Leyva, et al., 1995; Hussain, et al., 2016). In plants, male reproductive development is sensitive to stress. Upon exposure to stress, morphological, structural and metabolic alterations typically occur in male gametophytic organs that can lead to meiotic defects or premature spore abortion or male sterility (De Storme and Geelen, 2014).

Table 26: Number of DEGs in a comparison of different A_1-lines, B-lines, and F_1 hybrids. The significant DEGs were selected based on a $\log_2$ fold change ≥ 1 or ≤ -1 and an FDR corrected *p*-value <0.05.

Comparison	Total number of DEGs	Up-regulated in A_1-line	Down-regulated in A_1-line
AQL33 vs. B-QL33	51	8	43
AQL33 vs. AQL33×QL12	54	24	30
AQL33 vs. AQL33×QL36	125	40	85
AQL33 vs. AQL33×R931945-2-2	258	160	98
A992422 vs. B992422	22	12	10
A992422 vs. A992422×QL12	172	61	111
A992422 vs. A992422×QL36	101	39	62
A992422 vs. A992422×R931945-2-2	61	7	54
ATx623 vs. ATx623×RTx2783	83	31	52
ATx623 vs. ATx623×RTx436	63	47	16

Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway mapping

To identify DEG-enriched pathways associated with CMS and pollen fertility restoration, enrichment pathway analysis was performed. Although the number of DEGs were limited between sterile and fertile lines, a series of pathways did contain DEGs (Table 28). The up-regulation of genes *Sobic.001G318700*, *Sobic.001G514200*, *Sobic.010G106300*, and *Sobic.001G514400* in the glutathione metabolism pathway was found for some F_1 hybrid lines including AQL33×QL12, AQL33×QL36, and AQL33×R931945-2-2 (Appendix A). It has been postulated that the activation of

glutathione metabolism is a natural cell response to stress that can be induced by CMS (Fagard, et al., 2000). Elevated expression of *Arabidopsis* Glutathione S-transferase Φ8 *GSTF8* (*At2g47730*), the ortholog of *Sobic.001G514200,* has been reported as a marker for early stress and defense response (Alqurashi, et al., 2016). The effects of the glutathione pathway on CMS have also been supported by incorporation of the Glutathione S-Transferase (*GST*) gene into normal restorer line DES-HAF277 of cotton using *Agrobacterium*-mediated transformation, which significantly enhanced its fertility restoration potential (Wang and Li, 2002; Jiang, et al., 2007; Bibi, et al., 2014). Within the linoleic acid metabolism pathway, two genes (*Sobic.001G120400* and *Sobic.003G385900*) were up-regulated in CMS lines when compared to their F_1 hybrids (Appendix A). Fatty acids are crucial components of cellular membranes that provide structural barriers to the environment (Beisson, et al., 2007). They contribute to inducible stress resistance through the remodeling of membrane fluidity (Iba, 2002). Free linoleic acid is itself a stress signal (Blée, 2002). *Arabidopsis* gene *ATSPLA2-ALPHA* (*AT2G06925*) is the ortholog of *Sobic.001G120400,* and it has been shown to be involved in defense mechanisms (Froidure, et al., 2010). This gene is a member of the phospholipase A (PLA) superfamily that is involved in plant defense signaling (Canonne, et al., 2011). Previous studies indicate that the expression of the *Arabidopsis* ortholog i.e. *LOX1 (AT1G55020*) of *Sobic.003G385900* was induced by abiotic stresses (Melan, et al., 1993; Bu, et al., 2008).

The pathway protein processing in endoplasmic reticulum (ER) contained genes that were differentially expressed between CMS A_1 females and their F_1 hybrids. Among

those, *Sobic.001G420100*, *Sobic.002G243500*, *Sobic.003G039400*, *Sobic.003G082300*, *Sobic.008G136000*, and *Sobic.010G149300* were up-regulated in A_1 CMS parents (AQL33, A992422, and ATx623) when contrasted to their corresponding F_1 hybrids (AQL33×QL12, AQL33×QL36, AQL33×R931945-2-2, A992422×QL36, ATx623×RTx436, and ATx623×RTx2783). Genes *Sobic.004G228900*, *Sobic.001G426000*, *Sobic.003G081900*, *Sobic.003G350700*, *Sobic.006G005600*, *Sobic.007G216300*, and *Sobic.009G163900* showed up-regulation in A_1 CMS parent AQL33 and several F_1 hybrids including A992422×QL36 (*Sobic.004G228900*, *Sobic.001G426000*, and *Sobic.006G005600*), A992422×R931945-2-2 (*Sobic.003G081900*), AQL33×QL36 (*Sobic.003G350700*), ATx623×RTx436 (*Sobic.001G426000* and *Sobic.009G163900*), and ATx623×RTx2783 (*Sobic.009G163900* and *Sobic.001G426000*) (Appendix A). It has been reported that genes that are involved in protein processing in ER play an important role in flower formation and development (Ge, et al., 2016), and unfolded proteins in the ER are one type of plant sensor that can trigger different signal transduction events (McClung and Davis, 2010; Suzuki, et al., 2012). Protein processing in ER has been reported to regulate the balance of metabolic processes by modifying specific transcripts, proteins, metabolites or lipids (McClung and Davis, 2010; Suzuki, et al., 2012; Min, et al., 2014), and transcription factors that are associated with stress resistance have also been found in protein processing in the ER (Ge, et al., 2016). From the 13 genes that were categorized in protein processing in ER, 11 of them were annotated as heat shock protein genes.

Several genes assigned to photosynthesis-related processes were differentially expressed in the panicles of A_1-lines and their F_1 hybrids. Zhu, et al. (2015) reported the down-regulation of unigenes related to photosynthesis including photosynthesis-antenna proteins in SQ-1 induced male sterile wheat. The enrichment of photosynthesis and the photosynthesis-antenna proteins pathway has also been reported in a study of DEGs associated with fertility instability of S-type CMS in maize. The authors concluded that the DEGs with the function of photosynthesis may also play roles in the regulation of fertility instability (Su, et al., 2017). However, it should be noted that sorghum panicles in the early boot stage are largely devoid of chlorophyll and thus, contain etioplasts or proplastids rather than photosynthetically-active chloroplasts. Thus, the role of differentially expressed photosynthesis-related genes in immature panicles in CMS remains undetermined at this time.

Table 27: Enriched GO terms associated with DEGs in comparison of A_1-lines with B-lines and F_1 hybrids. The significant GO terms (FDR corrected *p*-value <0.05) were grouped into biological process, molecular function, and cellular component categories. *P*-values indicate the statistical significance of differential expression observed between compared lines for genes associated with each GO term. Qnum are the number of probe sets that belong to the GO term from the query list (DEGs). B/Rnum values are the number of probe sets that belong to the GO term from the background based on a genome-wide set of genes.

Comparison	GO term	GO type	GO name	Regulation[a]	Qnum	B/Rnum	FDR *p*-value
AQL33 vs. AQL33×QL12	GO:0003677	Molecular function	DNA binding	down	10	1598	0.00
AQL33 vs. AQL33×QL12	GO:0003676	Molecular function	nucleic acid binding	down	10	2374	0.00
AQL33 vs. AQL33×QL36	GO:0005515	Molecular function	protein binding	up	9	2890	0.04
AQL33 vs. AQL33×QL36	GO:0008152	Biological process	metabolic process	down	34	8076	0.00
AQL33 vs. AQL33×QL36	GO:0006629	Biological process	lipid metabolic process	down	7	681	0.03
AQL33 vs. AQL33×QL36	GO:0006508	Biological process	proteolysis	down	6	585	0.03
AQL33 vs. AQL33×QL36	GO:0055114	Biological process	oxidation reduction	down	10	1485	0.03
AQL33 vs. AQL33×QL36	GO:0003824	Molecular function	catalytic activity	down	33	7817	0.00
AQL33 vs. AQL33×QL36	GO:0016491	Molecular function	oxidoreductase activity	down	13	1669	0.00
AQL33 vs. AQL33×QL36	GO:0009055	Molecular function	electron carrier activity	down	7	576	0.01
AQL33 vs. AQL33×QL36	GO:0004175	Molecular function	endopeptidase activity	down	5	295	0.01
AQL33 vs. AQL33×QL36	GO:0070011	Molecular function	peptidase activity, acting on L-amino acid peptides	down	6	499	0.01
AQL33 vs. AQL33×QL36	GO:0008233	Molecular function	peptidase activity	down	6	520	0.01
AQL33 vs. AQL33×R931945-2-2	GO:0006457	Biological process	protein folding	up	5	108	0.00

Table 27: Continued

Comparison	GO term	GO type	GO name	Regulation[a]	Qnum	B/Rnum	FDR *p*-value
AQL33 vs. AQL33×R931945-2-2	GO:0006412	Biological process	translation	down	9	431	0.00
AQL33 vs. AQL33×R931945-2-2	GO:0003677	Molecular function	DNA binding	down	17	1598	0.00
AQL33 vs. AQL33×R931945-2-2	GO:0005198	Molecular function	structural molecule activity	down	8	343	0.00
AQL33 vs. AQL33×R931945-2-2	GO:0003735	Molecular function	structural constituent of ribosome	down	8	304	0.00
AQL33 vs. AQL33×R931945-2-2	GO:0003676	Molecular function	nucleic acid binding	down	21	2374	0.00
AQL33 vs. AQL33×R931945-2-2	GO:0030529	Cellular component	ribonucleoprotein complex	down	8	333	0.00
AQL33 vs. AQL33×R931945-2-2	GO:0005840	Cellular component	ribosome	down	8	304	0.00
AQL33 vs. AQL33×R931945-2-2	GO:0043232	Cellular component	intracellular non-membrane-bounded organelle	down	8	406	0.00
AQL33 vs. AQL33×R931945-2-2	GO:0043228	Cellular component	non-membrane-bounded organelle	down	8	406	0.00
AQL33 vs. AQL33×R931945-2-2	GO:0005622	Cellular component	intracellular	down	16	2129	0.02
AQL33 vs. AQL33×R931945-2-2	GO:0044444	Cellular component	cytoplasmic part	down	8	664	0.02
AQL33 vs. AQL33×R931945-2-2	GO:0032991	Cellular component	macromolecular complex	down	9	873	0.03
AQL33 vs. AQL33×R931945-2-2	GO:0043229	Cellular component	intracellular organelle	down	11	1325	0.04
AQL33 vs. AQL33×R931945-2-2	GO:0043226	Cellular component	organelle	down	11	1325	0.03
A992422 vs. A992422xQL12	GO:0044042	Biological process	glucan metabolic process	up	5	86	0.00

Table 27: Continued

Comparison	GO term	GO type	GO name	Regulation[a]	Qnum	B/Rnum	FDR *p*-value
A992422 vs. A992422×QL12	GO:0006073	Biological process	cellular glucan metabolic process	up	5	86	0.00
A992422 vs. A992422×QL12	GO:0044264	Biological process	cellular polysaccharide metabolic process	up	5	167	0.00
A992422 vs. A992422×QL12	GO:0005976	Biological process	polysaccharide metabolic process	up	5	197	0.00
A992422 vs. A992422×QL12	GO:0044262	Biological process	cellular carbohydrate metabolic process	up	5	381	0.03
AQL33 vs. AQL33×R931945-2-2	GO:0043226	Cellular component	organelle	down	11	1325	0.03
A992422 vs. A992422xQL12	GO:0044042	Biological process	glucan metabolic process	up	5	86	0.00
A992422 vs. A992422×QL12	GO:0006073	Biological process	cellular glucan metabolic process	up	5	86	0.00
A992422 vs. A992422×QL12	GO:0044264	Biological process	cellular polysaccharide metabolic process	up	5	167	0.00
A992422 vs. A992422×QL12	GO:0005976	Biological process	polysaccharide metabolic process	up	5	197	0.00
A992422 vs. A992422×QL12	GO:0044262	Biological process	cellular carbohydrate metabolic process	up	5	381	0.03
A992422 vs. A992422×QL12	GO:0016758	Molecular function	transferase activity, transferring hexosyl groups	up	6	427	0.03
A992422 vs. A992422×QL12	GO:0016757	Molecular function	transferase activity, transferring glycosyl groups	up	6	529	0.04
A992422 vs. A992422×QL12	GO:0008152	Biological process	metabolic process	down	36	8076	0.00

Table 27: Continued

Comparison	GO term	GO type	GO name	Regulation[a]	Qnum	B/Rnum	FDR *p*-value
A992422 vs. A992422×QL12	GO:0006629	Biological process	lipid metabolic process	down	8	681	0.01
A992422 vs. A992422×QL12	GO:0055114	Biological process	oxidation reduction	down	11	1485	0.03
A992422 vs. A992422×QL12	GO:0008610	Biological process	lipid biosynthetic process	down	5	340	0.03
A992422 vs. A992422×QL12	GO:0016614	Molecular function	oxidoreductase activity, acting on CH-OH group of donors	down	6	264	0.01
A992422 vs. A992422×QL12	GO:0016491	Molecular function	oxidoreductase activity	down	13	1669	0.02
A992422 vs. A992422×QL12	GO:0050662	Molecular function	coenzyme binding	down	6	411	0.02
A992422 vs. A992422×QL12	GO:0003824	Molecular function	catalytic activity	down	33	7817	0.02
A992422 vs. A992422×QL12	GO:0048037	Molecular function	cofactor binding	down	6	527	0.04
A992422 vs. A992422×QL36	GO:0008610	Biological process	lipid biosynthetic process	down	6	340	0.00
A992422 vs. A992422×QL36	GO:0006629	Biological process	lipid metabolic process	down	7	681	0.01
A992422 vs. A992422×R931945-2-2	GO:0008610	Biological process	lipid biosynthetic process	down	5	340	0.00
A992422 vs. A992422×R931945-2-2	GO:0006629	Biological process	lipid metabolic process	down	5	681	0.02
A992422 vs. A992422×R931945-2-2	GO:0055114	Biological process	oxidation reduction	down	7	1485	0.02

Table 27: Continued

Comparison	GO term	GO type	GO name	Regulation[a]	Qnum	B/Rnum	FDR *p*-value
A992422 vs. A992422×R931945-2-2	GO:0003824	Molecular function	catalytic activity	down	17	7817	0.02
A992422 vs. A992422×R931945-2-2	GO:0016491	Molecular function	oxidoreductase activity	down	7	1669	0.02
ATx623 vs. ATx623×RTx2783	GO:0006412	Biological process	translation	up	5	431	0.00
ATx623 vs. ATx623×RTx2783	GO:0008610	Biological process	lipid biosynthetic process	down	6	340	0.00
ATx623 vs. ATx623×RTx2783	GO:0044283	Biological process	small molecule biosynthetic process	down	5	253	0.00
ATx623 vs. ATx623×RTx2783	GO:0006629	Biological process	lipid metabolic process	down	6	681	0.02
ATx623 vs. ATx623×RTx436	GO:0044281	Biological process	small molecule metabolic process	down	6	720	0.00

[a]Up refers to genes whose expression was up-regulated in the A_1-line, whereas down refers to genes whose expression was down-regulated in the A_1-line.

Table 28: Pathways enriched in DEGs based on KOBAS in comparison of A_1-lines and F_1 hybrids. Input number indicates the total number of DEGs assigned to the particular pathway. Background indicates the number of non-differentially expressed genes mapped to the particular pathway based on the KEGG pathway database. DEG-enriched pathways were determined based on an FDR corrected *p*-value <0.05.

Comparison	Pathway	Regulation[a]	Database	Number of DEGs	Number of non-DEGs	FDR *p*-value
AQL33 vs. AQL33×QL12	Protein processing in endoplasmic reticulum	up	KEGG PATHWAY	6	200	0.00
AQL33 vs. AQL33×QL12	Glutathione metabolism	down	KEGG PATHWAY	1	120	0.01
AQL33 vs. AQL33×QL36	Protein processing in endoplasmic reticulum	up	KEGG PATHWAY	8	200	0.00
AQL33 vs. AQL33×QL36	Endocytosis	up	KEGG PATHWAY	3	163	0.01
AQL33 vs. AQL33×QL36	Spliceosome	up	KEGG PATHWAY	3	207	0.01
AQL33 vs. AQL33×QL36	Linoleic acid metabolism	up	KEGG PATHWAY	1	12	0.03
AQL33 vs. AQL33×QL36	Plant-pathogen interaction	up	KEGG PATHWAY	2	159	0.04
AQL33 vs. AQL33×R931945-2-2	Protein processing in endoplasmic reticulum	up	KEGG PATHWAY	9	200	0.00
AQL33 vs. AQL33×R931945-2-2	Plant-pathogen interaction	up	KEGG PATHWAY	3	159	0.03
AQL33 vs. AQL33×R931945-2-2	Endocytosis	up	KEGG PATHWAY	3	163	0.03
AQL33 vs. AQL33×R931945-2-2	Flavone and flavonol biosynthesis	up	KEGG PATHWAY	1	8	0.04
AQL33 vs. AQL33×R931945-2-2	Galactose metabolism	up	KEGG PATHWAY	2	71	0.04
AQL33 vs. AQL33×R931945-2-2	Ribosome	down	KEGG PATHWAY	10	348	0.00
AQL33 vs. AQL33×R931945-2-2	Isoquinoline alkaloid biosynthesis	down	KEGG PATHWAY	2	30	0.05
AQL33 vs. AQL33×R931945-2-2	Glutathione metabolism	down	KEGG PATHWAY	3	120	0.05

.

Table 28: Continued

Comparison	Pathway	Regulation[a]	Database	Number of DEGs	Number of non-DEGs	FDR *p*-value
A992422 vs A992422×QL12	Photosynthesis - antenna proteins	down	KEGG PATHWAY	4	15	0.00
A992422 vs A992422×QL12	Photosynthesis	down	KEGG PATHWAY	4	92	0.00
A992422 vs A992422×QL36	Protein processing in endoplasmic reticulum	down	KEGG PATHWAY	5	200	0.00
A992422 vsA992422×R931945-2-2	Linoleic acid metabolism	up	KEGG PATHWAY	1	12	0.00
ATx623 vs ATx623×RTx2783	RNA transport	up	KEGG PATHWAY	2	157	0.01
ATx623 vs ATx623×RTx2783	Linoleic acid metabolism	up	KEGG PATHWAY	1	12	0.01
ATx623 vs ATx623×RTx2783	Ribosome	up	KEGG PATHWAY	2	348	0.01
ATx623 vs ATx623×RTx2783	Flavonoid biosynthesis	down	KEGG PATHWAY	3	63	0.00
ATx623 vs ATx623×RTx2783	Photosynthesis-antenna proteins	down	KEGG PATHWAY	2	15	0.00
ATx623 vs ATx623×RTx2783	Circadian rhythm - plant	down	KEGG PATHWAY	2	43	0.01
ATx623 vs ATx623×RTx436	Phenylalanine metabolism	up	KEGG PATHWAY	2	50	0.04
ATx623 vs ATx623×RTx436	Protein processing in endoplasmic reticulum	down	KEGG PATHWAY	4	200	0.01
A992422 vs. A992422×QL12	Photosynthesis - antenna proteins	down	KEGG PATHWAY	4	15	0.00
A992422 vs. A992422×QL12	Photosynthesis	down	KEGG PATHWAY	4	92	0.00
A992422 vs. A992422×QL36	Protein processing in endoplasmic reticulum	down	KEGG PATHWAY	5	200	0.00
A992422 vs. A992422×R931945-2-2	Linoleic acid metabolism	up	KEGG PATHWAY	1	12	0.00

Table 28: Continued

Comparison	Pathway	Regulation[a]	Database	Number of DEGs	Number of non-DEGs	FDR *p*-value
ATx623 vs. ATx623×RTx2783	RNA transport	up	KEGG PATHWAY	2	157	0.01
ATx623 vs. ATx623×RTx2783	Linoleic acid metabolism	up	KEGG PATHWAY	1	12	0.01
ATx623 vs. ATx623×RTx2783	Ribosome	up	KEGG PATHWAY	2	348	0.01
ATx623 vs. ATx623×RTx2783	Flavonoid biosynthesis	down	KEGG PATHWAY	3	63	0.00
ATx623 vs. ATx623×RTx2783	Photosynthesis-antenna proteins	down	KEGG PATHWAY	2	15	0.00
ATx623 vs. ATx623×RTx2783	Circadian rhythm - plant	down	KEGG PATHWAY	2	43	0.01
ATx623 vs. ATx623×RTx436	Phenylalanine metabolism	up	KEGG PATHWAY	2	50	0.04
ATx623 vs. ATx623×RTx436	Protein processing in endoplasmic reticulum	down	KEGG PATHWAY	4	200	0.01

[a]Up refers to genes whose expression was up-regulated in the A_1-line, whereas down refers to genes whose expression was down-regulated in the A_1-line.

DEGs encoding transcription factors

Transcription factors (TFs) are essential for the regulation of gene expression (Yu, et al., 2003). Our results showed that, of the significant DEGs, there were 28 genes encoding transcription factors. These TFs showed up- and down-regulation in different lines (Table 29). Examples of differentially expressed TFs included myeloblastosis (*MYB*), basic helix-loop-helix family (*bHLH*), and AT-hook-containing TFs (*AHLs*). Six out of 7 DE *MYB* TFs (*Sobic.002G196000*, *Sobic.003G034300*, *Sobic.006G115200*, *Sobic.006G192100*, *Sobic.007G047400*, and *Sobic.007G132600*) showed up-regulation in F_1 hybrids and B-lines compared to A_1-lines, with only one *MYB* TF (*Sobic.004G070900*) showing higher expression (2-fold change) in the A_1-line A992422 in comparison with its F_1 hybrid i.e. A992422×QL12. Plant MYB proteins comprise a large and multifunctional family and play important roles in many plant processes such as pollen development (Higginson, et al., 2003; Zhang, et al., 2007; Dubos, et al., 2010).

Table 29: Differentially expressed transcription factors in different A_1-lines, B-lines, and F_1 hybrids. An FDR corrected *p*-value <0.05 and the absolute value of $\log_2$Ratio ≥1 were used as the thresholds to select significant DEGs.

FeatureID	Comparison	Annotation	Fold change[a]	FDR *p*-value	Gene ortholog	
					Arabidopsis	**Maize**
Sobic.001G054800	**AQL33 vs. AQL33×QL12**	N/A	2.05	0.00	*AT2G30620*	*GRMZM2G003002*
Sobic.001G243000	**A992422 vs. A992422×QL36**	similar to HSF-type DNA-binding domain containing protein, expressed	18.77	0.00	*AT3G22830*	*GRMZM2G010871*
Sobic.001G386000	**AQL33 vs. BQL33**	similar to Putative uncharacterized protein	2.57	0.01	*AT5G18270*	*GRMZM2G018436*
Sobic.001G468400	**AQL33 vs. BQL33**	similar to Homeobox domain containing protein, expressed	2.57	0.00	*AT2G18550*	*GRMZM2G005624*
Sobic.001G468400	**AQL33 vs. AQL33×QL36**	similar to Homeobox domain containing protein, expressed	2.66	0.00	*AT2G18550*	*GRMZM2G005624*
Sobic.002G087500	**A992422 vs. A992422×QL12**	Predicted protein	-3.65	0.00	N/A	N/A
Sobic.002G242000	**A992422 vs. A992422×QL12**	similar to Photosystem I reaction center subunit V, chloroplast precursor	3.07	0.00	*AT1G55670*	*GRMZM2G329047*
Sobic.002G324200	**AQL33 vs. AQL33×R931945-2-2**	similar to Protein translation factor SUI1 homolog	-2.91	0.00	*AT1G54290*	*GRMZM2G017966*
Sobic.003G018700	**AQL33 vs. AQL33xR931945-2-2**	similar to Transcription factor PCF5	2.58	0.00	*AT3G15030*	*AC205574.3_FG006*
Sobic.003G061400	**AQL33 vs. BQL33**	similar to ESTs C26093	2.72	0.00	*AT2G40200*	*GRMZM2G016039*
Sobic.004G070900	**A992422 vs. A992422×QL12**	similar to MYB transcription factor TaMYB1	-2.17	0.03	*AT5G67300*	*GRMZM2G050550*
Sobic.005G018500	**A992422 vs. B992422**	similar to NAC domain transcription factor, putative, expressed	2.57	0.03	*AT1G69490*	*GRMZM2G042494*

Table 29: Continued

FeatureID	Comparison	Annotation	Fold change[a]	FDR *p*-value	Gene ortholog	
					Arabidopsis	**Maize**
Sobic.005G018500	**AQL33 vs. BQL33**	similar to NAC domain transcription factor, putative, expressed	4.72	0.00	*AT1G69490*	*GRMZM2G042494*
Sobic.005G019800	**AQL33 vs. AQL33×R931945-2-2**	similar to ZF-HD protein dimerisation region containing protein, expressed	-2.67	0.03	*AT3G28917*	*GRMZM2G172586*
Sobic.005G019800	**A992422 vs. B992422**	similar to ZF-HD protein dimerisation region containing protein, expressed	2.07	0.00	*AT3G28917*	*GRMZM2G172586*
Sobic.005G019800	**AQL33 vs. BQL33**	similar to ZF-HD protein dimerisation region containing protein, expressed	4.50	0.00	*AT3G28917*	*GRMZM2G172586*
Sobic.006G115200	**ATx623 vs. ATx623×RTx436**	similar to H0418A01.1 protein	-3.24	0.00	*AT5G56110*	*GRMZM2G173633*
Sobic.006G115200	**A992422 vs. A992422×QL12**	similar to H0418A01.1 protein	12.71	0.02	*AT5G56110*	*GRMZM2G173633*
Sobic.007G000500	**AQL33 vs. AQL33×R931945-2-2**	similar to HMG type nucleosome/chromatin assembly factor D	3.09	0.00	*AT5G23420*	*GRMZM2G125648*
Sobic.007G132600	**AQL33 vs. AQL33×QL36**	N/A	16.05	0.00	*AT1G66230*	*GRMZM2G055158*
Sobic.007G132600	**ATx623 vs. ATx623×RTx2783**	N/A	85.22	0.00	*AT1G66230*	*GRMZM2G055158*
Sobic.008G021800	**AQL33 vs. AQL33×QL36**	similar to NAC domain transcription factor, putative, expressed	2.13	0.00	*AT1G61110*	N/A
Sobic.008G021800	**AQL33 vs. BQL33**	similar to NAC domain transcription factor, putative, expressed	3.38	0.00	*AT1G61110*	N/A

Table 29 : Continued

FeatureID	Comparison	Annotation	Fold change[a]	FDR *p*-value	Gene ortholog	
					Arabidopsis	**Maize**
Sobic.009G016300	**AQL33 vs. BQL33**	similar to C2H2 type zinc finger transcription factor ZFP16	5.37	0.00	*AT5G59820*	*AC206217.2_FG006*
Sobic.009G184400	**A992422 vs. A992422×QL12**	similar to Ethylene-responsive factor-like transcription factor ERFL2a	-2.00	0.01	*AT5G44210*	*GRMZM2G132185*
Sobic.009G195000	**A992422 vs. A992422×QL36**	similar to Putative uncharacterized protein	-3.10	0.00	*AT5G60970*	*GRMZM2G035944*
Sobic.010G269700	**A992422 vs. A992422×QL36**	similar to Putative uncharacterized protein	-2.26	0.00	*AT5G65640*	*GRMZM2G128807*
Sobic.010G269700	**AQL33 vs. AQL33×R931945-2-2**	similar to Putative uncharacterized protein	2.33	0.00	*AT5G65640*	*GRMZM2G128807*

[a]Positive values indicate greater expression in B-lines or F_1 hybrids compared to A_1-lines. Negative values indicate greater expression in A_1-lines compared to B-lines or F_1 hybrids.

Differentially expressed heat shock protein genes

In this study, we also found other DEGs potentially related to male-sterility in sorghum. There were 11 DEGs related to heat shock proteins among which, three (*Sobic.010G149300*, *Sobic.003G039400*, and *Sobic.003G082300*) were only up-regulated in A_1-line AQL33 compared to its F_1 hybrids (Table 30). Heat shock proteins have been considered as the hallmark of abiotic stress and play a key role in plant stress responsiveness (Kotak, et al., 2007). Our results were similar to Fujii, et al. (2010) and Wang, et al. (2016) who reported the association of heat shock proteins to CMS in rice and cabbage, respectively. Kuzmin, et al. (2004) showed that mitochondrial deficiencies in maize mitochondrial mutants (NCS2, NCS4, and NCS6) lead to constitutive expression of genes for heat shock proteins. In CMS, layers of interaction between mitochondrial and nuclear genes control male specificity (Chen and Liu, 2014). CMS proteins can cause mitochondrial deficiency; in response, mitochondrial retrograde signals can increase the expression of heat shock genes (Woodson and Chory, 2008; Chen and Liu, 2014).

Table 30: Heat shock proteins genes differentially expressed in different A_1-lines, B-lines, and F_1 hybrids. An FDR corrected *p*-value <0.05 and the absolute value of $\log_2$Ratio ≥1 were used as the thresholds to select significant DEGs.

Feature ID	Comparison	Annotation	Fold change[a]	FDR *p*-value	Gene ortholog	
					Arabidopsis	Maize
Sobic.001G420100	**AQL33 vs. AQL33×R931945-2-2**	similar to Heat shock cognate 70 kDa protein	-10.35	0.00	*AT3G12580*	*AC194017.3_FG001*
Sobic.002G243500	**AQL33 vs. AQL33×QL36**	similar to Heat shock protein 81-2	-2.94	0.00	*AT5G56000*	*GRMZM2G012631*
Sobic.003G039400	**AQL33 vs. AQL33×QL36**	similar to 17.8 kDa class II heat shock protein	-13.23	0.00	*AT5G12020*	*GRMZM2G012455*
Sobic.003G039400	**AQL33 vs. AQL33×R931945-2-2**	similar to 17.8 kDa class II heat shock protein	-3.73	0.00	*AT5G12020*	*GRMZM2G012455*
Sobic.003G039400	**A992422 vs. A992422×QL12**	similar to 17.8 kDa class II heat shock protein	-2.88	0.00	*AT5G12020*	*GRMZM2G012455*
Sobic.003G039400	**AQL33 vs. AQL33×QL12**	similar to 17.8 kDa class II heat shock protein	-2.49	0.00	*AT5G12020*	*GRMZM2G012455*
Sobic.003G081900	**AQL33 vs. AQL33×QL36**	similar to Heat shock protein 17.2	-15.58	0.00	*AT1G53540*	*AC208204.3_FG006*
Sobic.003G081900	**AQL33 vs. AQL33×R931945-2-2**	similar to Heat shock protein 17.2	-9.03	0.00	*AT1G53540*	*AC208204.3_FG006*
Sobic.003G081900	**AQL33 vs. AQL33×QL12**	similar to Heat shock protein 17.2	-2.30	0.00	*AT1G53540*	*AC208204.3_FG006*
Sobic.003G081900	**A992422 vs. A992422×R931945-2-2**	similar to Heat shock protein 17.2	9.18	0.00	*AT1G53540*	*AC208204.3_FG006*

Table 30: Continued

Feature ID	Comparison	Annotation	Fold change[a]	FDR *p*-value	Gene ortholog	
					Arabidopsis	Maize
Sobic.003G350700	**AQL33 vs. AQL33×QL36**	similar to Heat shock 70 kDa protein	-5.30	0.00	*AT3G12580*	*AC194017.3_FG001*
Sobic.003G350700	**AQL33 vs. AQL33×R931945-2-2**	similar to Heat shock 70 kDa protein	-4.40	0.00	*AT3G12580*	*AC194017.3_FG001*
Sobic.003G350700	**AQL33 vs. AQL33×QL12**	similar to Heat shock 70 kDa protein	-2.04	0.00	*AT3G12580*	*AC194017.3_FG001*
Sobic.003G350700	**ATx623 vs. ATx623×RTx436**	similar to Heat shock 70 kDa protein	2.05	0.03	*AT3G12580*	*AC194017.3_FG001*
Sobic.003G350700	**A992422 vs. A992422×QL36**	similar to Heat shock 70 kDa protein	2.33	0.00	*AT3G12580*	*AC194017.3_FG001*
Sobic.004G228900	**AQL33 vs. AQL33×R931945-2-2**	similar to Low molecular weight heat shock protein precursor	-17.98	0.03	*AT5G51440*	*GRMZM2G007729*
Sobic.004G228900	**A992422 vs. A992422×QL36**	similar to Low molecular weight heat shock protein precursor	27.32	0.00	*AT5G51440*	*GRMZM2G007729*
Sobic.006G005600	**AQL33 vs. AQL33×R931945-2-2**	similar to Heat shock protein 82	-12.59	0.00	*AT5G52640*	*GRMZM5G833699*
Sobic.006G005600	**A992422 vs. A992422×QL36**	similar to Heat shock protein 82	5.75	0.00	*AT5G52640*	*GRMZM5G833699*
Sobic.007G216300	**AQL33 vs. AQL33×QL36**	similar to Heat shock protein 81-2	-5.26	0.00	*AT5G56000*	N/A
Sobic.007G216300	**AQL33 vs. AQL33×R931945-2-2**	similar to Heat shock protein 81-2	-2.41	0.00	*AT5G56000*	*GRMZM2G012631*
Sobic.007G216300	**A992422 vs. A992422×QL36**	similar to Heat shock protein 81-2	2.46	0.00	*AT5G56000*	*GRMZM2G012631*
Sobic.008G136000	**AQL33 vs. AQL33×QL36**	similar to Heat shock cognate 70 kDa protein 2	-2.72	0.00	*AT3G12580*	*AC194017.3_FG001*

Table 30: Continued

Feature ID	Comparison	Annotation	Fold change[a]	FDR *p*-value	Gene ortholog	
					Arabidopsis	**Maize**
Sobic.009G163900	**AQL33 vs. AQL33×QL36**	similar to Heat shock cognate 70 kDa protein	-8.33	0.00	*AT3G12580*	*AC194017.3_FG001*
Sobic.009G163900	**AQL33 vs. AQL33×R931945-2-2**	similar to Heat shock cognate 70 kDa protein	-4.08	0.00	*AT3G12580*	*AC194017.3_FG001*
Sobic.009G163900	**AQL33 vs. AQL33×QL12**	similar to Heat shock cognate 70 kDa protein	-2.62	0.00	*AT3G12580*	*AC194017.3_FG001*
Sobic.009G163900	**ATx623 vs. ATx623×RTx436**	similar to Heat shock cognate 70 kDa protein	3.07	0.03	*AT3G12580*	*AC194017.3_FG001*
Sobic.009G163900	**ATx623 vs. ATx623×RTx2783**	similar to Heat shock cognate 70 kDa protein	4.03	0.00	*AT3G12580*	*AC194017.3_FG001*
Sobic.009G163900	**A992422 vs. A992422×QL36**	similar to Heat shock cognate 70 kDa protein	4.58	0.00	*AT3G12580*	N/A

[a]Positive values indicate greater expression in B-lines or F_1 hybrids compared to A_1-lines. Negative values indicate greater expression in A_1-lines compared to B-lines or F_1 hybrids.

C**onclusion**

This study provided an overview of differential gene expression in the A_1 CMS-*Rf* system of sorghum, and provides insight into the complex molecular changes that precede pollen abortion in A_1 CMS panicles. The expression of the transcriptomes of the A_1-lines and their corresponding maintainer lines was very similar as genome sequence differences between these near-isogenic lines are largely confined to the organellar genomes. Most of the DEG-enriched GO terms and pathways were discovered in a comparison of A_1 CMS lines with their F_1 hybrids, and the GO terms reflect the complex changes in panicle development that occur prior to pollen maturation/abortion. This study provides insights into the complex nuclear gene regulation in CMS lines and fertile maintainer and F_1 hybrid lines at an early stage of panicle maturation, and further investigation of multiple stages of panicle development (especially nearer pollen abortion in A_1 CMS lines) are necessary to better understand the nuclear-mitochondrial interactions and regulation of the A1 CMS-*Rf* system in sorghum.

CHAPTER IV

CONCLUSIONS

This project has two major components with each based on transcriptome analyses. In the first study, the transcriptional dynamics were profiled during stem development in two sorghum genotypes, R07020 and BTx623. These two genotypes differ in their maturity and height genes. R07020 is tall (likely 1-dwarf) and photoperiod-sensitive (dominant *Ma*$_1$ and *Ma*$_6$, recessive *ma*$_5$) whereas BTx623 is short (3-dwarf) and photoperiod-insensitive (recessive *ma*$_1$ and *ma*$_6$, dominant *Ma*$_5$). In the temperate Texas production environment, R07020 is grown as a biomass feedstock and BTx623 is used as a parent for either grain or forage hybrids. In the second study, transcriptome analysis of immature panicles was conducted to characterize the expression of the nuclear genome in a set of CMS lines (A_1-lines), maintainer lines (B-lines), and F_1 hybrids.

By employing the comprehensive transcriptome analyses of R07020 and BTx623, differentially expressed genes were identified along the sorghum stem that matures basipetally. Monocot stems provide a unique developmental system with intercalary meristems present from the immature shoot apex to the mature nodes at the stem base. A comparison of gene expression in nodal regions across the stem provided a profile of those genes that are differentially expressed as the stem matures. This included a range of differentially expressed genes (DEGs) from the apex where cells are rapidly dividing and initiating elongation to the non-elongating basal region where cellular processes associated with stem maturation occur. Numerous differences punctuate the cellular

events occurring in the immature and more mature regions of the stem. The stage of cell wall deposition differs prominently between the shoot apex where primary cell walls are being deposited and the shoot base where secondary wall maturation is occurring. Apart from the genes directly involved in lignin and cellulose biosynthesis, many genes related to cell wall deposition, particularly transcription factors, glucosyl transferases, and cell wall proteins were differentially expressed along the stem. The preferential expression of cell wall related genes differed between the two genotypes, which likely reflected the different developmental stages that existed after 60 days of growth in the two genotypes. This study provided fundamental information about cell wall biosynthesis genes in sorghum. A better understanding of cell wall biosynthesis may allow for breeding plants with increased efficiency of biomass conversion to biofuel. Specifically, there is promise for producing improved genotypes with altered lignin content (the major limiting factor in conversion) and altered lignin composition with the least side effects on the other fitness traits of the plant.

In the second study, a series of DEGs were identified in CMS A_1-lines, iso-cytoplasmic maintainer B-lines, and F_1 sorghum hybrids of immature panicles from the early boot stage (cream-colored florets, wispy panicles). Functional terms for these DEGs were significantly enriched in a series of Gene Ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways. The nuclear transcriptomes of A_1-lines and maintainer lines were very similar as genome sequence differences between these near-isogenic lines are largely confined to organelles. Most of the DEG-enriched GO terms and pathways were discovered by comparing A_1 CMS lines with their

F_1 hybrids. These terms reflect the complex changes in panicle development that occur prior to pollen maturation/abortion. Examination of the DEGs (and their homologs) within these terms provided evidence that the down-regulated genes in A_1-lines may relate to the eventual abortion of pollen that will occur within these panicles. These genes are involved in processes including lipid biosynthesis, anther and pollen development, flavonoid biosynthesis, and the glutathione metabolism pathway or were a member of the oxidoreductase gene family. These results suggest that these genes, including those from a large number of metabolic processes, are intimately linked to the production of viable pollen. The up-regulation of some stress related genes and pathways was found in CMS lines, including heat shock protein genes and genes involved in the plant-pathogen interaction pathway. The deleterious effects of CMS proteins might trigger stress signals in plants and activate plant defense metabolism. This study provided insight into the complex nuclear gene regulation in CMS lines and fertile maintainer and F_1 hybrid lines at an early stage of panicle maturation. Further investigation of multiple stages of panicle development (especially near pollen abortion in A_1 CMS lines) is necessary to better understand the nuclear-mitochondrial interactions and regulation of the A_1 CMS-*Rf* system in sorghum. Proteomics analysis may also help to elucidate the effects of CMS and *Rf* genes at the post-transcriptional level. This study provides a fundamental contribution to the knowledge of CMS and *Rf* genes that may allow for the development of more effective strategies for breeding CMS A-lines and R-lines. Such strategies are critical to the hybrid seed industry.

REFERENCES

Adams-Phillips, L., A.G. Briggs and A.F. Bent. 2010. Disruption of poly (ADP-ribosyl) ation mechanisms alters responses of *Arabidopsis* to biotic stress. Plant Physiology 152: 267-280.

Albersheim, P., A. Darvill, K. Roberts, R. Sederoff and A. Staehelin. 2010. Plant cell walls. Garland Science. New York. p. 430.

Allen, J.O., C.M. Fauron, P. Minx, L. Roark, S. Oddiraju, G.N. Lin, et al. 2007. Comparisons among two fertile and three male-sterile mitochondrial genomes of maize. Genetics 177: 1173-1192.

Allocco, D.J., I.S. Kohane and A.J. Butte. 2004. Quantifying the relationship between co-expression, co-regulation and gene function. BMC Bioinformatics 5: 1-10.

Alqurashi, M., C. Gehring and C. Marondedze. 2016. Changes in the *Arabidopsis thaliana* proteome implicate cAMP in biotic and abiotic stress responses and changes in energy Metabolism. International Journal of Molecular Sciences doi:10.3390/ijms17060852.

Alvira, P., E. Tomás-Pejó, M. Ballesteros and M. Negro. 2010. Pretreatment technologies for an efficient bioethanol production process based on enzymatic hydrolysis: a review. Bioresource Technology 101: 4851-4861.

An, H., Z. Yang, B. Yi, J. Wen, J. Shen, J. Tu, et al. 2014. Comparative transcript profiling of the fertile and sterile flower buds of *pol* CMS in *Brassica napus*. BMC Genomics 15: 258-267.

Anderson, N.A. and C. Chapple. 2014. Perturbing lignin biosynthesis: metabolic changes in response to manipulation of the phenylpropanoid pathway. Recent Advances in Polyphenol Research 4: 39-59.

Andersson-Gunnerås, S., E.J. Mellerowicz, J. Love, B. Segerman, Y. Ohmiya, P.M. Coutinho, et al. 2006. Biosynthesis of cellulose-enriched tension wood in *Populus*: global analysis of transcripts and metabolites identifies biochemical and developmental regulators in secondary wall biosynthesis. The Plant Journal 45: 144-165.

Appenzeller, L., M. Doblin, R. Barreiro, H. Wang, X. Niu, K. Kollipara, et al. 2004. Cellulose synthesis in maize: isolation and expression analysis of the cellulose synthase (*CesA*) gene family. Cellulose 11: 287-299.

Ariizumi, T., K. Hatakeyama, K. Hinata, R. Inatsugi, I. Nishida, S. Sato, et al. 2004. Disruption of the novel plant protein NEF1 affects lipid accumulation in the plastids of the tapetum and exine formation of pollen, resulting in male sterility in *Arabidopsis thaliana*. The Plant Journal 39: 170-181.

Barkan, A., M. Rojas, S. Fujii, A. Yap, Y.S. Chong, C.S. Bond, et al. 2012. A combinatorial amino acid code for RNA recognition by pentatricopeptide repeat proteins. PLoS Genet doi:org/10.1371/journal.pgen.1002910.

Barkan, A. and I. Small. 2014. Pentatricopeptide repeat proteins in plants. Annual Review of Plant Biology 65: 415-442.

Barr, C.M. and L. Fishman. 2010. The nuclear component of a cytonuclear hybrid incompatibility in *Mimulus* maps to a cluster of pentatricopeptide repeat genes. Genetics 184: 455-465.

Beaudoin, F., X. Wu, F. Li, R.P. Haslam, J.E. Markham, H. Zheng, et al. 2009. Functional characterization of the *Arabidopsis β*-ketoacyl-coenzyme A reductase candidates of the fatty acid elongase. Plant Physiology 150: 1174-1191.

Beisson, F., Y. Li, G. Bonaventure, M. Pollard and J.B. Ohlrogge. 2007. The acyltransferase *GPAT5* is required for the synthesis of suberin in seed coat and root of *Arabidopsis*. The Plant Cell 19: 351-368.

Bentolila, S., A.A. Alfonso and M.R. Hanson. 2002. A pentatricopeptide repeat-containing gene restores fertility to cytoplasmic male-sterile plants. Proceedings of the National Academy of Sciences 99: 10887-10892.

Bibi, N., S. Yuan, Y. Zhu and X. Wang. 2014. Improvements of fertility restoration in cytoplasmic male sterile cotton by enhanced expression of glutathione S-transferase (*GST*) gene. Journal of Plant Growth Regulation 33: 420-429.

Blée, E. 2002. Impact of phyto-oxylipins in plant defense. Trends in Plant Science 7: 315-322.

Boerjan, W., J. Ralph and M. Baucher. 2003. Lignin biosynthesis. Annual Review of Plant Biology 54: 519-546.

Bosch, M., C.-D. Mayer, A. Cookson and I.S. Donnison. 2011. Identification of genes involved in cell wall biogenesis in grasses by differential gene expression profiling of elongating and non-elongating maize internodes. Journal of Experimental Botany 62: 3545-3561.

Breton, C., S. Fournel-Gigleux and M.M. Palcic. 2012. Recent structures, evolution and mechanisms of glycosyltransferases. Current Opinion in Structural Biology 22: 540-549.

Broadhead, D. 1972. Registration of Rio Sweet Sorghum1 (Reg. No. 113). Crop Science 12: 716-716.

Broadhead, D. 1982. Registration of Keller Sweet Sorghum1 (Reg. No. 120). Crop Science 22: 1263-1263.

Brooking, I.R. 1976. Male sterility in *Sorghum bicolor* (L.) Moench induced by low night temperature. I. timing of the stage of sensitivity. Functional Plant Biology 3: 589-596.

Brooking, I.R. 1979. Male sterility in *Sorghum bicolor* (L.) Moench induced by low night temperature. II. genotypic differences in sensitivity. Functional Plant Biology 6: 143-147.

Brown, D.M., L.A.H. Zeef, J. Ellis, R. Goodacre and S.R. Turner. 2005. Identification of novel genes in *Arabidopsis* involved in secondary cell wall formation using expression profiling and reverse genetics. The Plant Cell Online 17: 2281-2295.

Brown, G.G., N. Formanová, H. Jin, R. Wargachuk, C. Dendy, P. Patil, et al. 2003. The radish *Rfo* restorer gene of Ogura cytoplasmic male sterility encodes a protein with multiple pentatricopeptide repeats. The Plant Journal 35: 262-272.

Browning, B. 1963. The composition and chemical reactions of wood, In: B. Browning, editor The Chemistry of Wood. John Wiley & Sons. New York. p. 58-101.

Bu, Q., H. Jiang, C.-B. Li, Q. Zhai, J. Zhang, X. Wu, et al. 2008. Role of the *Arabidopsis thaliana NAC* transcription factors *ANAC019* and *ANAC055* in regulating jasmonic acid-signaled defense responses. Cell research 18: 756-767.

Burow, G., R. Klein, C. Franks, P. Klein, K. Schertz, G. Pederson, et al. 2011. Registration of the BTx623/IS3620C recombinant inbred mapping population of sorghum. Journal of Plant Registrations 5: 141-145.

Canonne, J., S. Froidure-Nicolas and S. Rivas. 2011. Phospholipases in action during plant defense signaling. Plant Signaling & Behavior 6: 13-18.

Carlsson, J. and K. Glimelius. 2011. Cytoplasmic male-sterility and nuclear encoded fertility restoration, In: F. Kempken, editor Plant Mitochondria. Springer New York. p. 469-491.

Carlsson, J., M. Leino, J. Sohlberg, J.F. Sundstroem and K. Glimelius. 2008. Mitochondrial regulation of flower development. Mitochondrion 8: 74-86.

Carroll, A. and C. Somerville. 2009. Cellulosic biofuels. Annual Review of Plant Biology 60: 165-182.

Cassab, G.I. 1998. Plant cell wall proteins. Annual Review of Plant Biology 49: 281-309.

Cassan-Wang, H., N. Goué, M.N. Saidi, S. Legay, P. Sivadon, D. Goffner, et al. 2013. Identification of novel transcription factors regulating secondary cell wall formation in *Arabidopsis*. Frontiers in Plant Science doi:org/10.3389/fpls.2013.00189.

Chase, C.D. 2007. Cytoplasmic male sterility: a window to the world of plant mitochondrial-nuclear interactions. Trends in Genetics 23: 81-90.

Chen, L. and Y.-G. Liu. 2014. Male sterility and fertility restoration in crops. Annual Review of Plant Biology 65: 579-606.

Chen, P., S. Ran, R. Li, Z. Huang, J. Qian, M. Yu, et al. 2014. Transcriptome *de novo* assembly and differentially expressed genes related to cytoplasmic male sterility in kenaf (*Hibiscus cannabinus* L.). Molecular Breeding 34: 1879-1891.

Christiansen, M.W., P.B. Holm and P.L. Gregersen. 2011. Characterization of barley (*Hordeum vulgare* L.) *NAC* transcription factors suggests conserved functions compared to both monocots and dicots. BMC Research Notes 4: 302-314.

Courtial, A., M. Soler, A.-L. Chateigner-Boutin, M. Reymond, V. Méchin, H. Wang, et al. 2013. Breeding grasses for capacity to biofuel production or silage feeding value: an updated list of genes involved in maize secondary cell wall biosynthesis and assembly. Maydica 58: 67-102.

Dalal, M., K. Mayandi and V. Chinnusamy. 2012. Sorghum: Improvement of abiotic stress tolerance. Improving Crop Resistance to Abiotic Stress, Volume 1 & Volume 2: 923-950.

de Azevedo Souza, C., S.S. Kim, S. Koch, L. Kienow, K. Schneider, S.M. McKim, et al. 2009. A novel fatty acyl-CoA synthetase is required for pollen development and sporopollenin biosynthesis in *Arabidopsis*. The Plant Cell 21: 507-525.

De Storme, N. and D. Geelen. 2014. The impact of environmental stress on male reproductive development in plants: biological processes and molecular mechanisms. Plant, Cell & Environment 37: 1-18.

Delannoy, E., W. Stanley, C. Bond and I. Small. 2007. Pentatricopeptide repeat (PPR) proteins as sequence-specificity factors in post-transcriptional processes in organelles. Biochemical Society Transactions 35: 1643-1647.

Demirbas, M.F. 2009. Biorefineries for biofuel upgrading: a critical review. Applied Energy 86: 151-161.

Desloire, S., H. Gherbi, W. Laloui, S. Marhadour, V. Clouet, L. Cattolico, et al. 2003. Identification of the fertility restoration locus, *Rfo*, in radish, as a member of the pentatricopeptide-repeat protein family. EMBO Reports 4: 588-594.

Diamond, M. and P.F. McCabe. 2011. Mitochondrial regulation of plant programmed cell death, In: F. Kempken, editor Plant Mitochondria. Springer New York. p. 439-465.

Dixon, R.A., L. Achnine, P. Kota, C.J. Liu, M. Reddy and L. Wang. 2002. The phenylpropanoid pathway and plant defence—a genomics perspective. Molecular Plant Pathology 3: 371-390.

Dobritsa, A.A., Z. Lei, S.-i. Nishikawa, E. Urbanczyk-Wochniak, D.V. Huhman, D. Preuss, et al. 2010. *LAP5* and *LAP6* encode anther-specific proteins with similarity to chalcone synthase essential for pollen exine development in *Arabidopsis thaliana*. Plant Physiology doi:org/10.1104/pp.110.157446.

Dobritsa, A.A., S.-I. Nishikawa, D. Preuss, E. Urbanczyk-Wochniak, L.W. Sumner, A. Hammond, et al. 2009a. *LAP3*, a novel plant protein required for pollen development, is essential for proper exine formation. Sexual Plant Reproduction 22: 167-177.

Dobritsa, A.A., J. Shrestha, M. Morant, F. Pinot, M. Matsuno, R. Swanson, et al. 2009b. CYP704B1 is a long-chain fatty acid ω-hydroxylase essential for sporopollenin synthesis in pollen of *Arabidopsis*. Plant Physiology 151: 574-589.

Doering, A., R. Lathe and S. Persson. 2012. An update on xylan synthesis. Molecular Plant 5: 769-771.

Downes, R.W. and D.R. Marshall. 1971. Low temperature induced male sterility in *Sorghum bicolor*. Australian Journal of Experimental Agriculture 11: 352-356.

Du, K., Q. Liu, X. Wu, J. Jiang, J. Wu, Y. Fang, et al. 2016. Morphological structure and transcriptome comparison of the cytoplasmic male sterility line in *Brassica napus* (SaNa-1A) derived from somatic hybridization and its maintainer line SaNa-1B. Frontiers in Plant Science doi:10.3389/fpls.2016.01313.

Du, Z., X. Zhou, Y. Ling, Z. Zhang and Z. Su. 2010. agriGO: a GO analysis toolkit for the agricultural community. Nucleic Acids Research 38 Suppl: W64-70.

Dubos, C., R. Stracke, E. Grotewold, B. Weisshaar, C. Martin and L. Lepiniec. 2010. *MYB* transcription factors in *Arabidopsis*. Trends in Plant Science 15: 573-581.

Dykes, L., L. Hoffmann, O. Portillo-Rodriguez, W.L. Rooney and L.W. Rooney. 2014. Prediction of total phenols, condensed tannins, and 3-deoxyanthocyanidins in sorghum grain using near-infrared (NIR) spectroscopy. Journal of Cereal Science 60: 138-142.

Eckardt, N.A. 2011. Retrograde signaling: a new candidate signaling molecule. Plant Cell 23: 3870-3870.

Erichsen, A.W. and J.G. Ross. 1963. Inheritance of colchicine-induced male sterility in sorghum. Crop Science 3: 335-338.

Escamilla-Treviño, L.L., W. Chen, M.L. Card, M.-C. Shih, C.-L. Cheng and J.E. Poulton. 2006. *Arabidopsis thaliana* β-Glucosidases BGLU45 and BGLU46 hydrolyse monolignol glucosides. Phytochemistry 67: 1651-1660.

Fagard, M., T. Desnos, T. Desprez, F. Goubet, G. Refregier, G. Mouille, et al. 2000. *PROCUSTE1* encodes a cellulose synthase required for normal cell elongation specifically in roots and dark-grown hypocotyls of *Arabidopsis*. The Plant Cell Online 12: 2409-2423.

Farajollahi, S. and S. Maas. 2010. Molecular diversity through RNA editing: a balancing act. Trends in Genetics 26: 221-230.

Ferreira, S.S., C.T. Hotta, V.G. de Carli Poelking, D.C.C. Leite, M.S. Buckeridge, M.E. Loureiro, et al. 2016. Co-expression network analysis reveals transcription factors associated to cell wall biosynthesis in sugarcane. Plant Molecular Biology 91: 15-35.

Froidure, S., J. Canonne, X. Daniel, A. Jauneau, C. Brière, D. Roby, et al. 2010. *AtsPLA2-α* nuclear relocalization by the *Arabidopsis* transcription factor AtMYB30 leads to repression of the plant defense response. Proceedings of the National Academy of Sciences 107: 15281-15286.

Fu, C., J.R. Mielenz, X. Xiao, Y. Ge, C.Y. Hamilton, M. Rodriguez, et al. 2011. Genetic manipulation of lignin reduces recalcitrance and improves ethanol production from switchgrass. Proceedings of the National Academy of Sciences 108: 3803-3808.

Fujii, S., C.S. Bond and I.D. Small. 2011. Selection patterns on restorer-like genes reveal a conflict between nuclear and mitochondrial genomes throughout angiosperm evolution. Proceedings of the National Academy of Sciences 108: 1723-1728.

Fujii, S., M. Yamada, M. Fujita, E. Itabashi, K. Hamada, K. Yano, et al. 2010. Cytoplasmic–nuclear genomic barriers in rice pollen development revealed by comparison of global gene expression profiles among five independent cytoplasmic male sterile lines. Plant and Cell Physiology 51: 610-620.

Furtado, A., J.S. Lupoi, N.V. Hoang, A. Healey, S. Singh, B.A. Simmons, et al. 2014. Modifying plants for biofuel and biomaterial production. Plant Biotechnology Journal 12: 1246-1258.

Gabay-Laughnan, S. and K.J. Newton. 2005. Mitochondrial mutations in maize. Maydica 50: 349-359.

Gaborieau, L., G.G. Brown and H. Mireau. 2016. The propensity of pentatricopeptide repeat genes to evolve into restorers of cytoplasmic male sterility. Frontiers in Plant Science doi:10.3389/fpls.2016.01816.

Ge, W., Y. Zhang, Z. Cheng, D. Hou, X. Li and J. Gao. 2016. Main regulatory pathways, key genes and microRNAs involved in flower formation and development of moso bamboo (*Phyllostachys edulis*). Plant Biotechnology Journal doi:10.1111/pbi.12593.

Gray, J., D. Caparrós-Ruiz and E. Grotewold. 2012. Grass phenylpropanoids: regulate before using! Plant Science 184: 112-120.

Guillaumie, S., H. San-Clemente, C. Deswarte, Y. Martinez, C. Lapierre, A. Murigneux, et al. 2007. MAIZEWALL. Database and developmental gene expression profiling of cell wall biosynthesis and assembly in maize. Plant Physiology 143: 339-363.

Hammani, K. and P. Giege. 2014. RNA metabolism in plant mitochondria. Trends in Plant Science 19: 380-389.

Handakumbura, P.P. and S.P. Hazen. 2007. Transcriptional regulation of grass secondary cell wall biosynthesis: playing catch-up with *Arabidopsis thaliana*. Current Challenges in Plant Cell Walls 3: 43-47.

Handakumbura, P.P., D.A. Matos, K.S. Osmont, M.J. Harrington, K. Heo, K. Kafle, et al. 2013. Perturbation of *Brachypodium distachyon CELLULOSE SYNTHASE A4* or *7* results in abnormal cell walls. BMC Plant Biology 13: 131-145.

Hanson, M.R. and S. Bentolila. 2004. Interactions of mitochondrial and nuclear genes that affect male gametophyte development. Plant Cell 16 Suppl: S154-169.

He, S., Z.-H. Yu, C.E. Vallejos and S.A. Mackenzie. 1995. Pollen fertility restoration by nuclear gene *Fr* in CMS common bean: an *Fr* linkage map and the mode of *Fr* action. Theoretical and Applied Genetics 90: 1056-1062.

Higginson, T., S.F. Li and R.W. Parish. 2003. *AtMYB103* regulates tapetum and trichome development in *Arabidopsis thaliana*. The Plant Journal 35: 177-192.

Hirano, K., K. Aya, Y. Morinaka, S. Nagamatsu, Y. Sato, B.A. Antonio, et al. 2013. Survey of genes involved in rice secondary cell wall formation through a co-expression network. Plant and Cell Physiology 54: 1803-1821.

Ho-Yue-Kuang, S., C. Alvarado, S. Antelme, B. Bouchet, L. Cézard, P. Le Bris, et al. 2015. Mutation in *Brachypodium caffeic acid O-methyltransferase 6* alters stem and grain lignins and improves straw saccharification without deteriorating grain quality. Journal of Experimental Botany 67: 227-237.

Hodnett, G.L., A.L. Hale, D.J. Packer, D.M. Stelly, J. da Silva and W.L. Rooney. 2010. Elimination of a reproductive barrier facilitates intergeneric hybridization of *Sorghum bicolor* and *Saccharum*. Crop Science 50: 1188-1195.

Horn, R., K.J. Gupta and N. Colombo. 2014. Mitochondrion role in molecular basis of cytoplasmic male sterility. Mitochondrion 19: 198-205.

Hu, J., G. Chen, H. Zhang, Q. Qian and Y. Ding. 2016. Comparative transcript profiling of alloplasmic male-sterile lines revealed altered gene expression related to pollen development in rice (*Oryza sativa* L.). BMC Plant Biology 16: 175-190.

Hu, J., W. Huang, Q. Huang, X. Qin, C. Yu, L. Wang, et al. 2014. Mitochondria and cytoplasmic male sterility in plants. Mitochondrion 19: 282-288.

Hu, J., K. Wang, W. Huang, G. Liu, Y. Gao, J. Wang, et al. 2012. The rice pentatricopeptide repeat protein *RF5* restores fertility in Hong-Lian cytoplasmic male-sterile lines via a complex with the glycine-rich protein GRP162. The Plant Cell 24: 109-122.

Huang, W., C. Yu, J. Hu, L. Wang, Z. Dan, W. Zhou, et al. 2015. Pentatricopeptide-repeat family protein RF6 functions with hexokinase 6 to rescue rice cytoplasmic male sterility. Proceedings of the National Academy of Sciences 112: 14984-14989.

Hussain, A., B.-G. Mun, Q.M. Imran, S.-U. Lee, T.A. Adamu, M. Shahid, et al. 2016. Nitric oxide mediated transcriptome profiling reveals activation of multiple regulatory pathways in *Arabidopsis thaliana*. Frontiers in Plant Science doi:10.3389/fpls.2016.00975.

Iba, K. 2002. Acclimative response to temperature stress in higher plants: approaches of gene engineering for temperature tolerance. Annual Review of Plant Biology 53: 225-245.

Imoto, K., R. Yokoyama and K. Nishitani. 2005. Comprehensive approach to genes involved in cell wall modifications in *Arabidopsis thaliana*. Plant Molecular Biology 58: 177-192.

Ivanov, M.K. and G.M. Dymshits. 2007. Cytoplasmic male sterility and restoration of pollen fertility in higher plants. Russian Journal of Genetics 43: 354-368.

Jiang, P., X. Zhang, Y. Zhu, W. Zhu, H. Xie and X. Wang. 2007. Metabolism of reactive oxygen species in cotton cytoplasmic male sterility and its restoration. Plant Cell Reports 26: 1627-1634.

Jin, H., E. Cominelli, P. Bailey, A. Parr, F. Mehrtens, J. Jones, et al. 2000. Transcriptional repression by *AtMYB4* controls production of UV-protecting sunscreens in *Arabidopsis*. The EMBO Journal 19: 6150-6161.

Jing, B., S. Heng, D. Tong, Z. Wan, T. Fu, J. Tu, et al. 2012. A male sterility-associated cytotoxic protein ORF288 in *Brassica juncea* causes aborted pollen development. Journal of Experimental Botany 63: 1285-1295.

Jones, L., A.R. Ennos and S.R. Turner. 2001. Cloning and characterization of *irregular xylem4* (*irx4*): a severely lignin-deficient mutant of *Arabidopsis*. The Plant Journal 26: 205-216.

Jordan, D., R. Klein, K. Sakrewski, R. Henzell, P. Klein and E. Mace. 2011. Mapping and characterization of *Rf5*: a new gene conditioning pollen fertility restoration in A1 and A2 cytoplasm in sorghum (*Sorghum bicolor* (L.) Moench). Theoretical and Applied Genetics 123: 383-396.

Jordan, D.R., E.S. Mace, R.G. Henzell, P.E. Klein and R.R. Klein. 2010. Molecular mapping and candidate gene identification of the *Rf2* gene for pollen fertility restoration in sorghum [*Sorghum bicolor* (L.) Moench]. Theoretical and Applied Genetics 120: 1279-1287.

Joshi, C.P. and S.D. Mansfield. 2007. The cellulose paradox—simple molecule, complex biosynthesis. Current Opinion in Plant Biology 10: 220-226.

Jung, H.-J.G. and W. Ni. 1998. Lignification of plant cell walls: impact of genetic manipulation. Proceedings of the National Academy of Sciences 95: 12742-12743.

Jung, J.H., W.M. Fouad, W. Vermerris, M. Gallo and F. Altpeter. 2012. RNAi suppression of lignin biosynthesis in sugarcane reduces recalcitrance for biofuel production from lignocellulosic biomass. Plant Biotechnology Journal 10: 1067-1076.

Kal, A.J., A.J. van Zonneveld, V. Benes, M. van den Berg, M.G. Koerkamp, K. Albermann, et al. 1999. Dynamics of gene expression revealed by comparison of serial analysis of gene expression transcript profiles from yeast grown on two different carbon sources. Molecular Biology of the Cell 10: 1859-1872.

Kanehisa, M. and S. Goto. 2000. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Research 28: 27-30.

Kang, Q., L. Appels, T. Tan and R. Dewil. 2014. Bioethanol from lignocellulosic biomass: current findings determine research priorities. The Scientific World Journal doi:org/10.1155/2014/298153.

Kazama, T., T. Nakamura, M. Watanabe, M. Sugita and K. Toriyama. 2008. Suppression mechanism of mitochondrial ORF79 accumulation by Rf1 protein in BT-type cytoplasmic male sterile rice. The Plant Journal 55: 619-628.

Kazama, T. and K. Toriyama. 2003. A pentatricopeptide repeat-containing gene that promotes the processing of aberrant *atp6* RNA of cytoplasmic male-sterile rice. FEBS Letters 544: 99-102.

Kim, S.S., E. Grienenberger, B. Lallemand, C.C. Colpitts, S.Y. Kim, C. de Azevedo Souza, et al. 2010. *LAP6/POLYKETIDE SYNTHASE A* and *LAP5/POLYKETIDE SYNTHASE B* encode hydroxyalkyl α-pyrone synthases required for pollen development and sporopollenin biosynthesis in *Arabidopsis thaliana*. The Plant Cell 22: 4045-4066.

Kitazaki, K., T. Arakawa, M. Matsunaga, R. Yui-Kurino, H. Matsuhira, T. Mikami, et al. 2015. Post-translational mechanisms are associated with fertility restoration of cytoplasmic male sterility in sugar beet (*Beta vulgaris*). The Plant Journal 83: 290-299.

Klein, R.R., P.E. Klein, J.E. Mullet, P. Minx, W.L. Rooney and K.F. Schertz. 2005. Fertility restorer locus *Rf1* of sorghum (*Sorghum bicolor* L.) encodes a pentatricopeptide repeat protein not present in the colinear region of rice chromosome 12. Theoretical and Applied Genetics 111: 994-1012.

Klein, R.R., F.R. Miller, D.V. Dugas, P.J. Brown, A.M. Burrell and P.E. Klein. 2015. Allelic variants in the *PRR37* gene and the human-mediated dispersal and diversification of sorghum. Theoretical and Applied Genetics 128: 1669-1683.

Klein, R.R., J.E. Mullet, D.R. Jordan, F.R. Miller, W.L. Rooney, M.A. Menz, et al. 2008. The effect of tropical sorghum conversion and inbred development on genome diversity as revealed by high-resolution genotyping. Crop Science 48: S12-S26.

Koizuka, N., R. Imai, H. Fujimoto, T. Hayakawa, Y. Kimura, J. Kohno-Murase, et al. 2003. Genetic characterization of a pentatricopeptide repeat protein gene, *orf687*, that restores fertility in the cytoplasmic male-sterile Kosena radish. The Plant Journal 34: 407-415.

Komori, T., S. Ohta, N. Murai, Y. Takakura, Y. Kuraya, S. Suzuki, et al. 2004. Map-based cloning of a fertility restorer gene, *Rf-1*, in rice (*Oryza sativa* L.). The Plant Journal 37: 315-325.

Kotak, S., J. Larkindale, U. Lee, P. von Koskull-Döring, E. Vierling and K.-D. Scharf. 2007. Complexity of the heat stress response in plants. Current Opinion in Plant Biology 10: 310-316.

Kuhlman, L.C., B.L. Burson, P.E. Klein, R.R. Klein, D.M. Stelly, H.J. Price, et al. 2008. Genetic recombination in *Sorghum bicolor* x *S. macrospermum* interspecific hybrids. Genome 51: 749-756.

Kumar, P., N. Vasupalli, R. Srinivasan and S.R. Bhat. 2012. An evolutionarily conserved mitochondrial *orf108* is associated with cytoplasmic male sterility in different alloplasmic lines of *Brassica juncea* and induces male sterility in transgenic *Arabidopsis thaliana*. Journal of Experimental Botany 63: 2921-2932.

Kuzmin, E.V., O.V. Karpova, T.E. Elthon and K.J. Newton. 2004. Mitochondrial respiratory deficiencies signal up-regulation of genes for heat shock proteins. Journal of Biological Chemistry 279: 20672-20677.

Langfelder, P. and S. Horvath. 2008. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 9: 559-571.

Lapierre, C., B. Pollet, M. Petit-Conil, G. Toval, J. Romero, G. Pilate, et al. 1999. Structural alterations of lignins in transgenic poplars with depressed cinnamyl alcohol dehydrogenase or caffeic acid O-methyltransferase activity have an opposite impact on the efficiency of industrial kraft pulping. Plant Physiology 119: 153-164.

Leyva, A., J.A. Jarillo, J. Salinas and J.M. Martinez-Zapater. 1995. Low temperature induces the accumulation of *phenylalanine ammonia-lyase* and *chalcone synthase* mRNAs of *Arabidopsis thaliana* in a light-dependent manner. Plant Physiology 108: 39-46.

Li, H. and D. Zhang. 2010. Biosynthesis of anther cuticle and pollen exine in rice. Plant Signaling & Behavior 5: 1121-1123.

Li, J., S. Han, X. Ding, T. He, J. Dai, S. Yang, et al. 2015a. Comparative transcriptome analysis between the cytoplasmic male sterile line NJCMS1A and its maintainer NJCMS1B in soybean (*Glycine max* (L.) Merr.). PloS One 10: 1-18.

Li, S., D. Yang and Y. Zhu. 2007. Characterization and use of male sterility in hybrid rice breeding. Journal of Integrative Plant Biology 49: 791-804.

Li, Z., Y. Cheng, J. Cui, P. Zhang, H. Zhao and S. Hu. 2015b. Comparative transcriptome analysis reveals carbohydrate and lipid metabolism blocks in *Brassica napus* L. male sterility induced by the chemical hybridization agent monosulfuron ester sodium. BMC Genomics 16: 206-224.

Limayem, A. and S.C. Ricke. 2012. Lignocellulosic biomass for bioethanol production: current perspectives, potential issues and future prospects. Progress in Energy and Combustion Science 38: 449-467.

Lister, R., B.D. Gregory and J.R. Ecker. 2009. Next is now: new technologies for sequencing of genomes, transcriptomes, and beyond. Current opinion in plant biology 12: 107-118.

Liu, C., N. Ma, P.-Y. Wang, N. Fu and H.-L. Shen. 2013. Transcriptome sequencing and *de novo* analysis of a cytoplasmic male sterile line and its near-isogenic restorer line in chili pepper (*Capsicum annuum* L.). PloS One doi:10.1371/journal.pone.0065209.

Liu, Q., Y. Lan, C. Wen, H. Zhao, J. Wang and Y. Wang. 2016a. Transcriptome sequencing analyses between the cytoplasmic male sterile line and its maintainer line in welsh onion (*Allium fistulosum* L.). International Journal of Molecular Sciences doi:10.3390/ijms17071058.

Liu, Z., Z. Yang, X. Wang, K. Li, H. An, J. Liu, et al. 2016b. A mitochondria-targeted PPR protein restores *pol* cytoplasmic male sterility by reducing *orf224* transcript levels in oilseed rape. Molecular Plant 9: 1082-1084.

Lombard, V., H.G. Ramulu, E. Drula, P.M. Coutinho and B. Henrissat. 2014. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Research 42: 490-495.

MacMillan, C.P., S.D. Mansfield, Z.H. Stachurski, R. Evans and S.G. Southerton. 2010. Fasciclin-like arabinogalactan proteins: specialization for stem biomechanics and cell wall architecture in *Arabidopsis* and *Eucalyptus*. Plant Journal 62: 689-703.

Malone, J.H. and B. Oliver. 2011. Microarrays, deep sequencing and the true measure of the transcriptome. BMC biology 9: 34-42.

Marguerat, S. and J. Bähler. 2010. RNA-seq: from technology to biology. Cellular and molecular life sciences 67: 569-579.

Matsunaga, M., H. Nagano, T. Mikami and T. Kubo. 2011. Large 3 ' UTR of sugar beet *rps3* is truncated in cytoplasmic male-sterile mitochondria. Plant Cell Reports 30: 231-238.

Maunder, A.B. and R.C. Pickett. 1963. The genetic inheritance of cytoplasmic-genetic male sterility in grain sorghum. Crop Science 3: 481-483.

McClung, C.R. and S.J. Davis. 2010. Ambient thermometers in plants: from physiological outputs towards mechanisms of thermal sensing. Current Biology 20: R1086-R1092.

McKinley, B., W. Rooney, C. Wilkerson and J. Mullet. 2016. Dynamics of biomass partitioning, stem gene expression, cell wall biosynthesis, and sucrose accumulation during development of *Sorghum bicolor*. The Plant Journal doi:10.1111/tpj.13269.

Mei, S., T. Liu and Z. Wang. 2016. Comparative transcriptome profile of the cytoplasmic male sterile and fertile floral buds of radish (*Raphanus sativus* L.). International Journal of Molecular Sciences doi:10.3390/ijms17010042.

Melan, M.A., X. Dong, M.E. Endara, K.R. Davis, F.M. Ausubel and T.K. Peterman. 1993. An *Arabidopsis thaliana* lipoxygenase gene can be induced by pathogens, abscisic acid, and methyl jasmonate. Plant Physiology 101: 441-450.

Miller, D.A. and R.C. Pickett. 1964. Inheritance of partial male-fertility in *Sorghum vulgare* Pers. Crop Science 4: 1-4.

Miller, F.R., D.K. Barnes and H.J. Cruzado. 1968. Effect of tropical photoperiods on the growth of sorghum when grown in 12 monthly plantings. Crop Science 8: 499-509.

Min, L., Y. Li, Q. Hu, L. Zhu, W. Gao, Y. Wu, et al. 2014. Sugar and auxin signaling pathways respond to high-temperature stress during anther development as revealed by transcript profiling analysis in cotton. Plant Physiology 164: 1293-1308.

Minic, Z., E. Jamet, H. San-Clemente, S. Pelletier, J.-P. Renou, C. Rihouey, et al. 2009. Transcriptomic analysis of *Arabidopsis* developing stems: a close-up on cell wall genes. BMC Plant Biology 9: 1-17.

Monk, R.L., F.R. Miller and G.G. McBee. 1984. Sorghum improvement for energy production. Biomass 6: 145-153.

Morant, M., K. Jørgensen, H. Schaller, F. Pinot, B.L. Møller, D. Werck-Reichhart, et al. 2007. CYP703 is an ancient cytochrome P450 in land plants catalyzing in-chain hydroxylation of lauric acid to provide building blocks for sporopollenin synthesis in pollen. The Plant Cell 19: 1473-1487.

Mullet, J., D. Morishige, R. McCormick, S. Truong, J. Hilley, B. McKinley, et al. 2014. Energy Sorghum—a genetic model for the design of C4 grass bioenergy crops. Journal of Experimental Botany 65: 3479-3489.

Naoumkina, M.A., Q. Zhao, L. Gallego-Giraldo, X. Dai, P.X. Zhao and R.A. Dixon. 2010. Genome-wide analysis of phenylpropanoid defence pathways. Molecular Plant Pathology 11: 829-846.

Olson, S.N., K. Ritter, W. Rooney, A. Kemanian, B.A. McCarl, Y. Zhang, et al. 2012. High biomass yield energy sorghum: developing a genetic model for C4 grass bioenergy crops. Biofuels, Bioproducts and Biorefining 6: 640-655.

Overvoorde, P.J., Y. Okushima, J.M. Alonso, A. Chan, C. Chang, J.R. Ecker, et al. 2005. Functional genomic analysis of the *AUXIN/INDOLE-3-ACETIC ACID* gene family members in *Arabidopsis thaliana*. The Plant Cell 17: 3282-3300.

Park, J.Y., Y.-P. Lee, J. Lee, B.-S. Choi, S. Kim and T.-J. Yang. 2013. Complete mitochondrial genome sequence and identification of a candidate gene responsible for cytoplasmic male sterility in radish (*Raphanus sativus* L.) containing DCGMS cytoplasm. Theoretical and Applied Genetics 126: 1763-1774.

Paterson, A.H., J.E. Bowers, R. Bruggmann, I. Dubchak, J. Grimwood, H. Gundlach, et al. 2009. The *Sorghum bicolor* genome and the diversification of grasses. Nature 457: 551-556.

Pedersen, J., H.J. Gorz, F.A. Haskins and W. Ross. 1982. Variability for quality and agronomic traits in forage sorghum hybrids. Crop Science 22: 853-856.

Persson, S., H. Wei, J. Milne, G.P. Page and C.R. Somerville. 2005. Identification of genes required for cellulose synthesis by regression analysis of public microarray data sets. Proceedings of the National Academy of Sciences 102: 8633-8638.

Piffanelli, P., J.H. Ross and D. Murphy. 1998. Biogenesis and function of the lipidic structures of pollen grains. Sexual Plant Reproduction 11: 65-80.

Poovaiah, C.R., M. Nageswara-Rao, J.R. Soneji, H.L. Baxter and C.N. Stewart. 2014. Altered lignin biosynthesis using biotechnology to improve lignocellulosic biofuel feedstocks. Plant Biotechnology Journal 12: 1163–1173.

Prakasham, R.S., D. Nagaiah, K.S. Vinutha, A. Uma, T. Chiranjeevi, A.V. Umakanth, et al. 2014. Sorghum biomass: a novel renewable carbon source for industrial bioproducts. Biofuels 5: 159-174.

Quinby, J. and R. Karper. 1945. Inheritance of three genes that influence time of floral initiation and maturity date in milo. Journal of the American Society of Agronomy 37: 916-936.

Ralph, J., C. Lapierre, J.M. Marita, H. Kim, F. Lu, R.D. Hatfield, et al. 2001. Elucidation of new structures in lignins of *CAD*- and *COMT*-deficient plants by NMR. Phytochemistry 57: 993-1003.

Reddy, B.V., S. Ramesh and P.S. Reddy. 2006. Sorghum genetic resources, cytogenetics and improvement. Genetic Resources, Chromosome Engineering, and Crop Improvement 2: 309-363.

Reddy, B.V.S., S. Ramesh and R. Ortiz. 2010. Genetic and cytoplasmic-nuclear male sterility in sorghum. Plant Breeding Reviews 25: 139-172.

Rhoads, D.M., C.S. Levings and J.N. Siedow. 1995. *URF13,* a ligand-gated, pore-forming receptor for T-toxin in the inner membrane of *cms-T* mitochondria. Journal of Bioenergetics & Biomembranes 27: 437-445.

Rhoads, D.M. and C.C. Subbaiah. 2007. Mitochondrial retrograde regulation in plants. Mitochondrion 7: 177-194.

Richmond, T.A. and C.R. Somerville. 2000. The cellulose synthase superfamily. Plant Physiology 124: 495-498.

Robinson, M.D., D.J. McCarthy and G.K. Smyth. 2010. edgeR: a bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26: 139-140.

Robinson, M.D. and G.K. Smyth. 2008. Small-sample estimation of negative binomial dispersion, with applications to SAGE data. Biostatistics 9: 321-332.

Rooney, W. 2004. Sorghum improvement-integrating traditional and new technology to produce improved genotypes. Advances in Agronomy 83: 37-109.

Rooney, W.L. 2014. Sorghum, Cellulosic Energy Cropping Systems. John Wiley & Sons, Ltd. p. 109-129.

Rooney, W.L. and S. Aydin. 1999. Genetic control of a photoperiod-sensitive response in *Sorghum bicolor* (L.) Moench. Crop Science 39: 397-400.

Rooney, W.L., J. Blumenthal, B. Bean and J.E. Mullet. 2007. Designing sorghum as a dedicated bioenergy feedstock. Biofuels, Bioproducts and Biorefining 1: 147-157.

Saballos, A. 2008. Development and utilization of sorghum as a bioenergy crop, Genetic Improvement of Bioenergy Crops. Springer. p. 211-248.

Saballos, A., S.E. Sattler, E. Sanchez, T.P. Foster, Z. Xin, C. Kang, et al. 2012. *Brown midrib2* (*Bmr2*) encodes the major 4-coumarate: coenzyme A ligase involved in lignin biosynthesis in sorghum (*Sorghum bicolor* (L.) Moench). The Plant Journal 70: 818-830.

Saritha, M. and A. Arora. 2012. Biological pretreatment of lignocellulosic substrates for enhanced delignification and enzymatic digestibility. Indian Journal of Microbiology 52: 122-130.

Scheible, W.-R. and M. Pauly. 2004. Glycosyltransferases and cell wall biosynthesis: novel players and insights. Current Opinion in Plant Biology 7: 285-295.

Scheller, H.V. and P. Ulvskov. 2010. Hemicelluloses. Plant Biology 61: 263-289.

Schmitz-Linneweber, C. and I. Small. 2008. Pentatricopeptide repeat proteins: a socket set for organelle gene expression. Trends in Plant Science 13: 663-670.

Schnable, P.S. and R.P. Wise. 1998. The molecular basis of cytoplasmic male sterility and fertility restoration. Trends in Plant Science 3: 175-180.

Schwarzlaender, M., A.-C. Koenig, L.J. Sweetlove and I. Finkemeier. 2012. The impact of impaired mitochondrial function on retrograde signalling: a meta-analysis of transcriptomic responses. Journal of Experimental Botany 63: 1735-1750.

Shannon, P., A. Markiel, O. Ozier, N.S. Baliga, J.T. Wang, D. Ramage, et al. 2003. Cytoscape: A software environment for integrated models of biomolecular interaction networks. Genome Research 13: 2498-2504.

Shen, B., N. Carneiro, I. Torres-Jerez, B. Stevenson, T. McCreery, T. Helentjaris, et al. 1994. Partial sequencing and mapping of clones from two maize cDNA libraries. Plant Molecular Biology 26: 1085-1101.

Smith, C.W. and R.A. Frederiksen. 2000. Sorghum: Origin, History, Technology, and Production. John Wiley & Sons, Inc. New York. p. 825.

Stefaniak, T. and W. Rooney. 2013. Breeding Sorghum as a Bioenergy Crop, In: H. Saha, H. Bhandari and J. Bouton, editors. Bioenergy Feedstocks: Breeding and Genetics. John Wiley & Sons, Inc. Oxford, UK. p. 83-116.

Stefaniak, T.R., J.A. Dahlberg, B.W. Bean, N. Dighe, E.J. Wolfrum and W.L. Rooney. 2012. Variation in biomass composition components among forage, biomass, sorghum-sudangrass, and sweet sorghum types. Crop Science 52: 1949-1954.

Stephens, J., F. Miller and D. Rosenow. 1967. Conversion of alien sorghums to early combine genotypes. Crop Science 7: 396-396.

Stephens, J.C. and R.F. Holland. 1954. Cytoplasmic male sterility for hybrid sorghum seed production. Agronomy Journal 46: 20-23.

Su, A.-g., W. Song, Z. Shi, Y.-x. Zhao, J.-f. Xing, R.-y. Zhang, et al. 2017. Exploring differentially expressed genes associated with fertility instability of S-type cytoplasmic male-sterility in maize by RNA-Seq. Journal of Integrative Agriculture doi:10.1016/S2095-3119(16)61494-6.

Suh, M.C., A.L. Samuels, R. Jetter, L. Kunst, M. Pollard, J. Ohlrogge, et al. 2005. Cuticular lipid composition, surface structure, and gene expression in *Arabidopsis* stem epidermis. Plant Physiology 139: 1649-1665.

Supek, F., M. Bošnjak, N. Škunca and T. Šmuc. 2011. REVIGO summarizes and visualizes long lists of gene ontology terms. PloS One doi:10.1371/journal.pone.0021800.

Suzuki, N., S. Koussevitzky, R. Mittler and G. Miller. 2012. ROS and redox signalling in the response of plants to abiotic stress. Plant, Cell & Environment 35: 259-270.

Tanaka, K., K. Murata, M. Yamazaki, K. Onosato, A. Miyao and H. Hirochika. 2003. Three distinct rice cellulose synthase catalytic subunit genes required for cellulose synthesis in the secondary wall. Plant Physiology 133: 73-83.

Tang, H., D. Luo, D. Zhou, Q. Zhang, D. Tian, X. Zheng, et al. 2014. The rice restorer *Rf4* for wild-abortive cytoplasmic male sterility encodes a mitochondrial-localized PPR protein that functions in reduction of *WA352* transcripts. Molecular Plant 7: 1497-1500.

Tang, H.V., W. Chen and D.R. Pring. 1999. Mitochondrial *orf107* transcription, editing, and nucleolytic cleavage conferred by the gene *Rf3* are expressed in sorghum pollen. Sexual Plant Reproduction 12: 53-59.

Tang, H.V., D.R. Pring, L.C. Shaw, R.A. Salazar, F.R. Muza, B. Yan, et al. 1996. Transcript processing internal to a mitochondrial open reading frame is correlated with fertility restoration in male-sterile sorghum. Plant Journal 10: 123-133.

Tang, L.K., H. Chu, W.K. Yip, E.C. Yeung and C. Lo. 2009. An anther-specific *dihydroflavonol 4-reductase-like* gene (*DRL1*) is essential for male fertility in *Arabidopsis*. New Phytologist 181: 576-587.

Uyttewaal, M., N. Arnal, M. Quadrado, A. Martin-Canadell, N. Vrielynck, S. Hiard, et al. 2008. Characterization of *Raphanus sativus* pentatricopeptide repeat proteins encoded by the fertility restorer locus for Ogura cytoplasmic male sterility. Plant Cell 20: 3331-3345.

Valdivia, E.R., M.T. Herrera, C. Gianzo, J. Fidalgo, G. Revilla, I. Zarra, et al. 2013. Regulation of secondary wall synthesis and cell death by *NAC* transcription factors in the monocot *Brachypodium distachyon*. Journal of Experimental Botany 64: 1333-1343.

Van Der Meer, I.M., M.E. Stam, A.J. van Tunen, J. Mol and A.R. Stuitje. 1992. Antisense inhibition of flavonoid biosynthesis in petunia anthers results in male sterility. The Plant Cell 4: 253-262.

Vanderlip, R.L. and H.E. Reeves. 1972. Growth Stages of Sorghum [*Sorghum bicolor*, (L) Moench]. Agronomy Journal 64: 13-16.

Venuto, B. and B. Kindiger. 2008. Forage and biomass feedstock production from hybrid forage sorghum and sorghum–sudangrass hybrids. Grassland Science 54: 189-196.

Vogel, J. 2008. Unique aspects of the grass cell wall. Current Opinion in Plant Biology 11: 301-307.

Vogel, J.P., D.F. Garvin, T.C. Mockler, J. Schmutz, D. Rokhsar, M.W. Bevan, et al. 2010. Genome sequencing and analysis of the model grass *Brachypodium distachyon*. Nature 463: 763-768.

Vogt, T. 2010. Phenylpropanoid biosynthesis. Molecular Plant 3: 2-20.

Wang, L., K. Guo, Y. Li, Y. Tu, H. Hu, B. Wang, et al. 2010. Expression profiling and integrative analysis of the *CESA*/*CSL* superfamily in rice. BMC Plant Biology 10: 282-297.

Wang, S., C. Wang, X.-X. Zhang, X. Chen, J.-J. Liu, X.-F. Jia, et al. 2016. Transcriptome *de novo* assembly and analysis of differentially expressed genes related to cytoplasmic male sterility in cabbage. Plant Physiology and Biochemistry 105: 224-232.

Wang, X.-D. and Y.-Y. Li. 2002. Development of transgenic restorer of cytoplasmic male sterility in upland cotton. Agricultural Sciences in China 1: 375-380.

Wang, Z., M. Gerstein and M. Snyder. 2009. RNA-Seq: a revolutionary tool for transcriptomics. Nature Reviews Genetics 10: 57-63.

Wang, Z.H., Y.J. Zou, X.Y. Li, Q.Y. Zhang, L. Chen, H. Wu, et al. 2006. Cytoplasmic male sterility of rice with Boro II cytoplasm is caused by a cytotoxic peptide and is restored by two related PPR motif genes via distinct modes of mRNA silencing. Plant Cell 18: 676-687.

Wong, M.M., C.H. Cannon and R. Wickneswari. 2011. Identification of lignin genes and regulatory sequences involved in secondary cell wall formation in *Acacia auriculiformis* and *Acacia mangium* via *de novo* transcriptome sequencing. BMC Genomics 12: 342-354.

Woodson, J.D. and J. Chory. 2008. Coordination of gene expression between organellar and nuclear genomes. Nature Reviews Genetics 9: 383-395.

Wu, A.M., C. Rihouey, M. Seveno, E. Hörnblad, S.K. Singh, T. Matsunaga, et al. 2009. The *Arabidopsis IRX10* and *IRX10-LIKE* glycosyltransferases are critical for glucuronoxylan biosynthesis during secondary cell wall formation. The Plant Journal 57: 718-731.

Xie, C., X. Mao, J. Huang, Y. Ding, J. Wu, S. Dong, et al. 2011. KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. Nucleic Acids Research 39: W316-W322.

Xie, Y., W. Zhang, Y. Wang, L. Xu, X. Zhu, E.M. Muleke, et al. 2016. Comprehensive transcriptome-based characterization of differentially expressed genes involved in microsporogenesis of radish CMS line and its maintainer. Functional and Integrative Genomics 16: 529-543.

Yang, J., X. Liu, X. Yang and M. Zhang. 2010. Mitochondrially-targeted expression of a cytoplasmic male sterility-associated *orf220* gene causes male sterility in *Brassica juncea*. BMC Plant Biology 10: 231-241.

Yang, P., J. Han and J. Huang. 2014a. Transcriptome sequencing and *de novo* analysis of cytoplasmic male sterility and maintenance in JA-CMS cotton. PloS One doi:10.1371/journal.pone.0112320.

Yang, S., B.D. Weers, D.T. Morishige and J.E. Mullet. 2014b. CONSTANS is a photoperiod regulated activator of flowering in sorghum. BMC Plant Biology 14: 148-162.

Yang, S.S., Z.J. Tu, F. Cheung, W.W. Xu, J.F. Lamb, H.-J.G. Jung, et al. 2011. Using RNA-Seq for gene identification, polymorphism detection and transcript profiling in two alfalfa genotypes with divergent cell wall composition in stems. BMC Genomics 12: 199-217.

Yang, Y., C. Dong, J. Yu, L. Shi, C. Tong, Z. Li, et al. 2014c. *Cysteine Protease 51 (CP51*), an anther-specific cysteine protease gene, is essential for pollen exine formation in *Arabidopsis*. Plant Cell, Tissue and Organ Culture 119: 383-397.

Yokoyama, R. and K. Nishitani. 2006. Identification and characterization of *Arabidopsis thaliana* genes involved in xylem secondary cell walls. Journal of Plant Research 119: 189-194.

Yoon, J., H. Choi and G. An. 2015. Roles of lignin biosynthesis and regulatory genes in plant development. Journal of Integrative Plant Biology 57: 902-912.

Yu, H., N.M. Luscombe, J. Qian and M. Gerstein. 2003. Genomic analysis of gene expression relationships in transcriptional regulatory networks. Trends in Genetics 19: 422-427.

Yu, Y., R. Hu, H. Wang, Y. Cao, G. He, C. Fu, et al. 2013. *MlWRKY12*, a novel Miscanthus transcription factor, participates in pith secondary cell wall formation and promotes flowering. Plant Science 212: 1-9.

Zang, A., X. Xu, S. Neill and W. Cai. 2009. Overexpression of *OsRAN2* in rice and *Arabidopsis* renders transgenic plants hypersensitive to salinity and osmotic stress. Journal of experimental botany doi:org/10.1093/jxb/erp341.

Zhang, B. and S. Horvath. 2005. A general framework for weighted gene co-expression network analysis. Statistical Applications in Genetics and Molecular Biology 4: 1-45.

Zhang, J., S. Zhang, H. Li, H. Du, H. Huang, Y. Li, et al. 2016. Identification of transcription factors *ZmMYB111* and *ZmMYB148* involved in phenylpropanoid metabolism. Frontiers in Plant Science doi:10.3389/fpls.2016.00148.

Zhang, Z.B., J. Zhu, J.F. Gao, C. Wang, H. Li, H. Li, et al. 2007. Transcription factor *AtMYB103* is required for anther development by regulating tapetum development, callose dissolution and exine formation in *Arabidopsis*. The Plant Journal 52: 528-538.

Zhu, Q., Y. Song, G. Zhang, L. Ju, J. Zhang, Y. Yu, et al. 2015. *De novo* assembly and transcriptome analysis of wheat with male sterility induced by the chemical hybridizing agent SQ-1. PloS one 10. doi: doi:org/10.1371/journal.pone.0123556.

APPENDIX A

TABLE A-1: SELECTED DEGS IN DIFFERENT LINES OF SORGHUM. THE FDR CORRECTED *P*-VALUE <0.05 AND THE ABSOLUTE VALUE OF LOG_2RATIO ≥1 WERE USED AS THE THRESHOLDS TO SELECT SIGNIFICANT DEGS.

Feature ID	**Comparison**	**Annotation**	**GO terms**	**Fold change[a]**	**FDR *p*-value**	**Gene ortholog**	
						Arabidopsis	**Maize**
Sobic.001G012500	**A992422 vs. A992422×QL12**	similar to Glutathione S-transferase 1	protein binding	-2.71	0.00	*AT3G62760*	*GRMZM2G116273*
Sobic.001G012500	**AQL33 vs. AQL33×QL36**	similar to Glutathione S-transferase 1	protein binding	-2.34	0.00	*AT3G62760*	*GRMZM2G116273*
Sobic.001G012500	**ATx623 vs. ATx623×RTx2783**	similar to Glutathione S-transferase 1	protein binding	2.02	0.00	*AT3G62760*	*GRMZM2G116273*
Sobic.001G012500	**AQL33 vs. AQL33×R931945-2-2**	similar to Glutathione S-transferase 1	protein binding	2.13	0.00	*AT3G62760*	*GRMZM2G116273*
Sobic.001G120400	**A992422 vs. A992422×QL36**	similar to Phospholipase A2, putative, expressed	lipid catabolic process, calcium ion binding, phospholipase A2 activity	-2.14	0.01	*AT2G06925*	*GRMZM2G033820*
Sobic.001G120400	**A992422 vs. A992422×QL12**	similar to Phospholipase A2, putative, expressed	N/A	-2.02	0.01	*AT2G06925*	*GRMZM2G033820*
Sobic.001G215000	**AQL33 vs. AQL33×R931945-2-2**	N/A	N/A	-17.71	00.00	*AT1G02050*	*GRMZM2G380650*

Table A-1: Continued

Feature ID	Comparison	Annotation	GO terms	Fold change[a]	FDR *p*-value	Gene ortholog	
						Arabidopsis	**Maize**
Sobic.001G215000	**A992422 vs. A992422×QL36**	N/A	N/A	3.57	0.00	*AT1G02050*	*GRMZM2G380650*
Sobic.001G215000	**A992422 vs. A992422×R931945-2-2**	N/A	N/A	4.70	0.00	*AT1G02050*	*GRMZM2G380650*
Sobic.001G215000	**A992422 vs. A992422×QL12**	N/A	N/A	17.03	0.00	*AT1G02050*	*GRMZM2G380650*
Sobic.001G318700	**AQL33 vs. AQL33×QL36**	similar to Putative uncharacterized protein	protein binding	7.15	0.00	*AT1G10370*	*GRMZM2G016241*
Sobic.001G420100	**AQL33 vs. AQL33×R931945-2-2**	similar to Heat shock cognate 70 kDa protein	N/A	-10.35	0.00	*AT3G12580*	*AC194017.3_FG001*
Sobic.001G426000	**AQL33 vs. AQL33×QL36**	similar to 17.4 kDa class I heat shock protein 3	N/A	-12.38	0.00	*AT1G53540*	*AC208204.3_FG006*
Sobic.001G426000	**AQL33 vs. AQL33×R931945-2-2**	similar to 17.4 kDa class I heat shock protein 3	N/A	-9.06	0.00	*AT1G53540*	*AC208204.3_FG006*
Sobic.001G426000	**AQL33 vs. AQL33×QL12**	similar to 17.4 kDa class I heat shock protein 3	N/A	-2.47	0.00	*AT1G53540*	*AC208204.3_FG006*
Sobic.001G426000	**ATx623 vs. ATx623×RTx436**	similar to 17.4 kDa class I heat shock protein 3	N/A	2.21	0.00	*AT1G53540*	*AC208204.3_FG006*
Sobic.001G426000	**ATx623 vs. ATx623×RTx2783**	similar to 17.4 kDa class I heat shock protein 3	N/A	2.48	0.00	*AT1G53540*	*AC208204.3_FG006*
Sobic.001G426000	**A992422 vs. A992422×QL36**	similar to 17.4 kDa class I heat shock protein 3	N/A	4.34	0.00	*AT1G53540*	*AC208204.3_FG006*

Table A-1: Continued

Feature ID	Comparison	Annotation	GO terms	Fold change[a]	FDR *p*-value	Gene ortholog	
						Arabidopsis	**Maize**
Sobic.001G428300	**AQL33 vs. AQL33×R931945-2-2**	similar to Male fertility protein	N/A	-11.94	0.00	*AT3G59530*	*GRMZM2G430601*
Sobic.001G428300	**AQL33 vs. AQL33×QL36**	similar to Male fertility protein	N/A	2.33	0.00	*AT3G59530*	*GRMZM2G430601*
Sobic.001G428300	**A992422 vs. A992422×QL36**	similar to Male fertility protein	N/A	10.97	0.02	*AT3G59530*	*GRMZM2G430601*
Sobic.001G428300	**A992422 vs. A992422×R931945-2-2**	similar to Male fertility protein	N/A	19.33	0.00	*AT3G59530*	*GRMZM2G430601*
Sobic.001G428300	**A992422 vs. A992422×QL12**	similar to Male fertility protein	N/A	48.57	0.00	*AT3G59530*	*GRMZM2G430601*
Sobic.001G491400	**AQL33 vs. AQL33×R931945-2-2**	similar to Cytochrome P450-like protein	oxidation reduction, oxidoreductase activity	-21.23	0.00	*AT1G69500*	*AC211006.3_FG004*
Sobic.001G491400	**ATx623 vs. ATx623×RTx436**	similar to Cytochrome P450-like protein	oxidation reduction, oxidoreductase activity	-2.18	0.00	*AT1G69500*	*AC211006.3_FG004*
Sobic.001G491400	**AQL33 vs. AQL33×QL36**	similar to Cytochrome P450-like protein	oxidation reduction, oxidoreductase activity	2.15	0.00	*AT1G69500*	*AC211006.3_FG004*
Sobic.001G491400	**A992422 vs. A992422×QL36**	similar to Cytochrome P450-like protein	oxidation reduction, oxidoreductase activity	3.35	0.03	*AT1G69500*	*AC211006.3_FG004*

Table A-1: Continued

Feature ID	Comparison	Annotation	GO terms	Fold change[a]	FDR *p*-value	Gene ortholog	
						Arabidopsis	**Maize**
Sobic.001G491400	**A992422 vs. A992422×R931945-2-2**	similar to Cytochrome P450-like protein	oxidation reduction, oxidoreductase activity	4.53	0.00	*AT1G69500*	*AC211006.3_FG004*
Sobic.001G491400	**A992422 vs. A992422×QL12**	similar to Cytochrome P450-like protein	oxidation reduction, oxidoreductase activity	19.81	0.00	*AT1G69500*	*AC211006.3_FG004*
Sobic.001G492400	**AQL33 vs. AQL33×R931945-2-2**	similar to Putative male sterility protein	fatty-acyl-CoA reductase	-87.86	0.00	*AT3G11980*	*GRMZM2G120987*
Sobic.001G492400	**ATx623 vs. ATx623×RTx436**	similar to Putative male sterility protein	fatty-acyl-CoA reductase	-2.62	0.00	*AT3G11980*	*GRMZM2G120987*
Sobic.001G492400	**AQL33 vs. AQL33×QL36**	similar to Putative male sterility protein	fatty-acyl-CoA reductase	3.88	0.00	*AT3G11980*	*GRMZM2G120987*
Sobic.001G492400	**A992422 vs. A992422×QL12**	similar to Putative male sterility protein	fatty-acyl-CoA reductase	106.46	0.00	*AT3G11980*	*GRMZM2G120987*
Sobic.001G514200	**AQL33 vs. AQL33×QL12**	similar to Glutathione S-transferase GST 9	protein binding	2.40	0.00	*AT2G47730*	*GRMZM2G096247*
Sobic.001G514200	**AQL33 vs. AQL33×R931945-2-2**	similar to Glutathione S-transferase GST 9	protein binding	2.88	0.00	*AT2G47730*	*GRMZM2G096247*
Sobic.001G514300	**A992422 vs. A992422×QL36**	similar to Glutathione S-transferase GST 13	protein binding	-3.38	0.00	*AT2G47730*	*GRMZM2G096247*
Sobic.001G514300	**A992422 vs. A992422×QL12**	similar to Glutathione S-transferase GST 13	protein binding	-2.28	0.00	*AT2G47730*	*GRMZM2G096247*

Table A-1: Continued

Feature ID	Comparison	Annotation	GO terms	Fold change[a]	FDR *p*-value	Gene ortholog	
						Arabidopsis	**Maize**
Sobic.001G514300	**AQL33 vs. AQL33×R931945-2-2**	similar to Glutathione S-transferase GST 13	protein binding	2.84	0.04	*AT2G47730*	*GRMZM2G096247*
Sobic.001G514400	**AQL33 vs. AQL33×R931945-2-2**	similar to Glutathione S-transferase GSTF14	protein binding	3.18	0.00	*AT3G03190*	*GRMZM2G096153*
Sobic.002G115700	**AQL33 vs. AQL33×R931945-2-2**	similar to Putative chalcone synthase	N/A	-12.35	0.00	*AT4G34850*	*GRMZM2G108894*
Sobic.002G115700	**AQL33 vs. AQL33×QL36**	similar to Putative chalcone synthase	N/A	2.17	0.00	*AT4G34850*	*GRMZM2G108894*
Sobic.002G115700	**A992422 vs. A992422×QL36**	similar to Putative chalcone synthase	N/A	4.36	0.00	*AT4G34850*	*GRMZM2G108894*
Sobic.002G115700	**A992422 vs. A992422×R931945-2-2**	similar to Putative chalcone synthase	N/A	7.59	0.00	*AT4G34850*	*GRMZM2G108894*
Sobic.002G115700	**A992422 vs. A992422×QL12**	similar to Putative chalcone synthase	N/A	20.82	0.00	*AT4G34850*	*GRMZM2G108894*
Sobic.002G243500	**AQL33 vs. AQL33×QL36**	similar to Heat shock protein 81-2	unfolded protein binding, response to stress, protein folding, ATP binding	-2.94	0.00	*AT5G56000*	*GRMZM2G012631*

Table A-1: Continued

Feature ID	Comparison	Annotation	GO terms	Fold change[a]	FDR *p*-value	Gene ortholog	
						Arabidopsis	**Maize**
Sobic.002G251200	**AQL33 vs. AQL33×R931945-2-2**	similar to Dihydro-flavanoid reductase-like protein	coenzyme binding, catalytic activity	-19.18	0.00	*AT4G35420*	*GRMZM2G004683*
Sobic.002G251200	**AQL33 vs. AQL33×QL36**	similar to Dihydro-flavanoid reductase-like protein	coenzyme binding, catalytic activity	2.00	0.00	*AT4G35420*	*GRMZM2G004683*
Sobic.002G251200	**A992422 vs. A992422×QL36**	similar to Dihydro-flavanoid reductase-like protein	coenzyme binding, catalytic activity	12.73	0.00	*AT4G35420*	*GRMZM2G004683*
Sobic.002G251200	**A992422 vs. A992422×R931945-2-2**	similar to Dihydro-flavanoid reductase-like protein	coenzyme binding, catalytic activity	18.39	0.00	*AT4G35420*	*GRMZM2G004683*
Sobic.002G251200	**A992422 vs. A992422×QL12**	similar to Dihydro-flavanoid reductase-like protein	coenzyme binding, catalytic activity	71.74	0.00	*AT4G35420*	*GRMZM2G004683*
Sobic.002G384400	**AQL33 vs. AQL33×QL36**	similar to Thiazole biosynthetic enzyme 1-1, chloroplast precursor	N/A	18.51	0.00	*AT5G54770*	*GRMZM2G018375*
Sobic.003G004900	**AQL33 vs. AQL33×QL36**	similar to H0211A12.17 protein	metabolic process,catalytic activity	3.55	0.00	*AT1G62940*	*GRMZM2G014651*
Sobic.003G004900	**A992422 vs. A992422×QL36**	similar to H0211A12.17 protein	metabolic process,catalytic activity	4.31	0.01	*AT1G62940*	*GRMZM2G014651*
Sobic.003G004900	**A992422 vs. A992422×R931945-2-2**	similar to H0211A12.17 protein	metabolic process,catalytic activity	7.81	0.00	*AT1G62940*	*GRMZM2G014651*
Sobic.003G004900	**A992422 vs. A992422×QL12**	similar to H0211A12.17 protein	metabolic process,catalytic activity	30.44	0.00	*AT1G62940*	*GRMZM2G014651*
Sobic.003G039400	**AQL33 vs. AQL33×QL36**	similar to 17.8 kDa class II heat shock protein	N/A	-13.23	0.00	*AT5G12020*	*GRMZM2G012455*

Table A-1: Continued

Feature ID	Comparison	Annotation	GO terms	Fold change[a]	FDR *p*-value	Gene ortholog	
						Arabidopsis	**Maize**
Sobic.003G039400	**AQL33 vs. AQL33×R931945-2-2**	similar to 17.8 kDa class II heat shock protein	N/A	-3.73	0.00	*AT5G12020*	*GRMZM2G012455*
Sobic.003G039400	**A992422 vs. A992422×QL12**	similar to 17.8 kDa class II heat shock protein	N/A	-2.88	0.00	*AT5G12020*	*GRMZM2G012455*
Sobic.003G039400	**AQL33 vs. AQL33×QL12**	similar to 17.8 kDa class II heat shock protein	N/A	-2.49	0.00	*AT5G12020*	*GRMZM2G012455*
Sobic.003G071600	**A992422 vs. A992422×QL12**	N/A	oxidoreductase activity, lipid metabolic process, cytoplasm	2.40	0.00	*AT3G55360*	*GRMZM2G467242*
Sobic.003G071600	**AQL33 vs. AQL33×QL36**	N/A	oxidoreductase activity, lipid metabolic process, cytoplasm	2.49	0.00	*AT3G55360*	*GRMZM2G467242*
Sobic.003G081900	**AQL33 vs. AQL33×QL36**	similar to Heat shock protein 17.2	N/A	-15.58	0.00	*AT1G53540*	*AC208204.3_FG006*
Sobic.003G081900	**AQL33 vs. AQL33×R931945-2-2**	similar to Heat shock protein 17.2	N/A	-9.03	0.00	*AT1G53540*	*AC208204.3_FG006*
Sobic.003G081900	**AQL33 vs. AQL33×QL12**	similar to Heat shock protein 17.2	N/A	-2.30	0.00	*AT1G53540*	*AC208204.3_FG006*
Sobic.003G081900	**A992422 vs. A992422×R931945-2-2**	similar to Heat shock protein 17.2	N/A	9.18	0.00	*AT1G53540*	*AC208204.3_FG006*
Sobic.003G082300	**AQL33 vs. AQL33×R931945-2-2**	similar to Putative uncharacterized protein	N/A	-3.99	0.00	*AT1G53540*	*AC208204.3_FG006*
Sobic.003G082300	**AQL33 vs. AQL33×QL12**	similar to Putative uncharacterized protein	N/A	-2.9	0.00	*AT1G53540*	*AC208204.3_FG006*

Table A-1: Continued

Feature ID	Comparison	Annotation	GO terms	Fold change[a]	FDR *p*-value	Gene ortholog	
						Arabidopsis	**Maize**
Sobic.003G350700	**AQL33 vs. AQL33×QL36**	similar to Heat shock 70 kDa protein	N/A	-5.3	0.00	*AT3G12580*	*AC194017.3_FG001*
Sobic.003G350700	**AQL33 vs. AQL33×R931945-2-2**	similar to Heat shock 70 kDa protein	N/A	-4.4	0.00	*AT3G12580*	*AC194017.3_FG001*
Sobic.003G350700	**AQL33 vs. AQL33×QL12**	similar to Heat shock 70 kDa protein	N/A	-2.04	0.00	*AT3G12580*	*AC194017.3_FG001*
Sobic.003G350700	**ATx623 vs. ATx623×RTx436**	similar to Heat shock 70 kDa protein	N/A	2.05	0.03	*AT3G12580*	*AC194017.3_FG001*
Sobic.003G350700	**A992422 vs. A992422×QL36**	similar to Heat shock 70 kDa protein	N/A	2.33	0.00	*AT3G12580*	*AC194017.3_FG001*
Sobic.003G385900	**AQL33 vs. AQL33×QL36**	similar to Lipoxygenase	protein binding, oxidation reduction, metal ion binding	-2.58	0.00	*AT1G55020*	*GRMZM2G017068*
Sobic.003G385900	**A992422 vs. A992422×R931945-2-2**	similar to Lipoxygenase	protein binding, oxidation reduction, metal ion binding	-2.43	0.00	*AT1G55020*	*GRMZM2G017068*
Sobic.003G385900	**ATx623 vs. ATx623×RTx2783**	similar to Lipoxygenase	protein binding, oxidation reduction, metal ion binding	-2.34	0.00	*AT1G55020*	*GRMZM2G017068*
Sobic.004G001700	**AQL33 vs. AQL33×R931945-2-2**	N/A	N/A	-16.65	0.00	*AT2G16630*	*GRMZM2G130813*
Sobic.004G001700	**ATx623 vs. ATx623×RTx2783**	N/A	N/A	2.12	0.00	*AT2G16630*	*GRMZM2G130813*
Sobic.004G001700	**AQL33 vs. AQL33×QL36**	N/A	N/A	5.50	0.00	*AT2G16630*	*GRMZM2G130813*
Sobic.004G001700	**A992422 vs. A992422×QL12**	N/A	N/A	70.53	0.00	*AT2G16630*	*GRMZM2G130813*
Sobic.004G203900	**AQL33 vs. AQL33×QL36**	similar to B-keto acyl reductase	N/A	2.78	0.00	*AT1G67730*	*AC205703.4_FG006*
Sobic.004G220400	**AQL33 vs. AQL33×R931945-2-2**	similar to Phenylalanine ammonia-lyase	N/A	-2.18	0.00	*AT2G37040*	*GRMZM2G029048*

Table A-1: Continued

Feature ID	Comparison	Annotation	GO terms	Fold change[a]	FDR *p*-value	Gene ortholog	
						Arabidopsis	**Maize**
Sobic.004G228900	**AQL33 vs. AQL33×R931945-2-2**	similar to Low molecular weight heat shock protein precursor	N/A	-17.98	0.03	*AT5G51440*	*GRMZM2G007729*
Sobic.004G228900	**A992422 vs. A992422×QL36**	similar to Low molecular weight heat shock protein precursor	N/A	27.32	0.00	*AT5G51440*	*GRMZM2G007729*
Sobic.006G005600	**AQL33 vs. AQL33×R931945-2-2**	similar to Heat shock protein 82	unfolded protein binding, response to stress, protein folding, ATP binding	-12.59	0.00	*AT5G52640*	*GRMZM5G833699*
Sobic.006G005600	**A992422 vs. A992422×QL36**	similar to Heat shock protein 82	unfolded protein binding, response to stress, protein folding, ATP binding	5.75	0.00	*AT5G52640*	*GRMZM5G833699*
Sobic.006G079500	**AQL33 vs. AQL33×QL36**	N/A	metabolic process, catalytic activity	2.76	0.00	*AT1G62940*	*GRMZM2G014651*
Sobic.006G079500	**A992422 vs. A992422×QL12**	N/A	metabolic process, catalytic activity	23.35	0.00	*AT1G62940*	*GRMZM2G014651*
Sobic.006G148800	**A992422 vs. A992422×QL12**	similar to OSJNBa0073E02.14 protein	N/A	-2.01	0.00	*AT2G37040*	*GRMZM2G029048*
Sobic.006G148900	**A992422 vs. A992422×QL12**	similar to Phenylalanine ammonia-lyase	N/A	-4.05	0.00	*AT2G37040*	*GRMZM2G029048*
Sobic.007G029900	**AQL33 vs. AQL33×QL36**	similar to Putative cytochrome P450 family	N/A	3.35	0.00	*AT1G01280*	*GRMZM5G830329*

Table A-1: Continued

Feature ID	Comparison	Annotation	GO terms	Fold change[a]	FDR *p*-value	Gene ortholog	
						Arabidopsis	**Maize**
Sobic.007G029900	**A992422 vs. A992422×R931945-2-2**	similar to Putative cytochrome P450 family	N/A	6.43	0.00	*AT1G01280*	*GRMZM5G830329*
Sobic.007G029900	**A992422 vs. A992422×QL12**	similar to Putative cytochrome P450 family	N/A	39	0.00	*AT1G01280*	*GRMZM5G830329*
Sobic.007G029900	**AQL33 vs. AQL33×R931945-2-2**	similar to Putative cytochrome P450 family	oxidation reduction, oxidoreductase activity	-12.9	0.00	*AT1G01280*	*GRMZM5G830329*
Sobic.007G068700	**AQL33 vs. AQL33×QL36**	similar to Polyphenol oxidase	oxidation reduction, metabolic process, oxidoreductase activity	3.63	0.00	N/A	N/A
Sobic.007G206000	**AQL33 vs. AQL33×R931945-2-2**	similar to Putative dihydroflavonol reductase	coenzyme binding, catalytic activity	-4.50	0.00	*AT4G35420*	*GRMZM2G004683*
Sobic.006G148800	**A992422 vs. A992422×QL12**	similar to OSJNBa0073E02.14 protein	N/A	-2.01	0.00	*AT2G37040*	*GRMZM2G029048*
Sobic.007G206000	**AQL33 vs. AQL33×QL36**	similar to Putative dihydroflavonol reductase	coenzyme binding, catalytic activity	2.17	0.00	*AT4G35420*	*GRMZM2G004683*
Sobic.007G206000	**ATx623 vs. ATx623×RTx2783**	similar to Putative dihydroflavonol reductase	coenzyme binding, catalytic activity	2.39	0.00	*AT4G35420*	*GRMZM2G004683*
Sobic.007G206000	**A992422 vs. A992422×QL36**	similar to Putative dihydroflavonol reductase	coenzyme binding, catalytic activity	3.62	0.01	*AT4G35420*	N/A
Sobic.007G206000	**A992422 vs. A992422×R931945-2-2**	similar to Putative dihydroflavonol reductase	coenzyme binding, catalytic activity	6.01	0.00	*AT4G35420*	N/A

Table A-1: Continued

Feature ID	Comparison	Annotation	GO terms	Fold change[a]	FDR *p*-value	Gene ortholog	
						Arabidopsis	**Maize**
Sobic.007G206000	**A992422 vs. A992422×QL12**	similar to Putative dihydroflavonol reductase	coenzyme binding, catalytic activity	15.10	0.00	*AT4G35420*	*GRMZM2G004683*
Sobic.007G216300	**AQL33 vs. AQL33×QL36**	similar to Heat shock protein 81-2	unfolded protein binding, response to stress, protein folding, ATP binding	-5.26	0.00	*AT5G56000*	N/A
Sobic.007G216300	**AQL33 vs. AQL33×R931945-2-2**	similar to Heat shock protein 81-2	unfolded protein binding, response to stress, protein folding, ATP binding	-2.41	0.00	*AT5G56000*	*GRMZM2G012631*
Sobic.007G216300	**A992422 vs. A992422×QL36**	similar to Heat shock protein 81-2	unfolded protein binding, response to stress, protein folding, ATP binding	2.46	0.00	*AT5G56000*	*GRMZM2G012631*
Sobic.008G087300	**ATx623 vs. ATx623×RTx2783**	similar to 3-oxoacyl-reductase, chloroplast, putative, expressed	N/A	4.37	0.00	*AT1G24360*	*GRMZM2G043602*
Sobic.008G087300	**AQL33 vs. AQL33×QL36**	similar to 3-oxoacyl-reductase, chloroplast, putative, expressed	N/A	7.00	0.00	*AT1G24360*	*GRMZM2G043602*
Sobic.008G136000	**AQL33 vs. AQL33×QL36**	similar to Heat shock cognate 70 kDa protein 2	N/A	-2.72	0.00	*AT3G12580*	*AC194017.3_FG001*
Sobic.007G206000	**A992422 vs. A992422×R931945-2-2**	similar to Putative dihydroflavonol reductase	coenzyme binding, catalytic activity	6.01	0.00	*AT4G35420*	N/A

Table A-1: Continued

Feature ID	Comparison	Annotation	GO terms	Fold change[a]	FDR *p*-value	Gene ortholog	
						Arabidopsis	**Maize**
Sobic.009G162000	**AQL33 vs. AQL33×QL36**	similar to Putative uncharacterized protein	N/A	8.66	0.00	*AT3G11430*	*GRMZM2G059637*
Sobic.009G163900	**AQL33 vs. AQL33×QL36**	similar to Heat shock cognate 70 kDa protein	N/A	-8.33	0.00	*AT3G12580*	*AC194017.3_FG001*
Sobic.009G163900	**AQL33 vs. AQL33×R931945-2-2**	similar to Heat shock cognate 70 kDa protein	N/A	-4.08	0.00	*AT3G12580*	*AC194017.3_FG001*
Sobic.009G163900	**AQL33 vs. AQL33×QL12**	similar to Heat shock cognate 70 kDa protein	N/A	-2.62	0.00	*AT3G12580*	*AC194017.3_FG001*
Sobic.009G163900	**ATx623 vs. ATx623×RTx436**	similar to Heat shock cognate 70 kDa protein	N/A	3.07	0.03	*AT3G12580*	*AC194017.3_FG001*
Sobic.009G163900	**ATx623 vs. ATx623×RTx2783**	similar to Heat shock cognate 70 kDa protein	N/A	4.03	0.00	*AT3G12580*	*AC194017.3_FG001*
Sobic.009G163900	**A992422 vs. A992422×QL36**	similar to Heat shock cognate 70 kDa protein	N/A	4.58	0.00	*AT3G12580*	N/A
Sobic.010G106300	**AQL33 vs. AQL33×R931945-2-2**	similar to Ribonucleoside-diphosphate reductase small chain	oxidation reduction, deoxyribonucleoside diphosphate	2.72	0.04	*AT3G27060*	*GRMZM2G060163*
Sobic.010G149300	**AQL33 vs. AQL33×R931945-2-2**	similar to SUMO-conjugating enzyme UBC9	N/A	-3.22	0.00	*AT1G64230*	*AC233922.1_FG008*
Sobic.009G163900	**AQL33 vs. AQL33×QL36**	similar to Heat shock cognate 70 kDa protein	N/A	-8.33	0.00	*AT3G12580*	*AC194017.3_FG001*

[a]Positive values indicate greater expression in B-lines or F_1 hybrids compared to A_1-lines. Negative values indicate greater expression in A_1-lines compared to B-lines or F_1 hybrids.