BIOCHEMICAL, MOLECULAR, AND GENETIC ANALYSIS OF CELL WALL BIOGENESIS AND CARBOHYDRATE ACCUMULATION OF THE STEM DURING THE PHENOLOGY OF *SORGHUM BICOLOR*

A Dissertation

by

BRIAN ADAM MCKINLEY

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

DOCTOR OF PHILOSOPHY

Chair of Committee,	John Mullet
Committee Members,	Bill Park
	Scott Finlayson
	Vlad Panin
Head of Department,	Gregory Reinhart

December 2016

Major Subject: Biochemistry

ABSTRACT

Alternative sources of fuel such as bioethanol offer benefits such as energy independence and stability, reduction in green house gas emissions, and are a renewable resource. The C4 grasses are large contributors of fermentable carbohydrates for bioenthanol production. However, fermentable carbohydrate yields from grass stems have plateaued in recent years. Transgenic manipulation may provide solutions to engineering increased nonstructural carbohydrate yields, increase cell wall digestibility and accessibility, and reduce agricultural input costs. *Sorghum bicolor* is a prime candidate as a biofuel feedstock, but lacks the molecular characterization essential for transgenic manipulation. Therefore, the focus of this research was to conduct a comprehensive characterization of the biochemical, molecular and genetic dynamics of nonstructural carbohydrate accumulation and cell wall biogenesis in the sorghum stem. To facilitate this investigation a high biomass accumulating elite sweet sorghum named Della was selected as a model since this sorghum accumulates high biomass and therefore closely approximates a sweet energy sorghum.

Analysis of dry biomass accumulation revealed that Della invests heavily in stem growth starting at the beginning of the vegetative phase and ending at anthesis, subsequently accumulating sugar in its stem during flowering. To identify genes involved in cell wall biogenesis and carbohydrate accumulation, a time-course of Della internode transcriptome dynamics was conducted

beginning at floral initiation and ending ~20 days after grain maturity encompassing the time course of carbohydrate accumulation in the stem. This analysis identified candidate genes involved in cell wall biosynthetic processes and nonstructural carbohydrate partitioning and accumulation were identified.

Della synthesizes starch in its stem after anthesis. Since starch is more stable than sucrose, has favorable osmotic properties, and is already present in the sorghum stem, this pathway was selected as the focus for further analysis. This analysis revealed candidate genes that will be useful for future transgenic studies. Additionally, stem biomass from Della was observed to have accumulated starch up to ~15% of the total stem dry biomass and ~30% of total nonstructural carbohydrate content. Therefore, utilization of existing starch stores in sorghum stem biomass would increase fermentable carbohydrate yield significantly.

Lastly, genetic analysis of the grain*sweet sorghum cross RTx436*Della identified loci that underlie stem volume traits when mapping quantitative trait loci for internode carbohydrate content. This suggests that stem carbohydrate concentration reached a maximum level and was limited by volume. Also, carbohydrate was found to be correlated with internode volume and not stem juiciness and leaf area. Finally, radiation use efficiency analysis revealed that at certain points during development the canopy may be underutilized and that increasing the sink strength of the sorghum stem may benefit nonstructural carbohydrate yield.

NOMENCLATURE

AGPase	ADP-glucose pyrophosphorylase
CIN	Cytoplasmic invertase
DAE	Days after emergence
FDR	False discovery rate
GAX	Glucuronoarabinoxylan
GHG	Green house gas
GWD	Glucan-water dikinase
INV	Invertase
LAI	Leaf area index
LC	Lignocellulose
LG	Linkage group
LOD	Log of odds
MLG	Mixed linkage glucan
NIR	Near infrared spectroscopy
NREL	National Renewable Energy Laboratory
NSC	Nonstructural carbohydrate(s)
PAR	Photosynthetically active radiation
PCC	Pearson correlation coefficient
PWD	Phosphoglucan-water dikinase
QTL	Quantitative trait locus

RIL	Recombinant inbred line
RPKM	Reads per kilobase mapped per million mapped
RUE	Radiation use efficiency
SAM	Shoot apical meristem
SPS	Sucrose phosphate synthase
SUS	Sucrose synthase
TF	Transcription factor
VIN	Vacuolar invertase

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

INTRODUCTION*

The C4 grasses have excellent potential as bioenergy crops (Byrt *et al.* 2011, Mullet *et al.* 2014, Sage and Zhu 2011, van der Weijde *et al.* 2013). Currently, the C4 grasses maize, grain sorghum, and sugarcane provide most of the feedstock for production of bio-ethanol (Ranum *et al.* 2014, Waclawovsky *et al.* 2010). Sugarcane grown on ~25M ha is a source of sucrose, bio-ethanol produced from sucrose, and bio-power generated from lignocellulosic bagasse (Waclawovsky*, et al.* 2010). In 2011, approximately 54% of the world's sugarcane crop was used to generate 27B liters of bio-ethanol (Dal-Bianco *et al.* 2012). In 2013, the U.S. used 40% of the maize grain crop (Ranum*, et al.* 2014) and 29% of the sorghum grain crop (Godoy and Tesso 2013) to produce ~50B liters of bio-ethanol, approaching the mandated limit of grain-based ethanol production in the U.S. (Rahall 2007). While ethanol production from maize grain starch is economical, the resulting energy output:input ratio is low (Hill *et al.* 2006), and greenhouse gas (GHG) mitigation is significantly less than from lignocellulosic bio-ethanol derived from C4 grasses such as switchgrass, *Miscanthus*, sugarcane, and sweet sorghum (Byrt*, et al.* 2011). However, only small amounts of lignocellulosic (LC) biofuels are being produced due to the aggregate costs of LC-feedstock, biomass transport to biorefineries,

*Reprinted with permission from "Dynamics of biomass partitioning, stem gene expression, cell wall biosynthesis, and sucrose accumulation during development of Sorghum bicolor" by Brian McKinley, William Rooney, Curtis Wilkerson and John Mullet, 2016. *The Plant Journal*, 10.1111/tpj.13269. Copyright 2016 by John Wiley and Sons.

pretreatment, and LC-conversion technologies. Techno-economic analysis of ethanol production from lignocellulosic biomass found that the cost of LC-feedstock represents 36-38% of the cost of bio-ethanol production and conversion of LC-biomass to biofuels accounted for an additional 25% of production costs (i.e., pretreatment, enzymes, fermentation) (Gnansounou and Dauriat 2010). This indicates that commercial production of cost competitive biofuels from bioenergy crops will require development of higher yielding crops that produce feedstock at lower cost, optimization of feedstock composition, and improved technology for conversion of LC-feedstocks into next generation biofuels and bioproducts.

Bioenergy crops such as energy sorghum, switchgrass, *Miscanthus*, and poplar are being developed to support a sustainable and economically viable lignocellulosic biofuels industry (Mullet*, et al.* 2014, Somerville *et al.* 2010, van der Weijde*, et al.* 2013). The key design features of these bioenergy crops include high biomass yield, low input requirements, high water and nitrogen use efficiency, drought and heat resilience, pest resistance, and biomass composition that aids low cost high efficiency production of biofuels and bio-products. Commercial production of energy sorghum is a recent development because an efficient method for breeding energy sorghum hybrids was only developed during the past ten years (Rooney *et al.* 2007). Identification of sorghum lines with complementary flowering time genotypes (Rooney and Aydin 1999) and elucidation of sorghum's photoperiod responsive flowering time

regulatory pathway (Murphy *et al.* 2011, Murphy *et al.* 2014, Yang *et al.* 2014a, Yang *et al.* 2014b) is now enabling marker-assisted breeding of energy sorghum hybrids. First generation energy sorghum hybrids have high biomass yield (~20-40 dry Mg/ha) (Gill *et al.* 2014, Olson *et al.* 2012) and high nitrogen use efficiency (Olson *et al.* 2013) due to C4 photosynthesis (Byrt*, et al.* 2011), long duration of biomass accumulation (Olson*, et al.* 2012), and efficient N-recycling (Olson*, et al.* 2013). Depending on yield in a given year/location, energy sorghum biomass provides ~70-75% GHG displacement when used to produce bio-ethanol and >90% for bio-power (Olson*, et al.* 2012). Taken together, energy sorghum's high biomass yield, drought and heat resilience, water use efficiency, and low input requirements make this crop ideal for use on cultivated land that is marginal for annual food crops (Mullet*, et al.* 2014).

The optimization of energy sorghum biomass composition for low cost high efficiency conversion to biofuels and bio-products is needed to improve the economics of production (Dien *et al.* 2009, Martin *et al.* 2013). Sorghum germplasm has a large range of biomass composition that is available for breeding improved energy sorghum (Stefaniak *et al.* 2012). Sorghum breeding historically has focused on the development of short stature early flowering grain sorghum hybrids that accumulate large amounts of grain starch, and forage sorghums with increased leaf biomass yield, and digestibility (Rooney 2004). In contrast, stem biomass is the main product produced by energy sorghum. Some energy sorghum genotypes produce stems with high lignocellulose

content and low amounts of sugar whereas other energy sorghum genotypes, generally referred to as sweet sorghums, can accumulate up to ~50% of their stem biomass in the form of sugars, principally as sucrose (Mcbee and Miller 1982, Monk *et al.* 1984, Slewinski 2012, Wang *et al.* 2009). Sweet sorghum's capacity for stem sucrose accumulation is similar to sugarcane, a closely related perennial C4 grass. Research on sweet sorghum has shown that accumulation of stem sugars occurs after floral induction, and is correlated with completion of stem and leaf growth and a reduction in the activity of enzymes that mediate sucrose hydrolysis (Gutjahr *et al.* 2013a, HoffmannThoma *et al.* 1996, Lingle 1989, Tarpley *et al.* 1994b). The sucrose that accumulates in large pith parenchyma cell vacuoles of sorghum stems is derived from leaves via the phloem, unloaded into the apoplast, taken up by pith cells and sequestered in vacuoles with minimal hydrolysis (Tarpley and Vietor 2007). While the overall dynamics of sucrose accumulation in sweet sorghum stems have been characterized, the full complement and regulation of the genes involved require further analysis (Calvino *et al.* 2011).

Structural carbohydrates that comprise lignocellulosic cell walls represent a significant fraction of energy sorghum biomass that could be used for production of biofuels, bio-products and bio-power. Sorghum *brown mid-rib (bmr)* genotypes that have modified lignin content and chemistry have already been developed to enhance LC-biomass digestibility for forage use and to enhance saccarification and fermentation efficiency (Dien*, et al.* 2009, Jung *et*

al. 2012, Saballos *et al.* 2012, Sattler *et al.* 2009). Other modifications of cell wall lignin (i.e., zip-lignins) (Wilkerson *et al.* 2014), xylan content (Biswal *et al.* 2015), cell wall methylation (Lionetti *et al.* 2010) and mixed linkage glucan content (Vega-Sanchez *et al.* 2015), may also be useful to deploy in sorghum to improve LC-biomass conversion efficiency. Targeted engineering of sorghum stem cell walls will be aided by the identification of genes involved in stem cell wall biosynthesis, and characterization of their temporal and spatial patterns of expression, a goal of the current study.

Sorghum genotypes vary in their partitioning of photosynthate between lignocellulose, stem sugars, and grain starch providing the opportunity to select energy sorghum genotypes that accumulate high levels of non-structural carbohydrates (i.e., sugars, starch) and that have cell walls that can be converted efficiently to biofuels and bio-products (Calvino*, et al.* 2011, Slewinski 2012, Vermerris 2011). Stems can account for ~84% of energy sorghum biomass therefore optimization of stem composition is a central focus of research aimed at improving biomass conversion efficiency. However, our understanding of the molecular basis of sorghum stem biogenesis, partitioning between structural and non-structural carbohydrates, and information about genes that are involved in sorghum stem cell wall biosynthesis is incomplete. Previous studies of sorghum stem gene expression focused primarily on the mechanism of stem sucrose accumulation utilizing comparisons of grain and sweet sorghum genotypes and targeted analysis of gene expression (Qazi *et al.*

2012, Tarpley*, et al.* 1994b), microarrays, and cDNA/small RNA sequencing (Calvino *et al.* 2008, Calvino*, et al.* 2011, Jiang *et al.* 2013).

Sorghum is a non-model species of which little is known concerning the molecular biology and genetic of the traits being improved. Information regarding the candidate genes is essential for genetic engineering. The overall purpose of this investigation is to obtain this information as it relates to the composition of the sorghum stem to facilitate further genetic manipulation.

CHAPTER II

DYNAMICS OF BIOMASS PARTITIONING, STEM GENE EXPRESSION PROFILES, CELL WALL BIOSYNTHESIS, AND SUCROSE ACCUMULATION DURING DEVELOPMENT OF *SORGHUM BICOLOR**

INTRODUCTION

Sorghum development progresses through juvenile, vegetative and reproductive phases during its phenology with each phase imposing different requirements on carbohydrate allocation during its duration. During the juvenile and early vegetative phases, sorghum builds out the canopy and root system that will subsequently fuel stem growth followed by stem and grain filling. After the juvenile to vegetative phase transition (25-30DAE), the stem begins elongation which continues until anthesis with successive internodes completing their development as they are produced by the shoot apical meristem (SAM) (Murphy*, et al.* 2014). During this stage the plant invests large quantities of photoassimilate into stem metabolism and cell wall biosynthesis. At floral initiation the SAM undergoes the floral transition to produce the panicle which prevent the SAM from producing new internodes. Stem and canopy growth finalize at anthesis, which marks a period of time in which the demand for photoassimilate is low while the photosynthetic capacity of the plant is high due

to the mature canopy. Sorghum uses this excess photoassimilate to fill its stem with sucrose and the grain with starch. Previous studies have shown that the pattern of sucrose accumulation can be variable, peaking sometime before or after grain maturity, but there is generally a peak that occurs during the reproductive phase (Gutjahr *et al.* 2013b, Tarpley *et al.* 1994a).

The cessation of growth at anthesis and the subsequent accumulation of sucrose that follows is possibly affected by, and likely has a significant impact on, the internode transcriptome during the phenology. Identification of the genes that are involved in these processes will not only clarify fundamental aspects of the biology of the stem but will also identify specific members of the gene families involved in these processes that can be used for follow-on transgenic manipulation. To this end. a time-course of dynamic changes in the sorghum stem transcriptome using RNA-seq and quantified changes in stem composition during sorghum stem development. Sorghum stem gene expression was characterized starting at the transition from vegetative growth to floral initiation, and extending through anthesis, grain maturity, and for ~30 days post grain maturity. The results identified differentially expression genes involved in growth, cell wall biosynthesis, and sucrose accumulation that are sequentially expressed in stems during development.

EXPERIMENTAL PROCEDURES

Field experiments

Della [*Sorghum bicolor* (L.) Moench] (Reg. no. CV-130, PI1566819), a sweet sorghum developed from a cross of Dale and ATx622 (Harrison and Miller 1993) was planted at the Texas A&M Research Farm in College Station on April 2 in 2012 and March 28 in 2013. For each year, two ranges consisting of 10 plots with 50 m row length, of Della were planted in adjacent regions of the field. The row spacing was 76 cm and plants in plots were thinned to 10 cm spacing at the five-leaf stage. This spacing corresponded to ~132,000 plants per ha^{-1} calculated as previously described (Olson*, et al.* 2012). Prior to planting, diammonium phosphate (10-34-0) was applied at a rate of 100 kg N ha^{-1}. Basal tillers were removed until anthesis to maintain a plant density of one main culm per 10 cm. Basal tillering was not observed post anthesis. Apical tillers (or branches) produced by plants started emerging a week before grain maturity and were not removed from field grown plants. In both years, five plants were harvested from each of two experimental replications for a total of 10 plants per time-point. Plants were not harvested from the outside row or from the ends of plots to avoid plot edge effects. Harvesting began six days before anthesis in 2012 and 45 days before anthesis in 2013. The fresh weights of the leaves, leaf sheaths, stems and panicles were recorded. The tissues were placed in paper bags and dried in a forced air oven at 60°C for three days. The tissues were re-weighed to obtain dry weights. In 2012 and 2013, internodes 12-14 were

harvested for composition analysis. In 2012, internodes 12-14 were fully elongated at the first harvest. In 2013, harvests were started at an earlier stage of vegetative phase development. Due to the small size of the stem at the A-45d and A-29d time-points the entire stem was harvested. In later harvests the stem was divided into four sections; (1) internode 11 and all lower internodes, (2) internode 12-14, internode 13 and more apical internodes (if elongated) and (4) the peduncle (if present). This method of harvesting allowed for the total dry biomass of the stem to be obtained from the sum of the dry biomass of all the sections but also allowed for sampling in a way that could be aligned with internode harvesting used in RNA-seq. Once internodes 12-14 were present (A-19d and in older plants), those internodes were harvested for the remainder of the experiment. Internode 12-14 may not have been fully elongated at the time of harvest at A-19d.

Near infrared spectroscopic analysis of internode composition

Stem composition was analyzed using near infrared spectroscopy (NIR) and HPLC (see below). Leaves and leaf sheaths were removed from the stem prior to analysis. Internodes 12-14 from field grown plants were excised at the nodes producing stem sections that consisted of three internodes. Following excision the internodes were dried in an oven at 60°C, ground and analyzed with NIR for composition analysis as previously described (Stefaniak*, et al.* 2012, Wolfrum *et al.* 2013). The dry biomass of internode 12-14 was ground using a Wiley Mill (Thomas Scientific, Inc.) until the biomass particles could pass through a 2 mm

sieve. Before NIR analysis, the ground internode samples were re-dried to remove residual moisture, thereby minimizing variation in moisture content between samples. NIR spectra were acquired using the stationary module of a Foss XDS grating-monochromator (Foss North America, Eden Prairie, MN). The samples were scanned in reflectance mode across the wavelength range of 4,000 - 9,000 cm^{-1}. Two NIR spectra of each sample were collected on separate occasions to ensure reproducibility of measurement. The spectra were analyzed using a previously published composition prediction model developed for sorghum by NREL (Stefaniak*, et al.* 2012).

Soluble carbohydrate quantitation

To quantify the abundance of stem non-structural carbohydrates, internode stem tissue was ground to a small particle size in a Cyclone Sample Mill (Udy Corporation, Fort Collins, Colorado, USA) followed by removal of residual moisture overnight using a forced air oven at 60°C. Next, 200 mg of finely ground biomass was weighed (± 0.5 mg) using an analytical balance and transferred to a 15 mL conical glass tube. Water-soluble NSCs were extracted in 10 mL of water/Na azide (200 mgL^{-1}) solution at 50°C for 48 hours with agitation. This length of incubation was experimentally determined to be optimal for this experimental set-up. This extraction time allowed the extraction solvent to fully penetrate all of the biomass particles and extract the soluble carbohydrates. Aliquots of 20 µL were diluted 50X into 980 µL of de-ionized water. All HPLC samples were filtered using 0.45 µm cellulose acetate sterile

filters. Sucrose, glucose, and fructose concentrations were quantified using high performance anion exchange-pulsed amperometric detection (HPAE-PAD) with a Carbopac PA1 analytical column (Dionex) as well as a Borate trap (Dionex) and an Aminopac column (Dionex) to reduce borate and amino acid interference (Dionex 2003b, Murray *et al.* 2008). A solution of 75 mM NaOH was made from 50% NaOH solution (Sigma-Aldrich) instead of pellets to eliminate carbonate contamination. This running buffer was vacuum degassed overnight and stored under a helium bed for the duration of the chromatographic run. The standard curve was validated by the incorporation of curve validation samples of known concentration throughout the experiment in accordance to the NREL LAP (Sluiter and Sluiter 2005).

Stem starch quantitation

Stem starch was quantified using the NREL Laboratory Analytical Procedure for extracting starch from dry biomass (Sluiter and Sluiter 2005). After extraction of soluble carbohydrates, samples were washed four additional times with agitation with methanol-chloroform-water extraction solvent. The samples were centrifuged at 3500 rpm for 15 minutes after each wash. The samples were oven dried at 60°C. To gelatinize the starch, DMSO was added to the dry samples and allowed to permeate the samples for 24 hours in refrigeration. This step ensures complete saturation of the dry biomass with DMSO and complete gelatinization of the starch. The samples were then incubated in a water bath at 100°C for 15 minutes with occasional vortexing to facilitate starch gelatinization.

MOPS buffer and thermostable α-amylase were added to digest the starch to maltose. Following amylase digestion, sodium acetate buffer and amyloglucosidase were added to hydrolyze the maltose to glucose. Glucose was quantified with HPLC as described above using HPAE-PAD. The percentage of starch in the dry biomass was calculated from the glucose released during digestion.

Della stem tissue harvest for RNA-seq

To obtain tissue for RNA-seq, Della was grown in pots filled with Metromix in a greenhouse in 2011 with a day length of 14 hours and mid-day PAR of ~1100-1200 $\mu E/m^2/s$. Plants were thinned to one plant per pot. Osmocote was added to the soil at the beginning of the experiment and supplemented with Peter's nutrient solution every 30 days. The plants were watered with an automated watering system to maintain plants in water sufficient conditions. Leaves were numbered as they appeared ensuring that internode 10 was harvested at each of the eight sampling dates. Three biological replicates were harvested in the morning on each harvest date. Leaves and leaf sheaths were removed prior to harvesting internode tissue. Internode 10 is located in the stem between internode 9 (below) and internode 11 (above). Internode 10 was excised from the stem by cutting the nodes located above and below internode 10 at their midpoints, the internode was quickly sectioned into smaller pieces with a razor and placed into a 50 mL conical tube, frozen in liquid nitrogen, and stored at -80°C until processing.

Sequencing, transcriptome alignment, and calculation of differential expression

Internode 10 was pulverized into a fine powder in liquid nitrogen with a mortar and pestle. RNA extraction was performed using the Trizol extraction method. The high salt precipitation step was included because of the high concentration of carbohydrates in some of the samples (Molecular Research Center, http://mrcgene.com/rna-isolation/tri-reagent). Total RNA was further purified using the RNeasy kit from Qiagen which isolates RNAs >200 nucleotides in length. An on column DNase1 digestion was performed to remove any contaminating DNA. The quality of the RNA libraries was assessed using an Agilent Bioanalyzer. The mean RNA integrity number (RIN) from this quality control step was 8.9 and the minimum RIN was 8.3 indicating that the quality of the RNA libraries was high. RNA libraries were prepped for sequencing using the Illumina TruSeq® V2 mRNA Sample Preparation Kit, which selects for polyadenylated RNA and excludes rRNA and non-polyadenylated RNAs from the library. The 24 libraries were sequenced on one lane of an Illumina HiSeq 2500 in single end mode. The average sequence depth per sample was 5.8M reads. The resulting 70bp reads were aligned to the Sbicolor-v2.1_255 using CLC Genomics Workbench version 6.5.1 and version 7 (CLC Inc., Aarhus, Denmark). Sequence reads were aligned to the 33,032 gene models annotated in Sbicolor-v2.1_255 provided through Phytozome (Paterson *et al.* 2009).

Analysis of gene expression was performed on RPKM normalized data. Differential expression and the FDR-adjusted p-values were calculated using the edgeR package in CLC Genomics Workbench that uses non-normalized alignment data as the input for the edge statistical test. Further functional analysis was confined to genes that exhibited expression ≥ 2 RPKM at one or more time points, differential expression ≥ 3 between any two time-points of the experiment, and a FDR-adjusted p-value of differential expression < 0.05. The FDR adjusted p-value represents the p-value adjusted for multiple testing which was associated with the maximum fold-change values shown in the gene tables. In some cases, low expression values (RPKM<1) at one time-point led to a high differential expression value. To minimize this effect, the time-point with low expression was set to a RPKM = 1, followed by manual calculation of the fold change using the calculation (Fold Change = RPKM/1). These instances are indicated with the ">" sign in front of the fold change values in the gene tables. To analyze/visualize variation between samples and time-points, principal component analysis was performed. This analysis was conducted first in CLC Genomics Workbench followed by rendering in Excel.

Sorghum gene annotation

Gene function analysis utilized gene annotations from multiple resources. A detailed resource of annotated genes that contribute to cell wall biology was developed for Arabidopsis, maize, rice and sorghum by Penning et al, 2009 (https://cellwall.genomics.purdue.edu/families/index.html). The resource

provides genes clustered by gene families based on sequence similarity with manual curation to ensure high quality. The sorghum genes identified in this resource were extracted from the gene family clusters. For other families of genes involved in sucrose, callose, and MLG metabolism, annotations were downloaded from the Sbicolor-v2.1_255 reference genome from Phytozome. One difficulty of large datasets is ensuring a complete overview of each gene family/functional category. To ensure that our analysis was comprehensive and to aid in data visualization Mapman was utilized because of its ability to group sets of genes into general functional categories. Sorghum orthologs of GAX pathways were identified as the top TBLASTN hit to ether rice or Arabidopsis query genes identified from the literature (Dionex 2003b, Murray*, et al.* 2008).

RESULTS

Della development and biomass accumulation

Della, a sweet sorghum inbred line, was planted at the Texas A&M University Field Station in College Station Texas between April 2 and April 10 in 2012 and 2013. Plant growth and development was similar in both years although some variation was observed due to differences in environmental conditions and plot location. The timing and duration of developmental stages and the rate of biomass accumulation of Della plants under field conditions is shown in Figure 2.1. Della plants underwent floral initiation between 45-50 days after emergence (DAE), reached anthesis by ~75-80 DAE, and grain maturity between ~106-110 DAE (Figure 2.1a). The current study focused on changes in biomass

accumulation, composition and gene expression that occurred between floral initiation and 30 days post grain maturity (~100 days).

The number of days from sorghum seedling emergence to floral initiation can vary significantly (>100 days) depending on genotype, day length, temperature, and other environmental factors. However, the number of days from floral initiation to anthesis and from anthesis to grain maturity is relatively consistent in sorghum under good growing conditions (Vanderlip and Reeves 1972). Therefore, the timing of events and other data collected in this study is expressed in days relative to anthesis (i.e., days before anthesis (A-#d) or after anthesis (A+#d)), in addition to days from plant emergence (DAE). Following germination, seedlings accumulated biomass slowly for approximately 50 days and then at an increased rate until grain maturity at ~106 DAE (A+27d) (Figure 2.1b). Leaf area per plant began increasing rapidly between 35 and 40 DAE reaching a maximum of ~2,000 cm^2 between ~70-75 DAE, a few days before anthesis (0d). Stem length, excluding the peduncle, increased rapidly following floral initiation (~50 DAE, A-29d) due to production of longer internodes. Stem growth rate decreased starting ~10-14 days before anthesis with final stem length reached by anthesis when the last internode below the peduncle was full length.

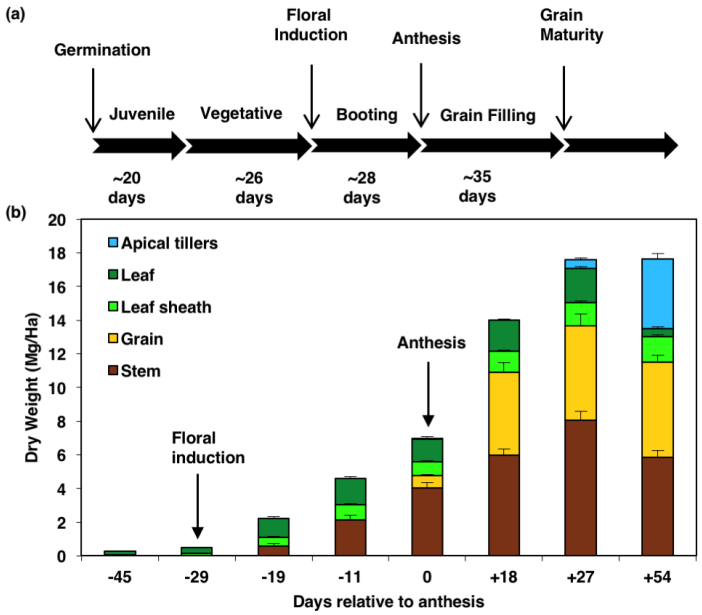

Figure 2.1. Time course of biomass accumulation during Della development.
(a) Duration of stages of Della development from germination to floral induction, floral induction and booting to anthesis, and anthesis to grain maturity.
(b) Time course of dry weight accumulation (Mg per hectare) in leaves (dark green), leaf sheaths (light green), stems (brown), grain (yellow) and apical tillers (blue) in the field in College Station, 2013. Under field conditions anthesis occurred at 79 days after emergence (DAE) and grain maturity at 106 DAE. The data shown are the mean ±SE of nine biological replicates.

Della shoots accumulated ~18 Mg ha^{-1} of dry biomass under field conditions in 2013 by grain maturity with grain accounting for 5.6 Mg ha^{-1} (Figure 2.1b, yellow bar). Dry biomass of leaves on the main stem reached a maximum between anthesis and grain maturity (~2 Mg ha^{-1}) (Figure 2.1b, green bar). Stem dry biomass was ~4 Mg ha^{-1} at anthesis and reached ~8 Mg ha^{-1} at grain maturity (A+27d; 106 DAE) (Figure 2.1b, brown bar). Stem dry biomass declined after grain maturity coincident with an increase in apical tillering. Apical tillers,

which began growing a few days before grain maturity, accumulated 4.1 Mg ha^{-1} by 133 DAE (A+54d) (Figure 2.1b, blue bar).

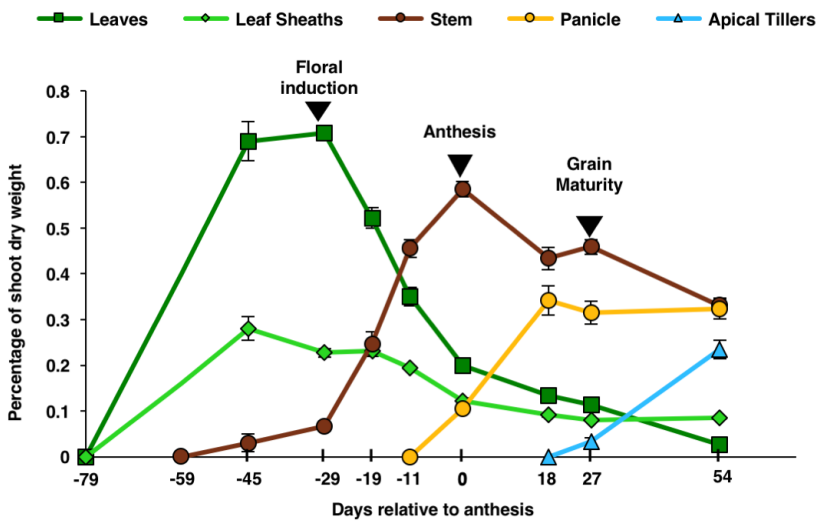

Figure 2.2. Relative dry weight of shoot tissues during Della development in the field at College Station in 2013. Dry weight of leaves, leaf sheaths, stems, panicles and tillers is plotted as a percentage of the total shoot dry biomass at each time point during development of Della. The data represent the mean ±SE of nine biological replicates. Arrows indicate the timing of floral initiation, anthesis and grain maturity. Stem growth began at 59 days before anthesis.

The relative rate of biomass accumulation in leaves, leaf sheaths, stems, panicles, and branches during development was analyzed in terms of harvest index (organ dry weight/total shoot dry weight) (Figure 2.2). The analysis showed that shoot biomass primarily accumulated in leaves and leaf sheaths from germination until ~45-50 DAE (A-29d) (Figure 2.2). From 50 DAE to anthesis (~75-80 DAE), biomass was partitioned preferentially into stems during rapid stem elongation. From anthesis to grain maturity biomass was

accumulated preferentially in the panicle/grain and post-grain maturity, in apical tillers.

Variation in stem composition during development

Grass stems contain lignocellulose (LC) comprised principally of the cell wall biopolymers cellulose, GAX, and lignin, non-structural carbohydrates (NSC) such as sucrose, glucose, fructose and starch, as well as proteins, lipids, water-soluble metabolites, and inorganic materials (ash). The relative amount of these compounds in Della stems at different stages of development was quantified using near infrared spectroscopy (NIR) (Stefaniak*, et al.* 2012). At 63 DAE (A-6d), cellulose, xylan, and lignin, the main cell wall biopolymers, accounted for ~55% of total stem dry weight (Figure 2.3). This fraction of stem dry weight declined to ~27% post-anthesis as sucrose and starch accumulated. Sorghum stem cell walls were comprised of ~47-51% cellulose, ~24-28% xylan, ~21-25% lignin, and <5% arabinan and galactan. The relative abundance of these biopolymers in stem cell walls changed to only a small extent during development (Figure 2.4).

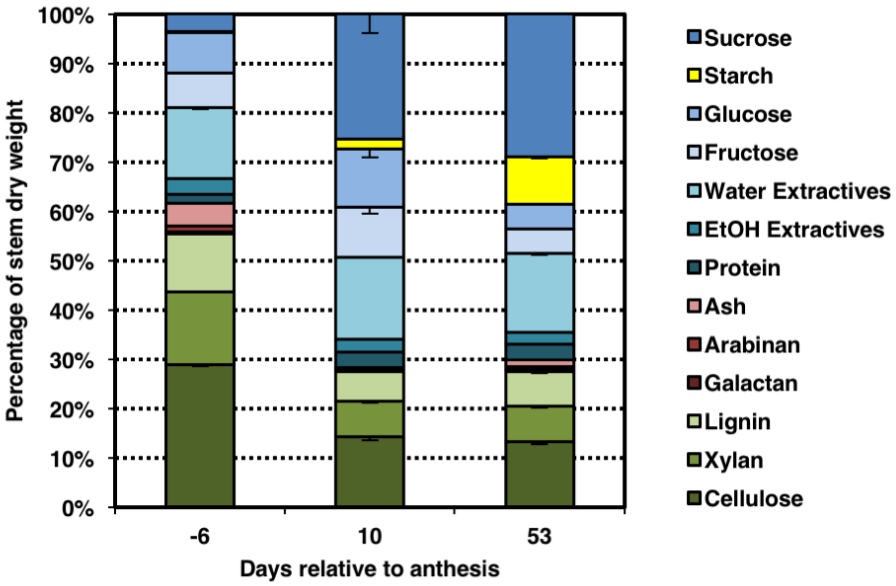

Figure 2.3. Composition of Della stems at three stages of development. Stems were collected from Della plants grown in College Station in 2012. Data shown represents the internode composition just after floral initiation (63 DAE), at anthesis (79 DAE) and post grain maturity (122 DAE). Colored bars comprising each analysis represent the average of NIR data collected from nine biological replicates. Error bars represent standard error. Sucrose, glucose, fructose and starch were determined independently by HPLC.

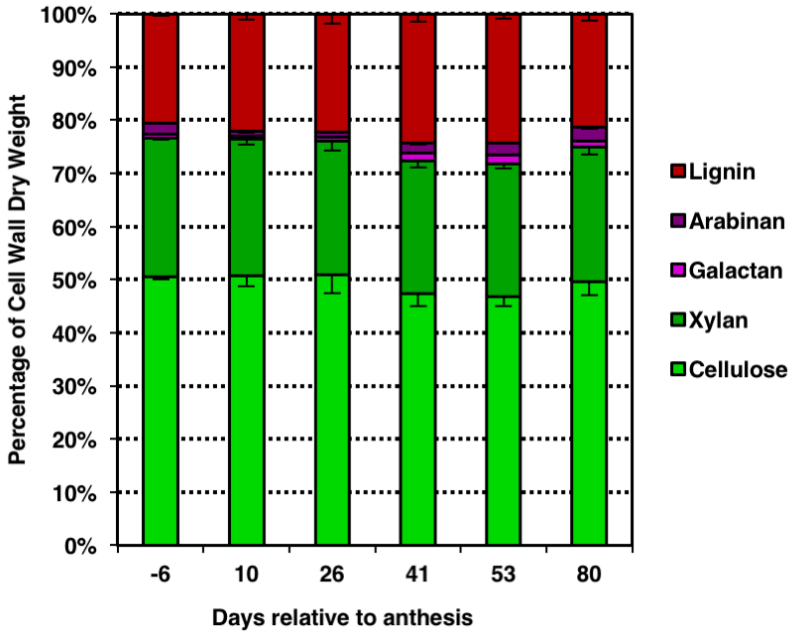

Figure 2.4. Lignocellulosic composition of Della stems. Stems of field grown Della sweet sorghum from College Station. Internodes were collected starting at 63 DAE and ending at 149 DAE. The data represent cellulose, xylan, lignin, arabinan and galactan composition as predicted by NIR and are the mean of nine biological replicates. The colored bars represent the relative composition of each fraction as a percent of total structural carbohydrate.

The NSC (sugars/starch) fraction of stem biomass increased from 20% six days before anthesis, to 30% at anthesis and to ~50% by grain maturity (A+30-35d) in field grown plants (Figure 2.3 and 2.5). Glucose and fructose were the most abundant sugars before anthesis, accounting for ~15% of the stem's dry weight (Figure 2.5). As a percentage of stem dry weight, monosaccharide levels peaked at anthesis then declined gradually until 53 days post anthesis (Figure 2.5). Sucrose levels increased from ~3% of stem biomass 6 days before anthesis to ~26% shortly after anthesis (A+10d) and reached a maximum of ~31% 26 days post anthesis. Sucrose levels remained elevated until after grain maturity, but then decreased between A+53d and A+80d. The decrease was

correlated with continued apical tiller outgrowth that started at grain maturity and accumulation of damage to leaves due to age. Low levels of starch were present in stems at anthesis and began increasing just before grain maturity reaching maximal levels (~10%) shortly after grain maturity (A+53d) (Figure 2.5).

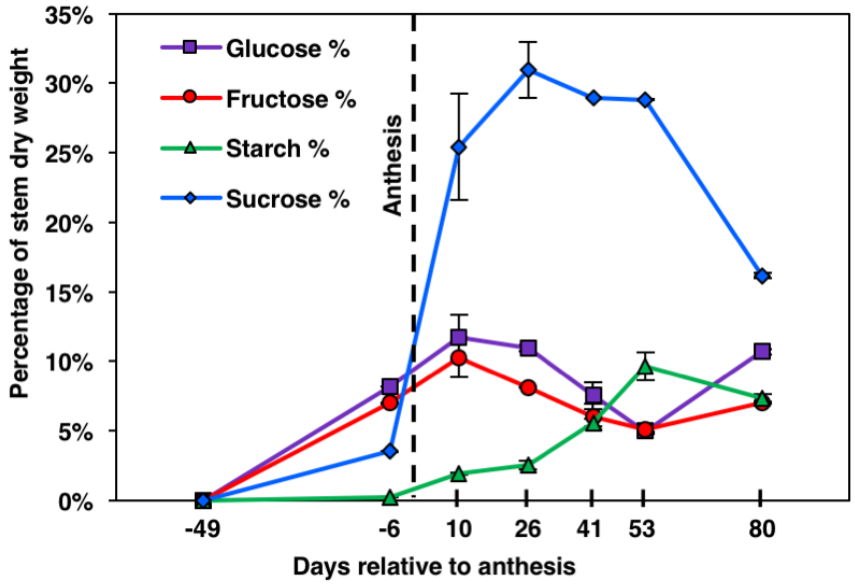

Figure 2.5. Nonstructural carbohydrate accumulation in the internode of Della during development in the field in College Station, 2012. Carbohydrate fractions were quantified with HPAE-PAD and plotted as a percent of total stem dry weight at each time point. The data represent the mean ±SE from two technical replicates of bulk samples consisting of nine biological replicates.

The ratio of LC-biomass to NSC biomass in Della sorghum stems was calculated at various stages of development (Figure 2.6). The stem's ratio of these constituents was ~0.5 at 6 days before anthesis and increased to ~2.8 after anthesis when NSC levels reached a maximum of ~52% of the stem dry

weight (Figure 2.6, red line). When data from stems plus panicles was combined, this ratio approached ~3.3 at grain maturity (Figure 2.6, blue line).

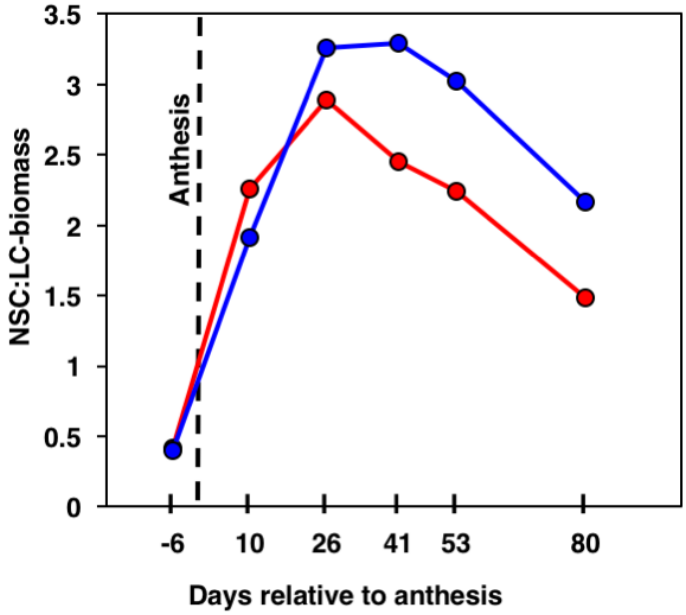

Figure 2.6. Ratio of stem non-structural carbohydrates (NSC) to LC-biomass during development of Della. The red line series represents the ratio of NSC:LC-biomass dry weights for the stem and the blue line series represents the ratio for the sum of the stem and panicle.

Stem internode transcriptome analysis

The next goal of this study was to obtain transcriptome profiles from stems of Della plants spanning ~100 days of development from floral initiation through 30 days post-grain maturity in order to identify genes involved in stem growth, cell wall synthesis, and sucrose accumulation. To minimize the influence of environmental variation on transcription profiles, Della plants were grown in a

controlled environment greenhouse. Della plants grown in the greenhouse traversed through the same stages of development and were of similar morphology and composition as field grown plants although in the greenhouse the vegetative phase was ~2 weeks longer. Della plants grown in the greenhouse reached anthesis at ~90 DAE, ~10-15 days later than field grown plants, most likely because in the greenhouse plants were exposed to constant 14h day lengths whereas field grown plants emerged in ~12.5h day lengths that increased during the season to 14h after floral initiation. Della exhibits a low level of photoperiod sensitivity and the constant long days in the greenhouse most likely delayed the onset of floral transition of these plants relative to field grown plants. While floral initiation occurred ~2 weeks later in the greenhouse, the time from floral initiation to anthesis and from anthesis to grain maturity were similar in the field and greenhouse. Therefore, transcriptome profiles, biomass composition and field data are presented on a time scale relative to anthesis to better align changes in composition and gene expression profiles. The first time point for transcriptome analysis was harvested from plants approximately at floral initiation, 29 days before anthesis (A-29d).

Sugar levels increased in stems after floral initiation in the field and greenhouse although the rate of sugar accumulation in stems was slower in the greenhouse (Figure 2.7). The slower rate of sucrose accumulation in the greenhouse was most likely due to lower light intensity (~1,200 μmol/m^2/sec at mid-day) compared to the field (~2,000 μmol/m^2/sec PAR at mid day). Soluble

sugars accumulated to high concentrations in stems in both experiments with maximum Brix values of 18.2°Bx (greenhouse) and 14.4°Bx (field). Analysis of stem sugar composition of greenhouse grown plants showed that glucose, fructose and sucrose levels were low prior to anthesis and higher in stems at anthesis (Figure 2.8), similar to field grown plants. Stems of greenhouse grown plants had higher Brix values after grain maturity compared to field grown plants most likely because apical tillers initiated after flowering were removed from these plants.

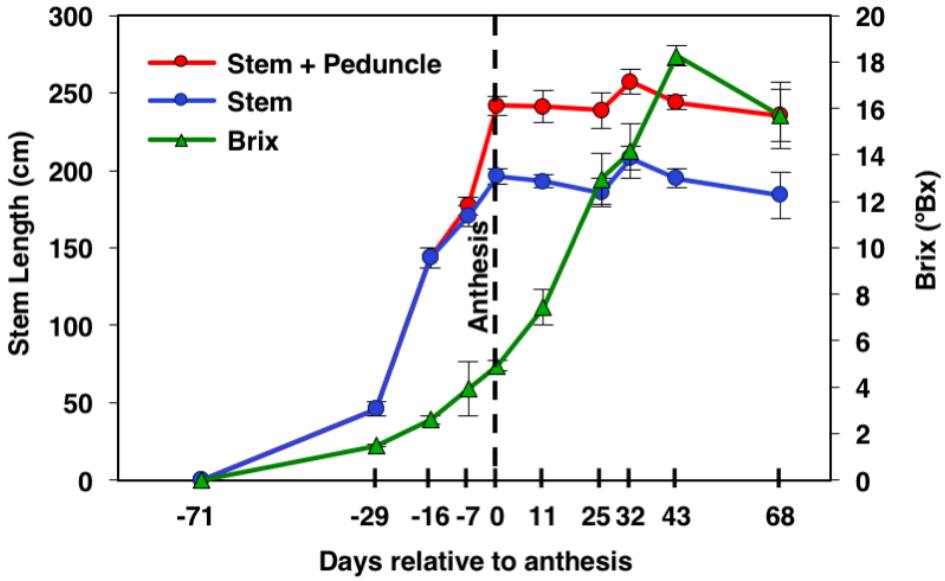

Figure 2.7. Time course of changes in stem length and accumulation of soluble solids during Della development in a greenhouse. Brix (green line), length of the stem (blue line) and the length of the stem plus the length of the peduncle (red line). Development is expressed in days relative to anthesis (0). Each data point represents the mean ±SE of 4 biological replicates.

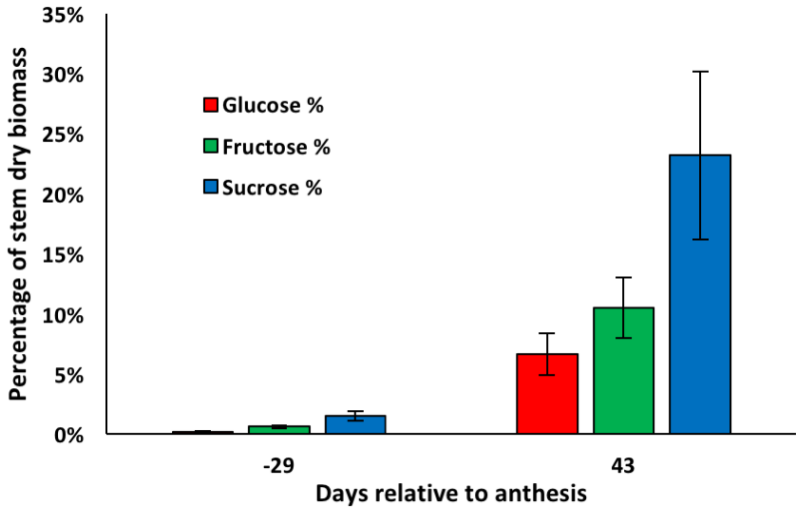

Figure 2.8. Sugar composition of Della stems in the greenhouse during development. Stems were collected from Della 29 days before anthesis (-29) or 43 days after anthesis and glucose, fructose and sucrose composition analyzed by HPLC. Each data point represents the mean of 3 biological replicates.

Sorghum stem internodes are produced sequentially by the shoot apical meristem (SAM) during the vegetative phase. Newly formed internodes increase slowly in size, then elongate due to the action of a meristem located at the base of the internode and an adjacent zone of cell elongation until final internode length is achieved (Knoller *et al.* 2010, Shen *et al.* 2013). Consequently, stems contain a gradient of internodes of varying development, with several newly formed, non-elongated internodes located just below the shoot apical meristem at floral initiation. New internode production stops upon floral initiation when the SAM is reprogrammed to produce a peduncle and panicle. From floral initiation to anthesis, internodes sequentially expand to full length until all internodes present at floral initiation have reached full length. To avoid pooling internodes at different stages of growth and development, tissue and RNA for transcriptome

profiling was collected from the tenth internode (Int10) of stems at eight time points during development. The tenth internode was located by numbering leaves sequentially based on their order of production and expansion. At the first time point (A-29d), Int10 was ~70% fully elongated. By A-16d, Int10 had reached full length while internodes closer to the shoot apical meristem were still elongating. Three biological replicates of Int10 RNA were collected and analyzed using RNA-seq. The first time point of stem tissue collection occurred 29 days before anthesis (A-29d) and samples were subsequently collected at A-16d (flag leaf stage), A-7d, and at anthesis (A). Stem Int10 samples were also isolated from plants at four time points post-anthesis including A+11d, A+25d (grain soft dough stage), A+43d, (~14 days post grain maturity) and A+68. RNA obtained from stem internodes analyzed by RNA-seq produced 5.8M reads on average per library, and ~17.3M reads per sample, that could be aligned to the reference sorghum genome. Principal component analysis of RNA-seq data showed that biological replicates clustered together (Figure 2.9). A total of 18,463 genes were expressed with an RPKM>2 and used in the analysis. Of these 6,802 genes were differentially expressed >3-fold (FDR ≤ 0.05) in stems between any two time points during the developmental time course.

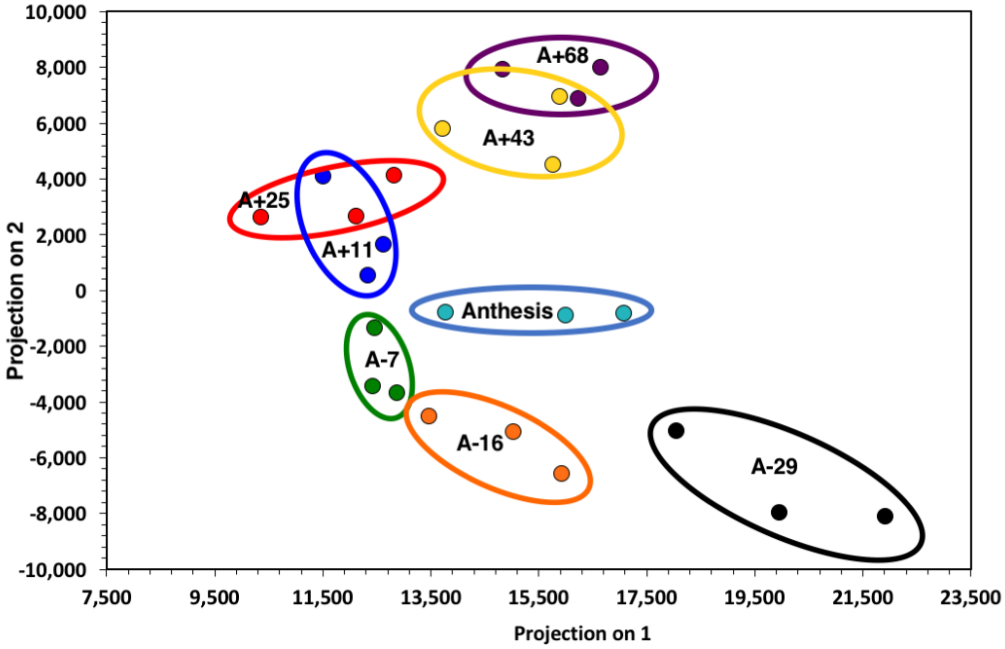

Figure 2.9. Principal components analysis of RNA-seq data derived from Della internodes harvested along the time course of development. Each data point represents a unique RNA-seq library. Libraries from the same time point have the same color. Circles delineate boundaries between clusters from different time points. The label in the center of each cluster represents the time of harvest (in days) relative to anthesis.

The analysis of differentially expressed genes in stems focused on pathways involved in growth, cell wall biology and stem sugar accumulation. Data on genes with relatively high expression that were differentially expressed during the developmental time course is summarized in Tables 2.1-2.4 and Figure 2.10. The genes were further grouped according to biochemical pathway and gene family. Genes in target pathways were identified in part through annotations provided by Phytozome. Annotations were further refined based on a prior analysis of genes involved in grass cell wall biology (Penning *et al.* 2009) and studies of specific groups of genes of interest (i.e., Mitchell clade) (Mitchell *et al.* 2007). The study by Penning et al (2009) identified a large set of sorghum

and maize genes involved in cell wall biology that were used in this study to identify sorghum genes involved in cell wall biology that were differentially expressed during Della stem and plant development. We anticipate that further discovery, gene validation experiments, and annotation refinement will occur over time.

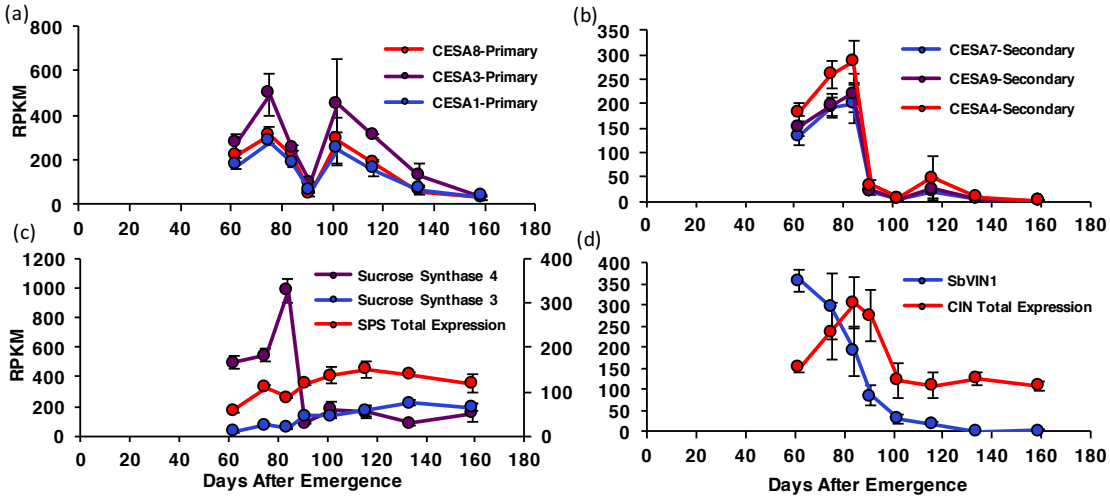

Figure 2.10. Relative expression of genes in the internode that modify sucrose accumulation in greenhouse grown Della during development.
(a) Expression (RPKM) of genes encoding cellulose synthases involved in primary cell wall biosynthesis (*CESA1, CESA3, CESA8*).
(b) Expression of genes encoding cellulose synthases involved in secondary cell wall biosynthesis (*CESA4, CESA7, CESA9*).
(c) Expression of *SUS4* and *SUS3*, genes encoding sucrose synthases and the sum of the expression of genes encoding sucrose phosphate synthase.
(d) Expression of *SbVIN1*, a gene encoding a vacuolar invertase and the sum of expression of four genes encoding cytosolic invertases (CIN). Data represent the mean ±SE of three biological replicates.

Genes involved in cell growth, cell wall proteins

Expansins facilitate cell growth by increasing cell wall extensibility (McQueen-Mason and Cosgrove 1995). Stem internode number 10 (Int10) was 70% of full

length at the first time point of the transcriptome analysis (A-29d) and had reached full length by the second time point (A-16d). Genes homologous to expansins were differentially expressed in Int10 at A-29d with significantly lower expression at later stages of development following cessation of elongation (Table 1, Table A.1). The maximum expression of individual expansin gene family members varied >10-fold possibly reflecting stem cell specific expression.

Table 2.1. Expression of sorghum genes encoding expansins and cell wall proteins.

Family/ID	Function	A-29	A-16	A-7	A	A+11	A+25	A+43	A+68	FC
Expansins										
Sobic.009G173700	α-expansin	145	8	8	1	9	11	0	1	>145
Sobic.006G191700	α-expansin	39	6	16	5	16	14	10	7	8
Sobic.001G033300	α-expansin	20	1	4	1	1	2	1	3	>20
Sobic.001G300400	β-expansin	180	10	10	2	8	21	1	1	233
Sobic.001G300800	β-expansin	168	6	10	0	5	4	0	1	>168
Sobic.001G301600	β-expansin	53	4	10	1	4	2	0	0	>53
Arabinogalactan-proteins (AGPs)										
Sobic.003G062000	FASCICLIN-like arabinogalactan	346	140	308	17	6	59	5	2	190
Sobic.003G250700	FASCICLIN-like arabinogalactan	280	28	41	21	11	6	3	2	129
Sobic.009G055900	FASCICLIN-like arabinogalactan	186	44	34	18	77	50	22	14	15
Proline-rich repeat Extensins										
Sobic.007G096700	POE130, extensin	182	84	89	54	87	101	95	83	4
Sobic.K008000	POE124, extensin	85	36	42	9	47	31	16	7	13

Sorghum stem internode samples were collected from greenhouse grown plants 29 days before anthesis (A-29) through 68 days post-anthesis (A+68). Expression values are presented as RPKM and each data point represents the mean of three biological replicates. Higher expression is represented by darker shades of blue. FC = fold change. FDR adjusted p-values were calculated for the maximum fold change in expression for each gene in the experiment.

Genes encoding several cell wall proteins were also differentially expressed in stem Int10 at A-29d. For example, six genes annotated as encoding fasciclin-like arabinogalactans were expressed at high levels in elongating internodes (A-29d) then were down-regulated to varying extents and kinetics at later stages of plant development (Table 1, Table A.2). Genes encoding proline rich extensins were differentially expressed during development in a similar manner (Table 1, Table A.2). Numerous genes

encoding xyloglucan endotransglucosylase/hydrolases were expressed in stems, and all showed their highest expression in Int10 at A-29d consistent with a role in cell wall modification during internode growth (Table A.7).

Cellulose synthases, cellulose-like synthases, COBRA, callose synthase

CESA genes that encode cellulose synthases were highly expressed in stem Int10 until 7 days before anthesis followed by a 4-10-fold decrease in RNA level by anthesis (Table 2.2, Table A.3). *CESA1, CESA3,* and *CESA8* are homologous to rice *CES* genes involved in primary cell wall biosynthesis (Tanaka *et al.* 2003) serving the same role as Arabidopsis *CESA1*, *CESA3* and *CESA6* (Persson *et al.* 2007). The expression of these genes peaked at A-16d, declined by anthesis, and then showed a second peak of expression post-anthesis followed by another gradual decline in expression (Table 2.2, Figure 2.10a). *CESA5* and *CESA6* were expressed at high levels in elongating internodes followed by a ~5-fold and ~26-fold decrease in RNA abundance during subsequent development (Table 2.2). *CESA4, CESA7* and *CESA9* are homologous to genes in rice involved in secondary cell wall biogenesis (Tanaka*, et al.* 2003) and Arabidopsis *CESA4*, *CESA7* and *CESA8* (Persson*, et al.* 2007). These genes were expressed at elevated levels in internodes until 7 days before anthesis and then ~5-10-fold lower at anthesis, and ~100-fold lower after grain maturity (Figure 2.10b). COBRA genes encode GPI anchored proteins that maintain anisotropic cell wall expansion (Roudier *et al.* 2005). Genes encoding COBRA proteins showed elevated expression prior to anthesis when *CESA*

genes were highly expressed followed by gradually decreasing expression post-anthesis (Table 2.2).

Table 2.2. Expression of genes encoding cellulose synthases, cellulose-like synthases and COBRA in sorghum stems.

Family/ID	Clade/Gene	Function	A-29	A-16	A-7	A	A+11	A+25	A+43	A+68	FC
Cellulose synthases											
Sobic.009G063400	*OsCESA1*	cellulose synthase	172	280	186	61	248	157	69	35	9
Sobic.002G118700	*OsCESA3*	cellulose synthase	277	494	253	95	450	310	129	30	17
Sobic.002G075500	*OsCESA8*	cellulose synthase	218	310	223	45	288	187	60	29	11
Sobic.001G021500	*OsCESA5*	cellulose synthase	191	91	94	79	103	99	92	50	4
Sobic.002G094600	*OsCESA6*	cellulose synthase	154	120	146	76	26	45	16	6	26
Sobic.003G296400	*OsCESA4*	cellulose synthase	183	260	287	33	7	48	8	1	229
Sobic.001G224300	*OsCESA7*	cellulose synthase	135	193	200	20	5	20	4	1	167
Sobic.002G205500	*OsCESA9*	cellulose synthase	154	198	221	22	5	25	4	1	203
COBRA											
Sobic.001G086200	COBRA	COBRA-like protein	442	479	494	340	436	321	224	118	5
Sobic.001G336700	COBRA	COBRA-like protein	47	75	90	13	9	26	6	3	36
Cellulose-like synthases											
Sobic.004G075900	*CSLA1*	mannan, glucomannan synthase	28	6	3	0	5	7	1	0	>28
Sobic.002G385800	*CSLA7*	mannan, glucomannan synthase	22	5	5	5	3	7	4	4	8
Sobic.007G137400	*CSLA11*	mannan, glucomannan synthase	24	123	95	106	76	38	44	36	5
Sobic.010G008600	*CSLD2*	mannan synthase	65	41	84	72	27	28	36	30	4
Sobic.007G050600	*CSLF6*	beta1,3;1,4 glucan synthase	215	188	318	241	164	146	121	62	6

Sorghum stem internode samples were collected from greenhouse grown plants 29 days before anthesis (A-29) through 68 days post-anthesis (A+68). Expression values are presented as RPKM and each data point represents the mean of three biological replicates. Higher expression is represented by darker shades of blue. FC = fold change. FDR adjusted p-values were calculated for the maximum fold change in expression for each gene in the experiment.

Nine genes encoding cellulose-like synthases (*CSL*-genes) were expressed in sorghum stems at variable and generally lower levels compared to *CESA* genes (Table 2.2, Table A.3). Several *CSLA* and *CSLD* genes are involved in mannan or glucomannan synthesis (Goubet *et al.* 2009, Li *et al.* 2009, Liepman and Cavalier 2012, Verhertbruggen *et al.* 2011). Four of the *CSLA*-genes showed elevated expressed in elongating stem internodes 29 days

before anthesis and at lower levels later in development (*CSLA1, CSLA4, CSLA6, CSLA7*). *CSLA11* and *CSLD2* expression increased from A-29d to A-7d before declining after anthesis (Table 2.2). A *CSLF* gene that encodes a (1,3;1,4) - *β*-D-glucan synthase (Burton and Fincher 2012, Ermawar *et al.* 2015) was expressed throughout the developmental time course. *CSLE6* and *CSLG2* were expressed at very low levels with a small increase in RNA level post grain maturity (Table A.3).

Three genes encoding callose synthases ((1,3)-*β*-D-glucan synthases) were expressed in stems with ~3-fold variation during the developmental time course (Table A.4). In addition, a large number of genes in glycoside hydrolase family 17 that are involved in callose turnover exhibited differentially elevated expression at A-29d, while other genes in this family showed peak expression at or just before anthesis, or post-grain maturity (Table A.4).

Genes involved in glucuronoarabinoxylan (GAX) synthesis

Grass cell walls contain significant amounts of glucuronoarabinoxylans (GAX) (Figure 2.3; Figure 2.4) (Carpita 1996, Penning*, et al.* 2009). Genes encoding enzymes involved in GAX synthesis and modification are found in the GT8, GT31, GT34, GT37, GT43, GT47 and GT61 gene families and the Mitchell clade (Mitchell*, et al.* 2007, Penning*, et al.* 2009). Numerous genes from these GT-families were differentially expressed in sorghum stem Int10 during development with a wide range of expression levels and patterns of expression (Table 2.3, Table A.5). Sorghum genes homologous to *IRX9*, *IRX10/IRX10L* (6 family

members), and *IRX14*, genes involved in xylan backbone biosynthesis (Rennie and Scheller 2014, Zhang *et al.* 2014) were differentially expressed in stems during development from A-29d to anthesis coincident with high rates of cell wall biosynthesis (Table 2.3). A sorghum homolog of *XAT1*, a gene encoding xylan arabinosyl transferase (Anders *et al.* 2012), was expressed in sorghum stems at relatively high levels throughout development (Table 2.3). Two sorghum homologs of *GUX1/2/4*, genes encoding glucuronoxylan transferases (Bromley *et al.* 2013, Rennie *et al.* 2012), were expressed in stems at different levels and patterns during development (Table 2.3). Sorghum homologs of *FUT1*, *XXT1/2* and *XAX1* (Chiniquy *et al.* 2012), genes encoding fucosyltransferase, xylosyltransferase, and an enzyme that mediates xylosyl substitution of arabinofurosylxylan, were also tentatively identified (Table 2.3). *TBL29 (AXY4)* genes encoding DUF231 family proteins are involved in o-acetylation of xylan (Gille *et al.* 2011) and *RWA*-genes encode putative acetyl-coA transporters required for acetylation of xylan (Lee *et al.* 2011). Three sorghum homologs of *TBL29* and two homologs of *RWA*-genes showed elevated expression during stem cell wall biosynthesis from A-29d to anthesis. Sorghum homologs of *UXT1/2/3*, genes encoding UDP-xylose transporters (Ebert *et al.* 2015), were differentially expressed in stems prior to anthesis (Table 2.3, Table A.5).

Table 2.3. Expression of sorghum genes involved in GAX biosynthesis in stems.

Family/ID	Clade/Gene	Function	A-29	A-16	A-7	A	A+11	A+25	A+43	A+68	FC
GT43											
Sobic.001G409100	*IRX9*	Xylan synthesis	66	42	110	47	19	25	17	4	28
Sobic.010G238800	*IRX14*	Xylan synthesis	318	366	428	305	145	178	141	129	4
GT47											
Sobic.003G410700	*IRX10L*	Exostosin, xylan synthesis	189	107	72	64	160	133	137	93	3
Sobic.003G410800	*IRX10L*	Exostosin, xylan synthesis	126	166	251	97	48	33	18	8	32
GT37											
Sobic.004G308400	*FUT1*	xyloglucan fucosyltransferase	33	32	68	73	16	30	39	18	6
GT34											
Sobic.001G401600	*XXT1/2*	UDP-xylosyltransferase	60	9	16	9	14	9	4	2	33
GT31											
Sobic.004G051300		galactosyltransferase	74	85	81	49	68	53	35	17	5
Sobic.002G068800		galactosyltransferase	40	31	83	104	11	22	32	9	11
GT61											
Sobic.003G376700	*XAT1*	Xylan arabinosyl transferase	433	408	448	498	241	225	207	124	4
Sobic.004G007600	*XAX1*	UDP-xylosyl substitution of arabinoxylan	123	117	146	59	42	48	36	32	5
GT8											
Sobic.001G479800	*GUX1/2/4*	Glucuronosyl substitution of Xylan	86	93	58	31	76	70	23	12	8
Sobic.009G144200	*GUX1/2/4*	Glucuronosyl substitution of Xylan	22	8	5	3	3	3	4	3	8
DUF231 family											
Sobic.001G406700	*TBL29*	o-acetylation of xylan	33	57	71	6	1	3	1	0	>71
Mitchell Clade 1											
Sobic.003G043600		HXXXD-type acyl-transferase	325	85	206	131	40	14	17	13	29
Sobic.009G034300		HXXXD-type acyl-transferase	94	194	232	104	81	103	25	43	11
Sobic.010G180100	*OsAT10*	p-coumaroyl CoA transferase	17	14	27	4	10	9	4	4	8
Transporters											
Sobic.006G149400	*UXT1*	UDP-Xylose transporter into the golgi	41	99	108	9	3	19	4	1	>108
Sobic.004G221300	*UXT1*	UDP-Xylose transporter into the golgi	155	130	150	37	25	21	16	12	14
Sobic.009G246700	*RWA1-4*	acetyl-CoA transporter	72	91	101	23	44	33	20	15	7

Sorghum stem internode samples were collected from greenhouse grown plants 29 days before anthesis (A-29) through 68 days post-anthesis (A+68). Expression values are presented as RPKM and each data point represents the mean of three biological replicates. Higher expression is represented by darker shades of blue. FC = fold change. FDR adjusted p-values were calculated for the maximum fold change in expression for each gene in the experiment.

A subset of the genes in Mitchell Clade 1 encode HXXXD type acyltransferases that are involved in modification of arabinoxylans through addition of ferulic acid or *para*-coumaric acid as well as other functions (Bartley *et al.* 2013, Mitchell*, et al.* 2007). Four genes in Mitchell Clade 1 were expressed in sorghum stems at various levels and patterns of expression during development (Table A.5). The sorghum homolog of *OsAT10*, a gene encoding a p-coumaroyl coenzyme A transferase involved in GAX modification in rice (Bartley*, et al.* 2013, Mitchell*, et al.* 2007) showed maximal expression seven days before anthesis, and lower expression post anthesis (Table 2.3).

Penning et al. (2009) identified a large number of genes that are involved in generating substrates for cell wall biosynthesis. Sorghum genes involved in nucleotide-sugar interconversion pathways were identified and in general were differentially expressed in stems from A-29d to A+11d (Table A.6).

Genes encoding glycosyl hydrolases, pectin lyases/esterases

Xyloglucan endotransglucosylase/hydrolases (XTHs) are active during cell growth and mediate cell wall remodeling (Park and Cosgrove 2015). Eleven members of the *XTH*-gene family were differentially expressed in sorghum stems, all with elevated expression at A-29d (Table A.7). Four family members showed a second peak of expression between A-7d and anthesis that could contribute to secondary cell wall biogenesis (Table A.7).

Genes encoding mixed linkage glucanases, B-galactosidases, and cellulases were differentially expressed in stem Int10 during development (Table A.7). The gene encoding a mixed linkage glucanase with the highest expression showed ~3-fold higher expression post-anthesis. Genes encoding endo-(1,4)-*β*-D-glucanses in general were expressed at higher levels between A-29d and A+11d. Genes encoding pectin lyases such as polygalacturonase, were expressed at lower levels compared to *CESA* genes, and with varied patterns of gene expression (Table A.8).

Genes involved in lignin biosynthesis

Lignin accounted for ~21-25% of Della stem cell wall biomass (Figure 2.4). Lignin biosynthesis is complex and developmentally regulated (Vanholme *et al.*

2010) More than 30 genes involved in lignin biosynthesis were differentially expressed in sorghum stems during development (Table A.9). Multi-gene families encoding *4CL* (5-members), *HCT* (7-members) and *CCR* (4-members) were expressed at varying levels and patterns of expression during development (Table A.9). Two sorghum genes were identified with sequence similarity to *OsPMT*, a rice enzyme that acylates monolignols with *p*-coumarate (Withers *et al.* 2012). Genes encoding enzymes for each step of lignin biosynthesis with the highest relative expression showed peak expression in stem internode 10 between A-16d to A-11d (Figure 2.11, Table 2.4) coincident with expression of *CESA* genes involved in secondary cell wall biosynthesis. Another group of genes involved in lignin biosynthesis were expressed at low levels prior to anthesis but at somewhat higher levels post-anthesis (Table A.9).

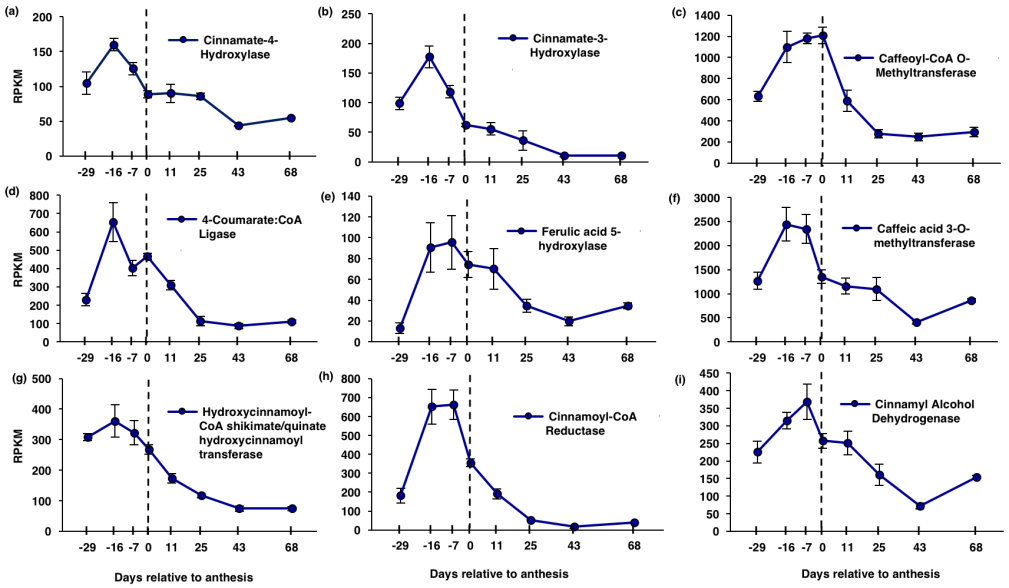

Figure 2.11. Expression (RPKM) of select family members from the phenylpropanoid pathway in stems during Della development. (a) cinnamate-4-hydroxylase, (b) cinnamate-3-hydroxylase, (c) caffeoyl-CoA O-methlytranferase, (d) 4-courmarate CoA ligase. (e) ferulic acid 5-hydroxylase, (f) caffeic acid 3-O-methyltranferase, (g) hydroxycinnamoyl-CoA shikimate/quinate hydroxycinnamoyltransferase, (h) cinnamoyl-CoA-reductase. (i) cinnamyl alcohol dehydrogenase. Each data point represents the mean and ±SE for three biological replicates.

Table 2.4. Expression of genes involved in lignin biosynthesis in stems.

Family/ID	Clade/Gene	A-30	A-16	A-7	A	A+11	A+25	A+43	A+68	FC
Phenylalanine ammonia-lyase										
Sobic.004G220300	*PAL*	4274	4883	3595	3239	1261	1267	160	62	34
Sobic.006G148800	*PAL*	317	1270	658	464	173	184	26	7	54
Cinnamate-4-hydroxylase										
Sobic.002G126600	*C4H*	105	160	125	89	90	86	44	55	5
Cinnamate-3-hydroxylase										
Sobic.009G181800	*C3H*	99	177	118	63	56	36	11	6	19
Sobic.003G327800	*C3H*	100	68	119	67	49	46	30	33	5
Caffeic acid 3-O-methyltransferase										
Sobic.007G047300	*COMT*	1270	2444	2346	1352	1157	1095	405	285	7
Ferulic acid 5-hydroxylase										
Sobic.001G196300	*F5H*	13	90	96	74	70	35	20	22	7
Cinnamoyl CoA reductase										
Sobic.007G141200	*CCR*	181	652	661	355	191	51	17	5	46
Sobic.004G065600	*CCR*	12	38	50	66	14	15	17	10	7
4-coumarate:CoA ligase										
Sobic.004G062500	*4CL*	230	653	403	467	309	113	86	85	8
Sobic.010G214900	*4CL*	79	58	70	59	56	45	41	41	3
Caffeoyl-CoA O-methyltransferase										
Sobic.010G052200	*CCoAOMT*	632	1099	1181	1207	591	276	247	187	6
Sobic.007G218800	*CCoAOMT*	442	643	494	138	88	85	61	66	12
Hydroxycinnamoyl-CoA transferase										
Sobic.004G212300	*HCT*	309	361	323	268	173	117	74	57	5
Sobic.006G136800	*HCT*	143	149	200	150	71	29	23	12	15
Cinnamyl alcohol dehydrogenase										
Sobic.006G014700	*CAD*	3	23	33	59	15	27	12	16	21
Sobic.010G071900	*CAD*	5	7	10	10	14	32	19	17	5
p-coumaroyl-CoA:monolignol transferase										
Sobic.009G034300	*PMT*	94	194	232	104	81	103	25	43	11
Sobic.010G180100	*PMT*	17	14	27	4	10	9	4	4	5

Sorghum stem internode samples were collected from greenhouse grown plants 29 days before anthesis (A-29) through 68 days post-anthesis (A+68). Expression values are presented as RPKM and each data point represents the mean of three biological replicates. Higher expression is represented by darker shades of blue. FC = fold change. FDR adjusted p-values were calculated for the maximum fold change in expression for each gene in the experiment.

Monolignol polymerization is mediated in the cell wall by radical coupling. Laccases and peroxidases localized in cell walls are a source of radicals used in this process (Berthet *et al.* 2011, Penning*, et al.* 2009). Seven genes encoding laccases and 25 genes encoding peroxidases were differentially expressed in sorghum stems (Table A.10). Four genes encoding laccases and numerous peroxidases were differentially expressed in elongating internodes and strongly

down regulated at anthesis. A second group of genes encoding laccases and peroxidases were expressed at relatively high levels until seven days before anthesis and then at lower levels during later stages of development.

Genes involved in sucrose metabolism

Stem sucrose levels were low during the vegetative phase when rapid stem growth was occurring (A-29d). Sucrose began accumulating after floral initiation and reached maximal levels by grain maturity in greenhouse grown plants (Figure 2.7). In the stem, sucrose levels can also be affected by varying rates of hydrolysis by invertases (INV) and sucrose synthases and synthesis by sucrose phosphate synthase (SPS) (Table A.11). Sucrose synthase (SUS) hydrolyzes sucrose producing UDP-glucose (and fructose) a substrate for cell wall biosynthesis. Expression of *SUS4*, the most highly expressed gene encoding SUS in stems, was very high from A-29d to A-7d, then down regulated ~10-fold at anthesis and thereafter in parallel with secondary cell wall biosynthesis (Figure 2.10c). *SbSUS3* in contrast was expressed in stems at lower levels compared to *SUS4* and this gene showed steadily increasing levels of expression during development (Figure 2.10c, Table A.11). A gene encoding sucrose phosphate synthase (SPS) was expressed at low and gradually increasing levels post anthesis (Figure 2.10c, Table A.11).

Lingle (1987) reported high invertase activity in sorghum stems during the vegetative phase and lower activity post-anthesis when high levels of sucrose accumulate. Four *CIN*-genes encoding cytoplasmic invertases and one *VIN*-

gene encoding a vacuolar invertase were differentially expressed in Della stems during development (Figure 2.10d; Table A.11). The stem *CIN*-genes showed elevated expression until anthesis and lower expression thereafter. The most highly expressed *INV*-gene encoded a vacuolar invertase (Sobic.004G004800; *SbVIN1*). This gene was expressed at high levels from A-29d to A-7d followed by >100-fold down regulation at later stages of stem and plant development (Figure 2.10d; A.11). The time course of down regulation of *SbVIN1* expression was inversely correlated with sucrose accumulation (Pearson's r = -0.83).

DISCUSSION

Energy crops can be improved by increasing biomass yield, enhancing resilience to adverse environmental conditions, reducing input requirements, and by optimizing biomass composition. Energy sorghum hybrids developed over the past 10 years can produce high biomass yield at low cost (~$40/Mg) (Gill*, et al.* 2014, Mullet*, et al.* 2014, Perlack and Stokes 2011, Rooney*, et al.* 2007). Energy sorghum's drought resilience and water use efficiency allows production of energy sorghum hybrids in drought prone regions that are sub-optimal for food crops (Mullet*, et al.* 2014, Olson*, et al.* 2013). An important next step in energy sorghum hybrid development is to improve biomass composition in order to reduce the cost and improve the efficiency of converting biomass to biofuels and bio-products. Sorghum germplasm exhibits extensive variation in biomass composition that is potentially useful for energy sorghum breeding (Stefaniak*, et al.* 2012). A more complete understanding of the molecular and genetic basis of

composition variation in sorghum germplasm would accelerate utilization of this diversity for energy sorghum breeding. Additional opportunities for engineering LC-biomass with improved conversion properties have been identified in other systems that could potentially be deployed in energy sorghum if the target genes and pathways in sorghum could be identified (McCann and Carpita 2015). Therefore one overall goal of this study was to obtain information on the regulation of processes, pathways, and genes that affect biomass partitioning, sucrose accumulation, and cell wall biology useful for optimizing energy sorghum biomass composition.

Figure 2.12 provides a diagram summarizing changes in plant growth and biomass partitioning that occurs during development to variation in stem growth, cell wall synthesis and pathways identified by profiling stem gene expression. The time course of biomass accumulation in leaves, stems, panicles and tillers of Della was characterized under field conditions from germination through 30 days post-grain maturity. Della accumulated biomass primarily in leaves, leaf sheaths and roots during the first ~50 days of vegetative growth. Following floral initiation leaf area and biomass continued to increase as leaves initiated prior to anthesis expanded, however biomass preferentially accumulated in stems during a phase of rapid stem elongation. Stem growth slowed starting 7-16 days prior to anthesis as the last several internodes produced before floral initiation completed elongation. Following anthesis, biomass accumulated in stems in the form of sucrose and preferentially in panicles in the form of grain starch until

grain maturity. Biomass accumulation slowed post-grain maturity and was reallocated to support apical tiller outgrowth (Figure 2.12b, right). The reduction in rate of biomass accumulation post grain maturity could be due to lack of available sinks, futile cycling, or reduced photosynthetic activity.

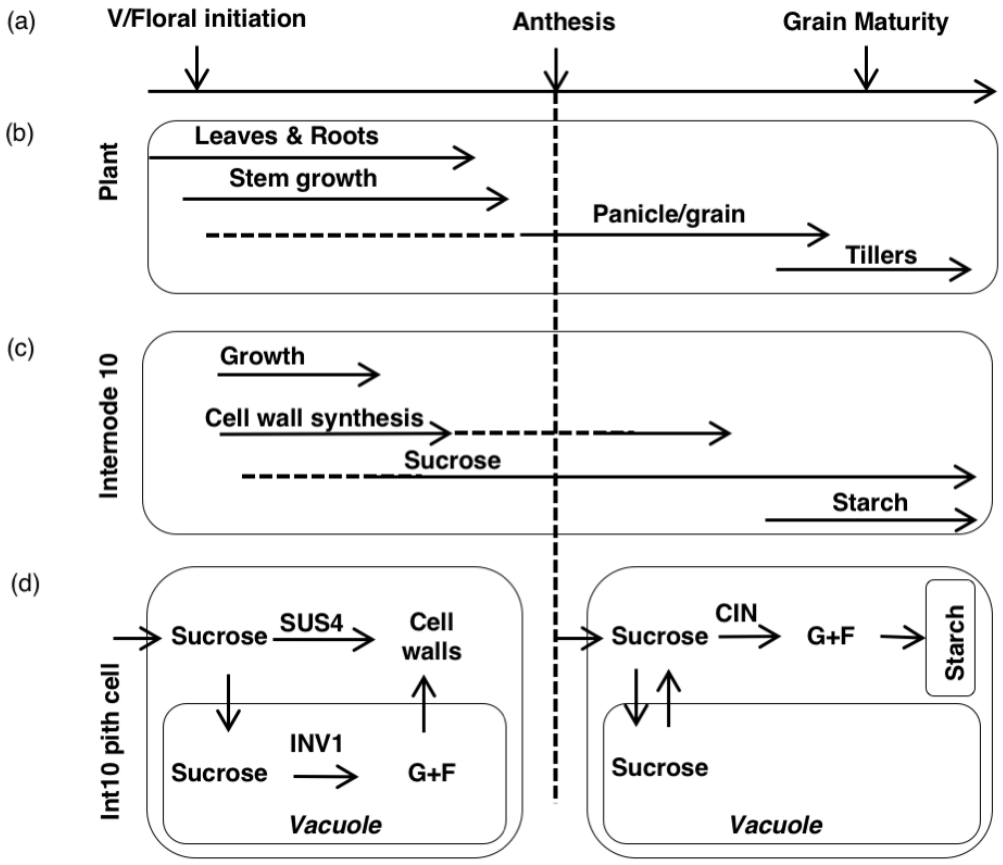

Figure 2.12. Diagram summarizing changes in biomass accumulation, composition, and gene expression during development of the sweet sorghum Della.
(a) A generalized timeline of sorghum development.
(b) The timing of biomass allocation to the leaves and roots, stem, panicle/grain and tillers. The dotted line for panicle/grain represents panicle development pre-grain filling.
(c) The timing of growth and changes in composition of internode 10 during development. The dotted line in cell wall biosynthesis and sucrose represents periods in which the deposition of these materials is reduced.
(d) Generalized mechanism by which sucrose metabolism genes regulate sucrose accumulation before and after anthesis.

Overall, the sequential partitioning of biomass to different organs comprising the shoot during development is a consequence of the timing and duration of organ growth, and the relative sink strength of leaves, leaf sheaths, stems, panicles, and tillers. The amount of biomass partitioned to different organs can be altered significantly by varying the length of different developmental phases or by changing the rate and extent of biomass accumulation in a given organ/tissue. For example, delaying floral initiation of energy sorghum hybrids for ~150 days increases total shoot biomass yield >2-fold and preferentially increases stem biomass compared to early flowering hybrids (Olson*, et al.* 2012). In some environments, delayed flowering will also reduce grain production. While the length of the vegetative phase can be readily increased in energy sorghum hybrids using alleles of *GHD7* and *PRR37* that increase photoperiod sensitivity (Murphy*, et al.* 2011, Murphy*, et al.* 2014), there is less variation in duration from floral initiation to anthesis and from anthesis to grain maturity (Quinby 1974). The productivity of sorghum following grain maturity is less well characterized, however many non-senescent sorghum genotypes remain photosynthetically active after grain maturity (Duncan *et al.* 1981) and some genotypes including Della can produce apical and/or basal tillers (‘ratooning’) post-anthesis. Increased biomass accumulation and modification of partitioning post grain maturity represent potentially useful ways to improve yield and composition in regions of production with long growing seasons. Taken together, modification of the timing of developmental transitions (juvenile to

adult, floral induction, anthesis, grain maturity) and the duration and rates of organ growth represent potential ways to optimize biomass composition for bioenergy applications.

Energy sorghum composition could also be improved by increasing the ratio of non-structural carbohydrates to LC-biomass in stems. Many energy sorghum genotypes such as Della accumulate high levels of sucrose in stems similar to sugarcane (Jackson *et al.* 1980, Murray *et al.* 2009, Wang*, et al.* 2009). Since sucrose, monosaccharides, and starch are converted to biofuels efficiently (>90%) at relatively low cost (Gnansounou and Dauriat 2010, Wu *et al.* 2010), and stem sucrose is a commercial product (Waclawovsky*, et al.* 2010) and important for grain filling (Slewinski 2012), numerous studies have investigated the molecular and genetic basis of stem sugar accumulation in sorghum and sugarcane (Calvino*, et al.* 2011, Gutjahr*, et al.* 2013a, Lingle 1987, Moore 1995, Tarpley*, et al.* 1994a, Tarpley and Vietor 2007, Waclawovsky*, et al.* 2010). Genetic studies of stem sugar accumulation in sorghum have identified a large number of QTL that modulate this trait (Murray*, et al.* 2008, Ritter *et al.* 2008, Shiringani and Friedt 2011). These results indicate that the regulation of sorghum stem sugar accumulation is complex. Unfortunately, most the genes/alleles underlying QTL that modulate sorghum stem sucrose accumulation are unknown.

The rate and extent of sucrose accumulation in stems could potentially be affected by numerous factors including the photosynthetic activity per plant, the

dynamics of sucrose transport from leaves to stems, and the utilization of sugars for growth, grain filling, respiration, futile cycling, or storage. In the current study, Della stem sugar levels increased following floral initiation and reached high levels (~55% of maximum) in the field by anthesis. In the greenhouse, where Della was grown at lower light intensity, accumulation of stem sugar levels reached ~25-30% of maximum by anthesis, and maximal levels after grain maturity. Leaf area and photosynthetic rates remain high in sweet sorghum under favorable field conditions through grain maturity (Gutjahr*, et al.* 2013a). Therefore accumulation of stem sugars following floral initiation is in part a consequence of cessation of phytomer production. This developmental transition eliminates a major sink for carbohydrate once all of the nascent leaves and internodes formed prior to floral initiation reach full size. Stem internodes initiated prior to floral induction expand sequentially so that the stem grows rapidly for 14-21 days after floral initiation, then more slowly by flag leaf stage, eventually stopping by anthesis (Figure 2.12b). Photosynthate from leaves supports stem growth by providing a source of sugars that help maintain turgor required for cell expansion and for cell wall biosynthesis. High levels of *VIN1* in elongating stem internodes may help maintain turgor pressure by hydrolyzing sucrose to glucose and fructose. Transcriptome analysis showed that genes encoding cellulose synthases, and genes involved in GAX and lignin biosynthesis were expressed at high levels in stems from floral initiation until down regulation began ~7-16 days before anthesis. Cell wall biosynthesis during

internode elongation and subsequent secondary cell wall biosynthesis utilizes sucrose transported to stems for cell wall polymer synthesis. Elevated expression of sucrose synthase (*SUS4*) in stem internodes from floral initiation until 7 days before anthesis could help provide UDP-glucose from sucrose for cell wall biosynthesis (Figure 12d). Interestingly, expression of *CESA* genes involved in primary cell wall biosynthesis peaked 16 days before anthesis, decreased ~10-fold by anthesis, then showed a second peak of expression between anthesis and grain maturity (Figure 12a). In contrast *CESA* genes involved in secondary cell wall biosynthesis (*CESA4/7/9*) and *SUS4* showed a peak of expression 7 days before anthesis and then lower expression by anthesis and at later stages of development. Differential regulation of *CESA* genes involved in primary and secondary cell wall biosynthesis in stems may provide a way to alter LC-biomass composition.

Prior research showed that acid and neutral invertase activity and sucrose synthase activity is high in elongating sorghum internodes and that these activities decrease prior to, or concomitant with accumulation of sucrose in stems (HoffmannThoma*, et al.* 1996, Lingle 1987, Tarpley*, et al.* 1994b). In the current study, four genes encoding cytoplasmic neutral invertases were identified that show elevated expression in stems until anthesis, followed by ~4- to ~20-fold lower expression post-anthesis (Figure 2.10d). In addition, a gene encoding a vacuolar acid invertase (*SbVIN1*) was highly expressed in elongating internodes, then ~5-fold lower at anthesis, and ~100-fold lower post-anthesis

with a time course inversely correlated with accumulation of sucrose (Pearson's r = -0.83) (Figure 2.10d). Decreased expression of *SbVIN1* could be a key factor enabling accumulation of sucrose in the large vacuoles of stem pith parenchyma. This result is consistent with previous analysis of invertase activity (Lingle 1987) and the finding that sucrose is transported from the phloem and stored in mature sorghum internode vacuoles via an apoplastic pathway with minimal cleavage (Tarpley and Vietor 2007). A gene encoding SPS was also induced in stems during development in parallel with sucrose accumulation, however, this gene showed relatively low expression, so the contribution of this gene to overall stem sucrose accumulation needs further study. Similar changes in the activity of these enzymes were documented in sugarcane during the transition to sucrose accumulation (Zhu *et al.* 1997).

Sorghum stem LC-biomass represents a significant fraction of total biomass harvested from energy sorghum hybrids. Recent advances in cell wall engineering indicate good potential for improving the efficiency of LC-deconstruction to sugars for fermentation and recovery of lignin for up-grading, consolidated bio-processing or thermochemical conversion of LC-biomass to products (McCann and Carpita 2015). Therefore, genes involved in the synthesis of sorghum stem LC-biomass were identified to enable LC-composition optimization for various conversion processes through directed engineering. Della's stem cell wall composition varied during development from ~47-50% cellulose, ~22-29% glucuronoarabinoxylan (GAX) and ~21-25% lignin,

an overall composition similar to other grasses (Arundale *et al.* 2015, Pauly and Keegstra 2008). A previous comprehensive study of genes involved in grass cell wall biology identified and annotated ~750 genes in maize and their homologs in sorghum (Penning*, et al.* 2009). Using this database of information, augmented by additional targeted analyses of gene families, we were able to identify ~200 genes involved in cell wall biosynthesis that were differentially expressed in stems during Della development. Genes in multi-gene families encoding enzymes involved in cellulose, GAX, and lignin biosynthesis were expressed in stems at varying levels and patterns during development. The complexity of the multi-gene family expression profiles is consistent with the presence of numerous cell types in sorghum stems (i.e., epidermal, pith parenchyma, fiber cells, xylem, phloem) that vary in abundance and cell wall composition (Hatfield *et al.* 1999).

Genes involved in cellulose, GAX and lignin biosynthesis were expressed at relatively high levels in Int10 from floral initiation until 11-25 days post anthesis with peak expression of most genes in these pathways occurring prior to anthesis. Stem Int10 was ~70% fully elongated when sampled 29 days before anthesis. Stem elongation in sorghum, like other C4 grasses, occurs in internodes produced sequentially by the shoot apical meristem following a period of delocalized cell division. When internodes transition to rapid elongation, an intercalary meristem located at the base of elongating internodes produces cells that elongate as they move away from the meristem, resulting in

more fully expanded cells located towards the apical end of each elongating internode (Knoller*, et al.* 2010, Shen*, et al.* 2013). As expected, genes encoding expansins that are involved in cell elongation showed peak expression in internodes at this stage of development. Numerous other genes involved in cell wall biology also showed peak expression in elongating internodes including members of gene families encoding fasciclin-arabinogalactan proteins, proline-rich extensins, glucomannan synthases (*CSLA1, CSLA7*), xyloglucan endotransglucosylase/hydrolases (XTHs), pectin lyases, and a few laccases and peroxidases. We assume, as occurs in rice (Hirano *et al.* 2013), that these genes contribute to varying extent to stem cell growth, cell wall remodeling, and growth cessation (i.e., laccases) in different regions of the elongating internode at this stage of development. A study is underway to analyze the spatial and temporal expression of sorghum genes in elongating internodes (Kebrom, McKinley and Mullet, in progress). A group of genes involved in callose turnover were also expressed at high levels in the elongating internodes that could play a role in phloem development (Chen and Kim 2009).

Cellulose is the largest constituent of Della stem cell walls by mass and represents a major sink for sugars during growth and development. Eight *CESA* genes encoding cellulose synthases were differentially expressed at high levels in Int10 from floral initiation through ~25 days post-anthesis although with distinct patterns of expression. Sorghum *CESA4*, *CESA7*, and *CESA9* were identified as homologs of rice *CESA* genes involved in secondary cell wall

synthesis (Tanaka*, et al.* 2003). These genes showed peak expression 7 days before anthesis, followed by a ~10-fold reduction in RNA level by anthesis, and ~100-fold lower expression post-anthesis (Figure 2.10b). *CESA1, CESA3* and *CESA8*, genes involved in primary cell wall biosynthesis (Tanaka*, et al.* 2003) showed peak expression ~16 days before anthesis, then transient down regulation at anthesis, followed by a second phase of relatively high expression post-anthesis (Figure 2.10A). Synthesis of cellulose post-anthesis may be required to increase stalk strength in parallel with grain filling and increasing panicle weight at the top of the stem. Transient down-regulation of these genes at anthesis and decreased synthesis of cellulose and cell walls could help maintain sugar supply to developing panicles during this critical period of reproductive growth (Boyle *et al.* 1991, Ruan *et al.* 2010). The biochemical basis of transient down regulation of these *CESA* genes, and several other cell wall biology genes with similar patterns of expression is unknown. Interestingly, five genes encoding cellulases were differentially expressed in stems during development in a biphasic pattern similar to *CESA1/3/8* genes. One of these genes was homologous to *KORRIGAN1* (*KOR1*), a gene that encodes an endo-(1,4)-*β*-D-glucanase that associates with the cellulose synthase complexes in the plasma membrane (Vain *et al.* 2014). While the role of these genes in sorghum stems needs further investigation, cleavage of cellulose is observed during synthesis and cell wall maturation (Park and Cosgrove 2015).

The cell walls of some grasses contain mixed linkage (1,3;1,4)-*β*-D-glucans (MLG) that are synthesized by enzymes encoded by *CSLF* and *CSLH* gene families (Burton *et al.* 2006, Ermawar*, et al.* 2015, Vega-Sánchez *et al.* 2012). This polymer also accumulates in seeds of some plants, growing organs, and in mature stems (Vega-Sánchez*, et al.* 2012, Vega-Sanchez *et al.* 2013). *SbCSLF6* was expressed in Della stems at levels similar to other *CESA* genes from A-29d until A+25d, followed by gradually decreasing expression until A+68d. *β*-D-glucanases involved in MLG turnover have been identified (Burton and Fincher 2012, Hrmova and Fincher 2001) (Table 2.2). Three genes encoding MLG glucanases were expressed in sorghum stems throughout development.

Improving biomass conversion efficiency through lignin modification has been investigated extensively (Li *et al.* 2008, McCann and Carpita 2015, Wilkerson*, et al.* 2014). In sorghum, brown mid-rib *bmr*-mutants that affect lignin level and chemistry have been identified and used to produce forage crops with increased digestibility (Dien*, et al.* 2009, Sattler *et al.* 2012, Vermerris *et al.* 2002). The identification of genes for all steps in the lignin biosynthesis that are expressed in sorghum stems at different stages of development provides the starting point for additional modification of stem lignin to enhance LC-biomass conversion efficiency. Similarly, numerous genes involved in GAX biosynthesis were differentially expressed in sorghum stems during growth and development. The identification of sorghum homologs involved in GAX biosynthesis expressed

in stems provides a first step towards validation of function and utilization for stem composition engineering. Modification of GAX composition and/or cross-linking appears to be a promising approach to optimize biomass composition for specified conversion processes (Biswal*, et al.* 2015, Jung and Phillips 2010).

Overall, this study characterized the time course of biomass accumulation and changes in stem composition during development of Della, an energy sorghum that accumulates high levels of stem sugars. Transcriptome analysis identified ~200 genes involved in cell wall biology that were differentially expressed during development including multi-gene families involved in cellulose, GAX and lignin biosynthesis. The identification of sorghum genes encoding *CINs, VIN1* and *SUS4* that were expressed at high levels in elongating internodes and down-regulated in parallel with accumulation of sucrose in stems confirms and extends prior studies on sucrose accumulation in sweet sorghum. The identification of genes involved in cell wall biology and accumulation of NSC in stems will enable the engineering of these pathways to increase yield and conversion efficiency. In addition, follow on studies will utilize co-expression analysis to begin the process of identifying regulators of genes involved in stem biogenesis using approaches successfully applied in other plant species (Hirano*, et al.* 2013, Taylor-Teeples *et al.* 2015).

CHAPTER III

DYNAMICS OF STEM STARCH ACCUMULATION POST ANTHESIS

INTRODUCTION

Sweet sorghums are capable of accumulating high concentrations of nonstructural carbohydrates in their stems (Murray*, et al.* 2009). In the sweet sorghum Della, approximately 60% of stem nonstructural carbohydrate is sucrose with the remaining being glucose, fructose and starch. Increasing sucrose concentration in the stem of sugarcane and sweet sorghum has been the focus of previous research but sucrose levels have reached an apparent plateau (Wu and Birch 2007). The reasons for this limitation remain unclear, but may be attributed to feedback inhibition of photosynthesis, turgor constraints due to the high concentration of sucrose (~0.5-0.7M) within pith parenchyma cells in the stem, or sucrose signaling mechanisms (Wu and Birch 2007). However, increasing sucrose, if it could be achieved, may have its disadvantages. In sorghum, the sucrose in the stem is unstable at the time of harvest due to metabolic and microbial degradation (Wu*, et al.* 2010). Due to this instability, sorghum and sugarcane grown for harvesting sucrose from stems requires a nearby processing plant and efficient processing from harvest to sucrose extraction and stabilization (Wu*, et al.* 2010). Both of these problems can potentially be solved if the sink strength of the stem is increased through increasing flux into a more stable polysaccharide. Increasing the existing starch content of the stem is prime candidate for this purpose since (1) starch has low

osmotic activity and will have minimal impact on turgor as the polymer accumulates, (2) the hexose degradation products derived from starch are readily fermentable, and (3) the requisite pathways are already expressed in the sorghum stem.

Starch biosynthesis and degradation have been extensively studied in multiple experimental systems such as the Arabidopsis leaf, the potato tuber and rice and maize endosperms (Sonnewald and Kossmann 2013). In heterotrophic tissues that import sucrose, the process of starch biosynthesis begins when sucrose is hydrolyzed via the action of invertases or sucrose synthases. If hydrolyzed by invertase yielding glucose and fructose, glucose is phosphorylated by hexokinase producing glucose-6-phosphate. In many dicot species, G-6-P is imported into the amyloplast via a hexose phosphate transporter followed by isomerization to glucose-1-phosphate by phosphoglucomutase. The heterotetrameric enzyme ADP-glucose pyrophosphorylase (AGPase) catalyzes the conversion of glucose-1-phosphate to ADP-glucose representing the first committed step in starch biosynthesis (Hädrich *et al.* 2011). In leaf tissue, this reaction is rendered irreversible by the action of plastidial alkaline pyrophosphatase which hydrolyses inorganic pyrophosphate into orthophosphate (Stitt and Zeeman 2012). However, in grasses, an extraplastidial AGPase has been identified that is responsible for the synthesis of 95% of the ADP-glucose in the grain of maize (Denyer *et al.* 1996). This cytosolic ADP-glucose is then imported into the plastid by the action of

Brittle-1, an adenylate translocator capable of transferring ADP-glucose (Shannon *et al.* 1998). With both strategies, plastid localized ADP-glucose is utilized directly by starch synthase which is responsible for synthesizing the α-1,4-linkages of starch (Stitt and Zeeman 2012). Following linear polysaccharide synthesis, starch branching enzyme hydrolyses an oligosaccharide from the terminal end just below the granule surface and attaches it further into the growing granule, catalyzing the formation of an α-1,6-linked branch point (Tetlow and Emes 2014). This action leads to the formation of semi-crystalline amylopectin and increases the granule density, crystallinity, and number of non-reducing ends available for chain elongation by starch synthase (Tetlow and Emes 2014). The action of starch synthase in combination with glucan trimming by debranching enzymes fine tune the structure and crystallinity of the granule and it physical properties (Lorberth *et al.* 1998, Tetlow and Emes 2014).

The synthesis of starch creates a molecule that is highly stable (Hejazi *et al.* 2008). Because of this stability industrial processing and degradation of starch requires high heat to gelatinize starch. The addition of heat adds thermal energy to the polymer chains of the molecule increasing the physical space between them allowing for solvation followed by hydrolysis by starch hydrolytic enzymes. Since plants cannot readily add heat to disrupt granule crystallinity, destabilization in vivo is thought to be achieved through phosphorylation of the granule surface initially by glucan water-dikinase (GWD) which phosphorylates the C-3 position of terminal glucose monomers, followed by phosphorylation by

phosphoglucan water-dikinase (PWD) which phosphorylates the C-6 positions of different glucosyl residues (Blennow *et al.* 2002, Hejazi*, et al.* 2008, Ritte, 2006 #505, Stitt and Zeeman 2012). Phosphorylation by PWD is strictly dependent on prior phosphorylation by GWD (Ritte *et al.* 2006). Hydrolysis begins once the reducing ends of the amylopectin chains are accessible to β-amylases (exo-amylases that liberate maltose) and de-branching enzymes (which hydrolyze α-1,6-linked branch points). Additional degradation of oligosaccharides liberated from the granule is performed by the α-amylases (endo-amylases) and disproportionating enzyme (Critchley *et al.* 2001).

Numerous studies have attempted to increased starch content. Stark *et al*., 1992 tried to increase the activity of AGPase by transforming a variant of the gene from *E. coli* into potato that was less sensitive to effectors such as 3-PGA and phosphate (Stark *et al.* 1992). An increase in starch content was not observed in subsequent studies even though enzyme activity was four times higher than in the controls (Sweetlove and Burrell 1996). Other variants of AGPase that are less sensitive to phosphate (an inhibitor of activity) have also been used to increase the specific activity of the enzyme. This led to a consistently higher starch content per seed but not increased starch concentration per cell. Many other studies have attempted to increase starch by modulating this gene but overall the results have been mixed (Smith 2008). A small number of studies have focused on increasing starch through modulation of starch synthase. Problems associated with targeting this enzyme arise due to

the presence of multiple isoforms of the gene, each with unique roles in starch synthesis. Reduced expression of the individual isoforms usually results in alterations to the structure of the granule and overexpression of the individual isoforms may have a similar impact (Smith 2008). Increases in adenylate transporter activity to increase the supply of ATP to the plastid increased starch content in potato. Potato being a dicot utilizes plastidial ADP-glucose biosynthesis making this approach viable (Geigenberger *et al.* 2001, Smith 2008).

There is a lack of information in the sorghum and grass literature regarding stem starch. Therefore, in order to facilitate further research in this area, a biochemical and molecular characterization of stem starch biology was carried out. This investigation characterized the accumulation of stem starch during the reproductive phase of a sweet sorghum Della, which revealed the timing of starch accumulation during development. Microscopic analysis revealed the radial location of starch in the stem as it accumulates. Transcriptome analysis was performed to identify candidate genes involved in starch synthesis and turnover in sorghum stems. Together, this information will enable the design of energy sorghum with optimal levels of stem starch.

EXPERIMENTAL PROCEDURES

Field analysis

Della [*Sorghum bicolor* (L.) Moench] (Reg. no. CV-130, PI1566819), a sweet sorghum developed from a cross of Dale and ATx622 (Harrison and Miller 1993)

was planted at the Texas A&M Field station near College Station, Texas on March 28, 2013. The experimental design consisted of two ranges of 10 plots each of the sweet sorghum Della planted in adjacent regions of the field. Five plants were harvested from random locations within each of the experimental replications for a total of 10 plants per time-point. Before planting, diammonium phosphate (10-34-0) was applied at a rate of 100 kg N ha^{-1}. The spacing between rows was 76 cm and plants were thinned to 10 cm spacing at the five-leaf stage. This spacing corresponded to ~132,000 plants per ha^{-1} calculated as previously described (Olson*, et al.* 2012). Basal tillers were removed until anthesis when basal tillering ceased to maintain a planting density of one main culm per 10 cm. Apical tillers (referred to as branches) produced by plants starting a week before grain maturity, were not removed from field grown plants. Plants were harvested a minimum of 2 feet in from the end of the plot to avoid edge effects. Additionally, two boarder rows were planted on ether side of the experiment to prevent edge effects. During the experiment plants were irrigated as needed.

Harvesting began six days before anthesis in 2012 and 45 days prior to anthesis in 2013. The fresh weights of the leaves, leaf sheaths, stems and panicles were recorded. The tissues were placed in paper bags and dried in a forced air oven at 60°C for three days. The tissues were re-weighed to obtain the dry weight. In 2012 and 2013, internodes 12-14 were harvested for composition analysis. Since harvests began slightly before anthesis in 2012,

towards the end of stem elongation, internodes 12-14 were fully elongated at the first harvest. In 2013 harvests were started early in development. Due to the small size of the stem at the A-45 and A-29 time-points the entire stem was harvested. In later harvests the stem was sectioned into four sections; (1) internode 11 and under, (2) internode 12-14, (Penning*, et al.*) internode 13 and higher (if present) and (4) the peduncle (if present). This method of harvesting allowed for the total dry biomass of the stem to be obtained from the sum of the dry biomass of all the sections but also allowed for sampling that more accurately represented the internode section used in RNA-seq. Once internodes 12-14 were present (A-19 and beyond), those internodes were harvested for the remainder of the experiment. Internode 12-14 may not have been fully elongated at the time of harvest at day (A-19).

Della stem tissue harvest for RNA-seq

To obtain tissue for RNA-seq, Della was grown in a greenhouse in 2011 in day length of 14 hours and mid-day PAR of ~1100-1200 $\mu E/m^2/s$. Plants were thinned to one plant per pot. Plants were germinated in 5 gallon pots containing Metromix potting soil. Osmocote was added to the soil at the beginning of the experiment and additional fertilization was supplemented with Peter's nutrient solution every 30 days. The plants were watered with the use of an automated watering system to eliminate the possibility of drought stress. Leaves were numbered as they appeared ensuring that internode 10 was harvested at each of the eight sampling dates. Three biological replicates were harvested in the

morning on each harvest date. Internode 10 was excised at the upper and lower nodes to ensure that a complete internode was harvested for analysis. The internode was quickly sectioned into smaller pieces with a razor for storage and ease of processing after freezing. The internode sections were placed into a 50 mL conical tube, frozen in liquid nitrogen, and stored at -80°C until processing.

Starch quantification

The starch in stem biomass was quantified using the NREL Laboratory Analytical Procedure (with minor modifications) for extracting starch from solid biomass (Sluiter and Sluiter 2005). After the soluble NSC extraction, samples were washed four times in methanol-chloroform-water extraction solvent at room temperature for 15 minutes then centrifuged at 3,500 rpm for 15 minutes. After the MCW washes, the samples were oven dried at 65°C for 48 hours until dry. Next, DMSO was added to the dry samples and incubated overnight to allow the DMSO to fully permeate the samples. This imbibition step was essential to ensure a full permeation of the tissue with DMSO and therefore a complete gelatinization of starch. After imbibition, the samples were incubated at 100°C for 10 minutes with agitation to gelatinize the starch. Then MOPS buffer and thermostable α-amylase were added to digest the starch to maltose at 95°C. Subsequently, sodium acetate buffer and amyloglucosidase will be added to hydrolyze the maltose to glucose. Glucose was quantified with HPAE-PAD. The percentage of starch in the dry biomass was calculated from the glucose released during digestion. Starch content and concentration were calculated by

multiplying starch percentage of the dry weight by total internode dry weight of the sample thereby obtaining internode starch content. Internode starch concentration was obtained by dividing internode starch content by total internode volume to obtain the starch concentration.

Stereomicroscopy of internode cross-sections

To visualize the location of starch in the stem by stereomicroscopy field grown plants were harvested at anthesis and post grain maturity. A 2mm slice of stem tissue was excised with a razor blade from the middle of internode 11. These 2mm internode sections were infused with Lugol's solution in a vacuum created using a 50mL syringe. After infusion, the internode sections were washed and incubated in water to remove excess iodine and starch localization visualized using stereomicroscope.

Sequencing and alignment of the transcriptome

Internode 10 was pulverized into a fine powder in liquid nitrogen with a mortar and pestle. RNA extraction was performed using the Trizol extraction method. The high salt precipitation step was included because of the high concentration of carbohydrates in the sample (Molecular Research Center, 1995). Total RNA was further purified using the RNeasy kit from Qiagen which isolates RNAs >200 nucleotides in length. An on column DNase1 digestion was performed to remove any contaminating DNA. The quality of the RNA libraries was assessed using an Agilent Bioanalyzer. The mean RNA integrity number (RIN) from this quality control step was 8.9 and the minimum RIN was 8.3 indicating that the

quality of the RNA libraries was high. RNA libraries were prepped for sequencing using the Illumina TruSeq® V2 mRNA Sample Preparation Kit, which selects for polyadenylated RNA and excludes rRNA and non-polyadenylated RNAs from the library. The 24 libraries were sequenced on one lane of an Illumina HiSeq 2500 in single end mode. The average sequence depth per sample was 5.8M reads. The resulting 70bp reads were aligned to the *Sorghum bicolor* V3.1 genome using CLC Genomics Workbench version 6.5.1 and version 7 (CLC Inc., Aarhus, Denmark). Sequence reads were aligned to the 33,032 gene models annotated in Sbicolor-v2.1_255 provided through Phytozome (Patterson et al., 2009).

Analysis of gene expression was performed on RPKM normalized data. Differential expression and the FDR-adjusted p-values were calculated using the edgeR package in CLC Genomics Workbench, which uses non-normalized alignment data as the input for the edge statistical test. Further functional analysis was confined to genes that exhibited expression ≥ 2 RPKM at one or more time points, differential expression ≥ 3 between any two time-points of the experiment, and a FDR-adjusted p-value of differential expression < 0.05. The FDR adjusted p-value represents the p-value adjusted for multiple testing which was associated with the maximum fold-change values shown in the gene tables. In some cases, low expression values at one time-point led to a very high differential expression value. To minimize this effect, the time-point with low expression was set to a RPKM = 1, followed by manual calculation of the fold

change using the calculation (Fold Change = RPKM/1). These instances are indicated with the ">" sign in front of the fold change values in the gene tables. To analyze/visualize variation between samples and time-points, Principal component analysis was performed.

Sorghum gene annotation

Analysis of gene function and families utilized gene annotations generated by DOE-JGI for the *Sorghum bicolor* V3.1 genome available through Phytozome.

RESULTS

In order to determine the extent and kinetics of accumulation of starch in the sweet sorghum stem during development a phenology study was conducted in the field during 2012. During this experiment, internodes from mid stem were harvested for composition analysis. An analysis of nonstructural carbohydrates abundance in internode dry biomass revealed that the Della stem is capable of accumulating appreciable quantities of starch. The abundance of starch generally increased throughout development (Figure 3.1b). During the vegetative phase during rapid stem growth, only low levels of stem starch were detectable. At the flag leaf stage, stem starch began to accumulate gradually in parallel with increasing stem sucrose concentrations. Starch concentrations did not increase further during the period of grain filling. However, following the soft dough stage, ~95 DAE, the stem began to accumulate starch at a faster rate. This rate increase coincided with maximum stem sucrose concentrations observed during development. Starch continued to accumulate until post grain

maturity (~120 DAE, ~10 days after grain maturity). At its peak, starch accounted for ~9% of the stem's dry weight. After grain maturity levels of stem sucrose and starch decreased in parallel, with starch concentration declining ~30%.

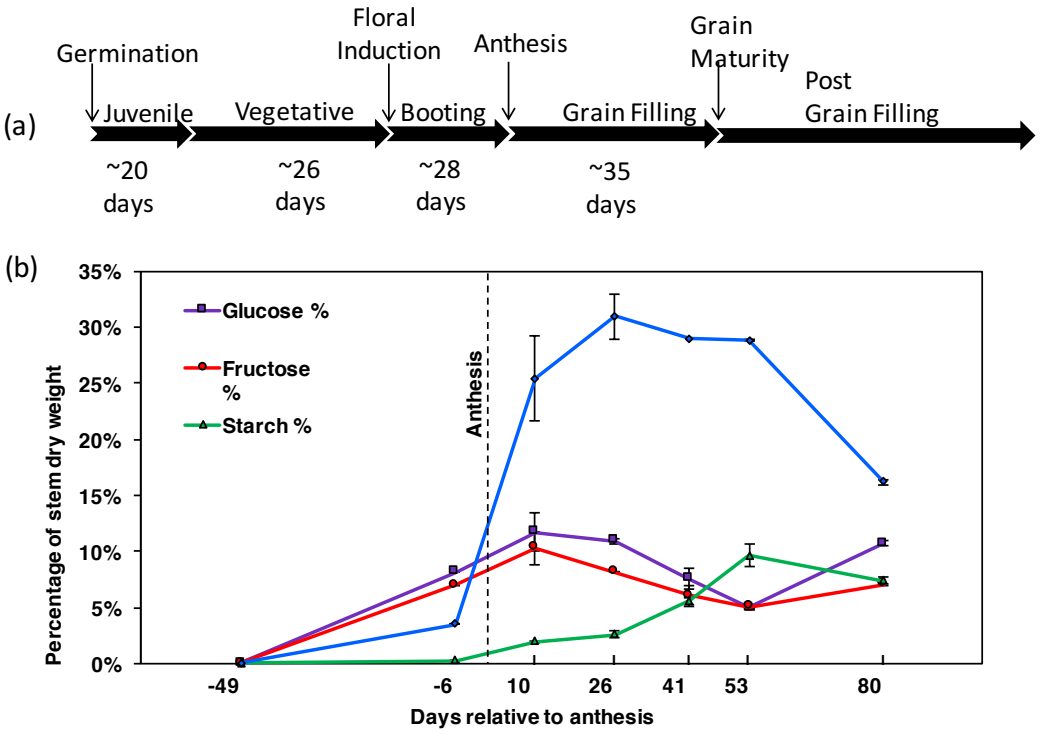

Figure 3.1. Developmental changes in sorghum NSC composition during the Della phenology conducted in the field in College Station, 2012. Anthesis occurs in Della at ~75 DAE. (a) The timeline of Della development. (b) The accumulation profiles of starch in the Della stem. (c) the accumulation profile of glucose, fructose, and sucrose during throughout time. Error bars represent S.E.M. Each data point represents the mean of 9 data points.

Nonstructural carbohydrates were quantified for each internode along the length of the stem at grain maturity. The results show that stem carbohydrate profiles vary depending on the position of the internode along the length of the stem. Internodes in the middle of the stem (internodes 10-16) have similar

carbohydrate concentrations while the internodes at the ends of the stem have carbohydrate profiles that deviate from the average. Sucrose concentrations were highest mid stem and declined toward the apical and basal regions similar to starch. In almost every internode, fructose concentration was slightly higher than glucose concentration with the abundance of both molecules increasing at the ends of the stem suggesting that additional sucrose hydrolysis occurs in these tissues. Starch concentration is lowest in the peduncle and is 5-7 fold higher in the internode below the peduncle. Proceeding down the stem (internode 17-9), there is a general trend of increasing starch concentration peaking at internode 11. The starch concentration was virtually the same with a slight decline in internode 9 and 10 relative to internode 11. Excluding the peduncle the percentage of the dry biomass occupied by starch is >5% in every internode of the stem at grain maturity excluding the peduncle and at its maximum was ~12.5% of the stem's dry weight.

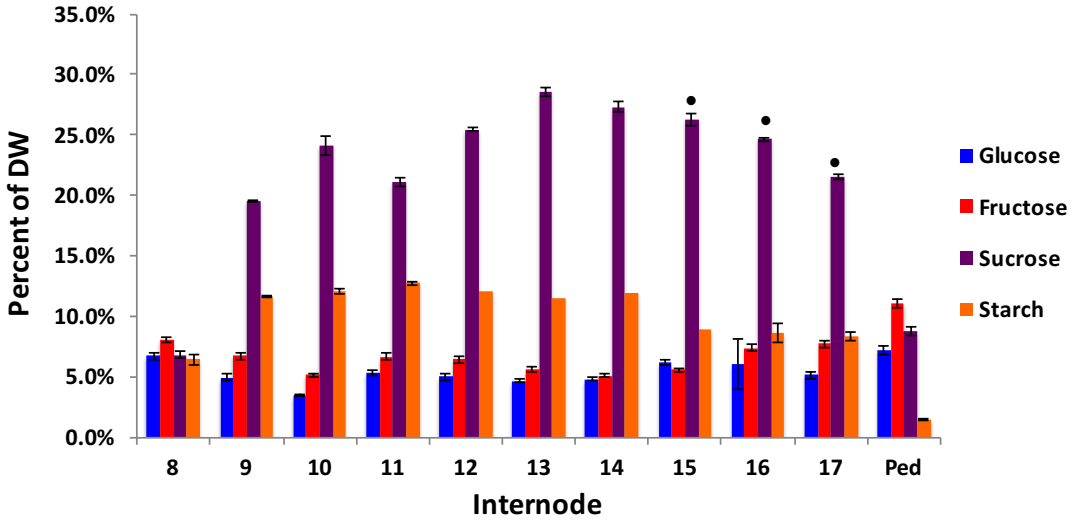

Figure 3.2. Longitudinal distribution of nonstructural carbohydrates in the stem of Della at grain maturity. Internode 8 represents the most basal internode and increasing numbers represent successive internodes higher up the stem proceeding to the peduncle. Asterisks represent internodes that had apical tillers at the time of harvest. Error bars represent S.E.M. Each data point represents the mean of 9 data points

Stereomicroscopy of internode sections infused with Lugol's solution revealed the radial distribution of starch in the middle internode of the stem of Della. Starch accumulated to higher concentrations in regions of high vascular bundle density. Starch concentration decreased with distance away from the vascular bundles. The results, shown in Figure 3.3, were of a plant that had moderate levels of starch in its stem. Staining experiments with other genotypes, that had higher starch concentration at the time of harvest, revealed greater starch densities in the ground pith parenchyma tissue. Interestingly starch did not accumulate within the vascular bundles.

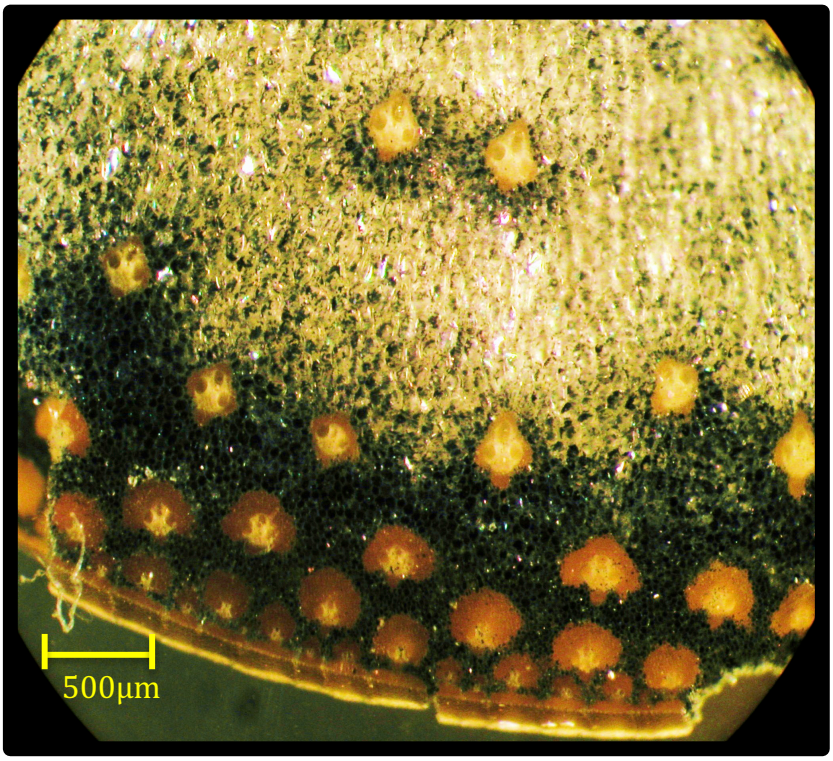

Figure 3.3. Radial distribution of starch in the sorghum stem at grain maturity. Lugol's solution was infuses under vacuum into mid stem internode sections harvested from mid internode at grain maturity. Starch accumulates around vascular bundles but does not accumulate within vascular bundles. Starch concentration is highest around the periphery of the stem indicating that starch biosynthesis is confined to the pith parenchyma cells.

These results show that the sweet sorghum Della is capable of accumulating significant quantities of starch throughout its stem beginning before anthesis and that accumulation occurs in parallel with sucrose accumulation. The time-course of accumulation shows that in this genotype starch content increased when stem sucrose concentration was high and declined as sucrose concentration declined. Microscopy results show that starch accumulated in the pith parenchyma cells of the stem and not in the vascular bundles. This indicates that starch biosynthesis genes are probably

differentially expressed in the pith parenchyma cells. To better understand the molecular events associated with the accumulation of starch in the internode, RNA-seq analysis was performed on whole internodes starting at floral initiation and ending after grain maturity. To this end, whole internode tissue, excluding leaves and leaf sheaths, was collected from Internode 10 during Della's developmental time-course to track changes in gene expression within that tissue. Since tissue collection started just before floral initiation in an early flowering genotype, internode 10 was selected since it was present at the earliest time-point. Field results show that this internode has the capacity to accumulate appreciable amounts of starch, sucrose, glucose, and fructose.

Sequences from RNAseq analysis were aligned to the *Sorghum bicolor* V3 genome. Genes that were expressed above an RPKM ≥ 2, fold change > 3, and an FDR < 0.05 were selected for further analysis. Genes from the starch biosynthesis and degradation pathways were identified using the Arabidopsis and rice functional annotations for the sorghum genome. Genes involved in the starch biosynthesis pathway such as ADP-glucose pyrophosphorylase (AGPase), starch synthase, and starch branching enzyme, were differentially induced during the accumulation of starch in the stem. Members of the AGPase gene family were expressed at very low levels in vegetative stems with gradually increasing expression as development progressed, peaking around grain maturity (A+25-A+43). The gene Sobic.009G245000 was differentially expressed 44 fold and was the most highly expressed member of this family

peaking in expression after grain maturity (A+43, Table 3.1). Interestingly, two members of the AGPase gene family had peak expression before anthesis during the stem growth phase (Table 3.1). Members of the starch synthase and starch branching enzyme gene families were expressed after anthesis, peaking at or after grain maturity (A+43-A+68) with similar expression profiles when compared to the AGPase gene family.

Table 3.1. Expression of sorghum gene involved in starch biosynthesis in stems

Function	TranscriptID	A-29	A-16	A-7	A	A+11	A+25	A+43	A+68	FC	FDR
Glucose-1-phosphate adenylyltransferase	Sobic.001G424500.1	82	40	17	20	13	16	6	5	-20	6.E-14
Glucose-1-phosphate adenylyltransferase	Sobic.001G424500.2	72	33	31	40	30	26	28	31	-4	2.E-06
ADP-glucose pyrophosphorylase	Sobic.001G462200.1	20	27	31	19	25	18	13	10	-3	3.E-06
ADP glucose pyrophosphorylase large subunit 1	Sobic.001G100000.1	5	10	9	8	13	29	19	15	4	6.E-03
glucose-1-phosphate adenylyltransferase large subunit	Sobic.003G230500.3	2	8	11	11	18	17	13	10	6	4.E-05
glucose-1-phosphate adenylyltransferase large subunit	Sobic.007G101500.3	25	24	25	20	46	67	59	47	3	4.E-03
glucose-1-phosphate adenylyltransferase large subunit	Sobic.002G160400.1	13	18	20	26	56	63	136	59	8	2.E-08
glucose-1-phosphate adenylyltransferase large subunit	Sobic.009G245000.1	10	21	65	48	295	413	555	216	42	1.E-13
starch synthase 3	Sobic.007G068200.1	1	3	2	3	5	6	3	2	4	3.E-04
starch synthase	Sobic.010G047700.1	3	7	5	4	51	108	106	96	26	1.E-08
starch synthase	Sobic.002G116000.1	27	43	175	48	328	348	354	277	10	2.E-09
starch synthase 2	Sobic.004G238600.1	16	12	11	7	13	16	12	19	3	3.E-03
Starch branching enzyme	Sobic.004G163700.1	1	5	5	3	6	7	8	4	4	7.E-06
Starch branching enzyme	Sobic.006G066800.1	18	33	37	31	172	348	559	347	23	1.E-12
Starch branching enzyme	Sobic.010G273800.1	2	5	4	4	15	18	38	20	13	4.E-10
Starch branching enzyme	Sobic.010G273800.2	3	5	5	4	15	20	37	21	10	1.E-08
Starch branching enzyme	Sobic.010G273800.3	2	6	4	4	14	19	38	20	13	1.E-10
Starch branching enzyme	Sobic.010G273800.4	2	6	6	5	16	23	67	21	21	1.E-12

Expression values are RPKM normalized values. Time points are designates in days relative to anthesis. Darker shades of blue represent higher expression. FDR p-value represents the minimum pairwise FDR corrected P-value between any two time points of the experiment.

In general, the expression of genes of the starch degradation pathway was lower than starch biosynthetic pathway genes but the pattern of expression of genes in both pathways was correlated (Table 3.1 and 3.2). Four members of the β-amylase gene family were expressed at low levels during stem growth and development (A-29 - Anthesis). These genes reached their maximum expression

after grain maturity between A+43 – A+68 coincident with a decline in soluble sucrose concentration in the stem juice. Interestingly, the most highly expressed members of this family exhibited peak expression early in the stem growth phase (A-29), and were subsequently down regulated ~50-60 fold during the reproductive phase of development. Genes such as an α-amylase, glucan-water dikinase (GWD) and phosphoglucan-water dikinase (PWD) were differentially up-regulated post grain maturity. GWD was ~60 fold induced at A+43 (Table 3.2). Interestingly the gene encoding PWD was expressed before GWD starting at A-16 even though evidence from other experimental systems suggests that its recruitment is dependent on prior phosphorylation by GWD. Additionally, PWD expression was reduced before the peak of GWD expression.

Table 3.2. Expression of sorghum gene involved in starch degradation in stems

Function	Transcript ID	A-29	A-16	A-7	A	A+11	A+25	A+43	A+68	FC	FDR
α-amylase	Sobic.003G276400.1	1.4	1.3	2.3	0.8	1.5	1.5	0.6	0.6	-4	7.E-03
α-amylase	Sobic.007G204600.1	2	4	10	6	14	21	39	20	12	4.E-13
β-amylase	Sobic.001G372100.1	50	3	2	2	1	2	1	2	-75	4.E-23
β-amylase	Sobic.001G293800.1	129	12	7	6	3	3	5	6	-54	9.E-21
β-amylase 1	Sobic.001G226600.1	0	0	0	1	2	2	1	3	>3	2.E-04
β-amylase 1	Sobic.001G508800.1	1	1	1	1	2	3	6	10	20	2.E-09
β-amylase 5	Sobic.002G329400.1	0	0	0	0	2	4	7	10	>10	2.E-14
β-amylase 5	Sobic.002G329500.1	1	1	1	2	1	1	3	2	4	3.E-02
α-glucan water-dikinase (GWD)	Sobic.010G143500.1	1	4	4	4	17	35	90	50	63	1.E-15
phosphoglucan water-dikinase (PWD)	Sobic.004G120100.2	2	8	7	12	13	10	10	6	6	1.E-03
phosphoglucan water-dikinase (PWD)	Sobic.004G120100.3	1	3	2	3	5	6	3	3	4	1.E-05
debranching enzyme 1	Sobic.001G107600.3	5	12	16	11	21	22	17	12	3	3.E-03
α-glucan phosphorylase isozyme	Sobic.001G083900.1	2	4	6	2	33	41	47	12	20	1.E-07
Disproportionating enzyme	Sobic.002G383700.1	1	2	2	1	3	5	6	5	5	2.E-05
Disproportionating enzyme 2	Sobic.002G407000.1	4	10	17	9	25	34	41	29	9	2.E-08
like SEX4 1	Sobic.007G116600.1	11	9	11	7	9	8	7	5	-3	3.E-03

Sorghum stem internode samples were collected from greenhouse grown plants 29 days before anthesis (A-29) through 68 days post-anthesis (A+68). Expression values are presented as RPKM and each data point represents the mean of three biological replicates. Higher expression is represented by darker shades of blue. FC = fold change. FDR adjusted p-values were calculated for the maximum fold change in expression for each gene in the experiment.

DISCUSSION

Timing and location of stem starch accumulation and it ramifications on yield

In this study, the rate of sucrose accumulation in stems increased substantially by anthesis with the onset of accumulation correlated with cessation of stem growth (Figure 3.1c). Stem starch began accumulating simultaneously with sucrose although at much slower rates. The amount of stem starch plateaued during grain filling and then increased thereafter with the rate of deposition being greatest when the concentration of sucrose reached its maximum (~30% of the dry weight) slightly before grain maturity. Starch levels peaked post grain maturity and began to decline when sucrose concentration declined. These results show that sorghum utilizes starch as an additional carbohydrate storage reservoir in tandem with sucrose. At grain maturity, stem starch levels were in excess of 12.5% of the dry weight and it is likely that other sweet sorghums are capable of accumulating even higher amounts of stem starch. The results from this study show that stem starch is a significant component of the fermentable biomass of the stem and should be considered in future yield calculations. The glucose from stem starch would increase the current fermentable carbohydrate yield of the sweet sorghum by as much as 10-15% (as a fraction of nonstructural carbohydrates; excluding lignocellulosic biomass).

Starch accumulates in stems initially in areas closest to the vascular bundles, which are predominantly located in the cortical region of the stem but

starch can also accumulate throughout the ground tissue. One hypothesis is that the sucrose concentration is highest in regions closest to the vascular bundle and therefore the flux of starch synthesis/degradation in the cells in that region is greater. Additionally, since starch is synthesized in the pith parenchyma cells of the stem and not in the vascular bundles this indicates that the genes of this pathway are also differentially expressed exclusively in this cell type and that the identification of pith parenchyma specific promoters are needed to engineer the pathway. Analysis of tissue specific gene expression will enable the identification of promoters that are active in the pith parenchyma cells and are expressed after anthesis.

CHAPTER IV

EVIDENCE OF SINK LIMITATION IN SORGHUM

INTRODUCTION

A central problem facing bioenergy crop research is maximizing biomass yield, particularly in crops other than maize (Olson*, et al.* 2012). Species targeted for improvement include Poplar, Switchgrass, Miscanthus, Sorghum, and Sugarcane. These species have been selected due to their ability to accumulate high biomass. Excluding Poplar, all of these species are grasses that utilize C4 photosynthesis which accumulates dry biomass at a faster rate than C3 photosynthesis. Additional qualities that support high biomass yield in sorghum include drought and heat tolerance as well as resistance to biotic stressors.

During sorghum development the plant partitions a significant amount of energy to building the stem, root system, and canopy. Following anthesis the plant begins to fill the stem and grain with nonstructural carbohydrates (McKinley *et al.* 2016). At peak carbohydrate concentration the biomass of the stem is composed of ~50-60 percent of the dry weight as nonstructural carbohydrates. Increases in NSC yield of stem crops have stalled in recent years with plateaus observed in sucrose concentrations in stems of sugarcane (Gutjahr*, et al.* 2013a, Wu and Birch 2007). Considering that the sorghum stem is composed of ~70% water, significant yield gains could be made if this fraction of the biomass yield was increased (Mullet*, et al.* 2014).

In sorghum, stem dry biomass accumulation consists of two main phases; the growth phase and the storage phase (Chapter II, Figure 2.1a and b, Figure 2.5). During the growth phase, the plant primarily builds cell walls. Once growth is complete, the stem transitions into a storage organ capable of accumulating sugar and starch (Ferraris 1981, Gutjahr*, et al.* 2013b, HoffmannThoma*, et al.* 1996). The onset and kinetics of carbohydrate accumulation in the stem seem to be flexible and dependent on genotype as well as environmental factors. Some sorghum genotypes accumulate sugar beginning at floral initiation. However, in other sorghum genotypes the accumulation of stem sugar is separated in time from the growth phase (Ferraris 1981, Gutjahr*, et al.* 2013b, HoffmannThoma*, et al.* 1996). The accumulation of high concentrations of stem NSC is particularly apparent in sweet sorghums, which have historically been selected for high nonstructural carbohydrate concentrations in stem at maturity (Murray*, et al.* 2009).

Numerous studies have been conducted to determine what plant phenotypes affect total carbohydrate yield of the stem. Genetic studies have found an association between brix and stem pithiness. Evidence suggests that stem pithiness (dead air-filled cells) reduces the aqueous volume of the stem available for carbohydrate accumulation this apparently increases NSC concentration in the remaining volume of the biomass (Felderhoff *et al.* 2012, Murray*, et al.* 2008). A related observation has been made regarding midrib coloration. Midrib coloration (ether brown or white), which is commonly used as

a qualitative indicator of the lignin content of the plant, has also been found to be a general indicator of the pithiness of the plant. The greater stem pithiness results in greater concentration of NSC in the remaining biomass due to inverse relationship between pithiness and NSC concentration as mentioned above (Gutjahr*, et al.* 2013b). Interactions between source and sink tissues appears to impact NSC yield as well. One study found a negative correlation between leaf dry biomass and stem carbohydrate concentration suggesting that the ratio of source to sink may be important (Sylvain Gutjahr et al., 2013). Another study suggested that selection of non-senescent genotypes may be desirable for stem carbohydrate accumulation suggesting that retaining source tissue would be advantageous (McBee *et al.* 1983).

Past studies have investigated the potential trade offs between stem carbohydrate yield and grain yield. One study suggested that the dynamics of carbohydrate partitioning between these two organs is due to the timing of developmental events. In experiments in which the stem filled with carbohydrates before anthesis there were no trade-offs between stem and grain filling (Gutjahr*, et al.* 2013b). Other authors have suggested that each sink acts independently with a sink-specific strength, and when excess photoassimilate exists, sinks do not compete (citation). Indeed evidence for this has been observed in genetic studies in grain/sweet sorghum crosses, the loci that impact grain NSC yield and stem NSC yield did not co-align under non stressed conditions but di co-align under water-limited conditions (Murray*, et al.* 2008).

Efforts to further understand limitation to carbohydrate yield of sorghum are ongoing. For this investigation a developmental study of the sweet sorghum Della was conducted in order to analyze of a broad range of traits that potentially impact carbohydrate yield. Additionally, we conducted an analysis of carbohydrate yield in a segregating grain*sweet sorghum cross RTx436*Della that exhibited a broad range of phenotypic variation to better understand which traits influence carbohydrate yield. These results suggest that increasing sink strength of the sorghum stem would be beneficial.

EXPERIMENTAL PROCEDURES

Della field experiments

Field experiments of the Della phenology were conducted in College Station during the 2013 season. The experimental setup consisted of two ranges of 10 plots each of the sweet sorghum Della. Each experimental replication was flanked by tow boarder rows. Rows were spaced at 76cm. Plots were thinned to culm per 10 cm spacing at the five-leaf stage. This spacing corresponds to 132,000 plants per hectare. Basal tillers were removed as they appeared to maintain the desired culm density. Five plants were randomly selected from each experimental replication for a total of 10 plants harvested per time point. The inner row at least 2 ft. from the end of the row. Harvested plants were taken to an air-conditioned field house and phenotyped for a broad range of traits. To account for the total dry biomass of each plant, plant material was dried in a

forced air oven at 60°C and weighed to obtain dry weight. The total above ground biomass was the sum of the dry biomasses of the individual tissues.

Radiation use efficiency analysis

Radiation use efficiency (RUE) is a measurement of how efficiently a crop converts intercepted photosynthetically active radiation (PAR) into dry biomass (g DW MJ^{-1}) for a given unit of leaf area. RUE can commonly calculated using ether the total solar incident radiation (S_t) which includes near infrared radiation (NIR) or it can be based exclusively on PAR (S_{PAR}). RUE_{PAR} was calculated for each harvest interval in order to identify differences in RUE during plant development. To calculate the RUE_{PAR}, the total above ground dry biomass of Della was derived from the sum of individual tissue dry weight measurements. The above ground dry biomass was used instead of total root and shoot biomass due to practical limitations of accurately quantifying root dry biomass under field conditions.

To obtain the total daily solar radiation intercepted by the crop (S_i) the fractional radiation interception $(\mathcal{E}_i)$ was calculated as shown in equation 1. The conversion factor S_{par}/S_t = 0.45 was used (Howell, 1983). The daily extinction coefficient of 0.5 was used since this is considered a reasonable approximation for this canopy type (Olson et al., 2012).

(1) $$\mathcal{E}_i = 1 - e^{-0.5*LAI}$$

In addition, in equation 1, the leaf area index, (LAI) is square meters of leaf area per square meter of land. To obtain the LAI, the total green leaf area was measured using a LiCOR LI-3100C (Lincoln, Nebraska). To accurately represent the leaf area available for light interception during a given harvest interval, the average leaf area was used in RUE calculations and was calculated as the average LAI between two time points.

(2) $$S_i = S_t \mathcal{E}_i$$

Using equation (2), the solar radiation intercepted by the canopy (S_i) per square meter of leaf area per day was calculated by multiplying the total daily incident solar radiation (S_t) by $\mathcal{E}_i$. S_t was downloaded from NASA at (http://power.larc.nasa.gov/ cgibin/cgiwrap/solar/agro.cgi?email=agroclim@larc.nasa.gov). Once (S_i) was

calculated for each day of the experiment, the solar radiation intercepted during each interval of plant growth was calculated by summing (S_i) for all days in a given interval.

Genetic mapping of RTx436*Della F_4

To investigate the genetic basis of carbon partitioning and sugar accumulation in the stem, a cross was made between the elite grain sorghum inbred RTx436 and the elite sweet sorghum inbred Della. This population was expected to segregate for traits that impact stem sugar concentration. The progeny were selfed to the F_4 generation, which resulted in the creation of 88 recombinant inbred lines (RILs). This population was grown in College Station in the summer of 2013 in a randomized, double replication design. The population was thinned at the 5-leaf stage to a density of 10 cm spacing between plants. The spacing between rows was 76 cm. The plants were irrigated when needed. To account for variation in sugar accumulation at different stages of development, flowering time was recorded to facilitate harvesting at 10 days post-grain maturity. Three plants were harvested from each RIL in each replication, for a total of six biological replicates. At the time of harvest, a number of phenotypes were measured such as the fresh and dry biomass of each tissue, stem length and diameter and leaf area. To measure dry biomass tissues samples were dried in a forced air oven at 65°C for three days and the re-weighed. To quantify the composition of the stem, internode sections were harvested from the middle of the stem, which is the region of the stem that has the highest carbohydrate

concentrations as well as the fewest extraneous influences on NSC concentration. For each RIL, a three-internode section was harvested from the middle of the stem (excluding the peduncle from the measurement of stem length). This internode section was also phenotyped for length, diameter, fresh weight and dry weight. The average internode length was the length of the three-internode section divided by three. The internode volume was calculated using average single internode length and radius at the middle of the three-internode section.

DNA was extracted from the combined juvenile leaf tissues of 10 plants per recombinant inbred line using the MP Biosciences Fastprep-96 DNA extraction protocol. Next, the DNA concentration was quantified using the AccuBlue protocol. DNA sequencing template preparation was carried out by the method described in Morishige et al., 2013. The enzyme Fse1 was used in order to create a genetic map for composite interval mapping. Adaptors were ligated that contained barcode sequences used to de-multiplex the reads following sequencing. Prepared template was sequenced on an Illumina HiSeq 2500. The reads were aligned to the Sbicolor-v2.1_255 reference genome and markers were called based on the identification of polymorphisms between the parental genotypes. The genotype data from the QTL analysis package R/qtl package was used to create the genetic map as well as to map QTL. Prior to any quality control steps in the creation the genetic map missing genotype data was filled in using the fill.geno() to reduce the number of markers eliminated by

subsequent steps of the map curation process. Markers that were genotyped in less than 75% of the RILs were eliminated from the map. Map distances were estimated using the Kosambi map function. Following genetic map creation, composite interval mapping was performed for all the traits measured.

Nonstructural carbohydrate quantitation

Soluble carbohydrates in the dry biomass used for NIR analysis were quantified by grinding the biomass to a small particle size in a Cyclone Sample Mill (Udy Corporation, Fort Collins, Colorado, USA) followed by removal of residual moisture. Next, 200mg of finely ground biomass was weighed (± 0.5mg) and transferred to a 15mL conical glass tube. Water-soluble NSCs were extracted in 10mL of water/Na azide solution at 50°C for 48 hours with agitation. This extraction time was determined empirically and allowed the extraction solvent to fully penetrate and extract soluble carbohydrates. The prepared samples for HPLC, 20 µL samples were diluted 50X into de-ionized water. All HPLC samples were sterile filtered using 0.45 µm cellulose acetate sterile filters. Sucrose, glucose, and fructose concentrations were quantified using high performance anion exchange-pulsed amperometric detection (HPAE-PAD) with a Carbopac PA1 analytical column (Dionex) as well as a Borate trap (Dionex) and a Aminopac column (Dionex) to reduce borate and amino acid interference (Murray et al., 2008; (Dionex 2003a, Murray*, et al.* 2008). A solution of 75mM NaOH was made from 50% NaOH solution to eliminate carbonate contamination. This running buffer was vacuum degassed and stored under a

helium bed for the duration of the chromatographic run. The standard curve was validated by the incorporation of curve validation samples throughout the experiment in accordance to the NREL laboratory analytical procedure (Wolfrum et al, 2013).

Stem starch was quantified using the NREL Laboratory Analytical Procedure for extracting starch from dry biomass (Sluiter and Sluiter 2005). After soluble carbohydrate extraction, samples were washed four times with methanol-chloroform-water. The samples were centrifuged at 3500 rpm for 15 minutes after each wash. The samples were oven dried at 60°C. To gelatinize the starch, DMSO was added to the dry samples and allowed to permeate the samples for 24 hours in refrigeration. This step ensures complete saturation of the dry biomass with DMSO and therefore ensures complete gelatinization of the starch. The samples were then incubated in a water bath at 100°C for 15 minutes with occasional vortexing to facilitate starch gelatinization. MOPS buffer and thermostable α-amylase were added to digest the starch to maltose. Following amylase digestion, sodium acetate buffer and amyloglucosidase were added to hydrolyze the maltose to glucose. Glucose was quantified with HPLC as described above using HPAE-PAD. The percentage of starch in the dry biomass was calculated from the glucose released during digestion.

RESULTS

Della development and biomass accumulation

Della, a sweet sorghum inbred line, was planted at the Texas A&M University Field Station in College Station Texas between April 2 and April 10 in 2012 and 2013. Plant growth and development was similar in both years although some variation was observed due to differences in environmental conditions and plot location. The timing and duration of developmental stages and the rate of biomass accumulation of Della plants under field conditions is shown in Figure 4.1. Della plants underwent floral initiation between 45-50 days after emergence (DAE), reached anthesis by ~75-80 DAE, and grain maturity between ~106-110 DAE (Figure 4.1a). The current study focused on changes in biomass accumulation, composition and gene expression that occurred between floral initiation and 30 days post grain maturity (~100 days).

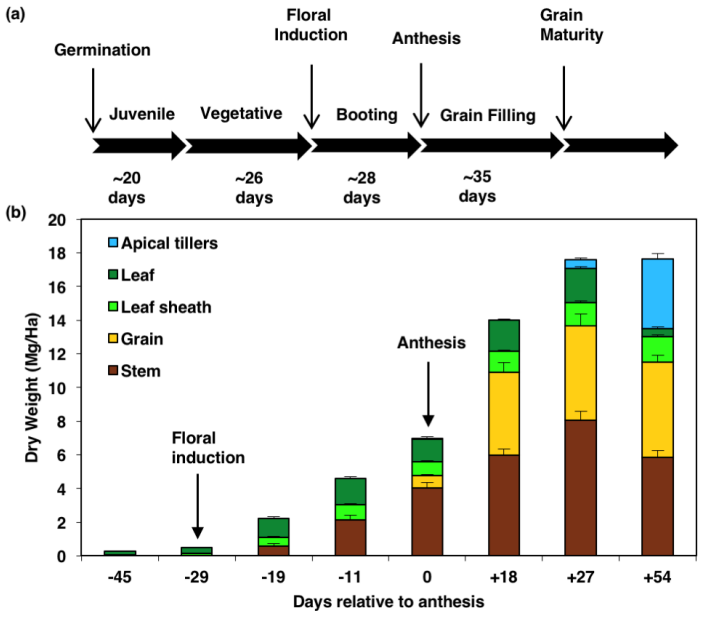

Figure 4.1. Time course of biomass accumulation during Della development.
(a) Duration of stages of Della development from germination to floral induction, floral induction and booting to anthesis, and anthesis to grain maturity.
(b) Time course of dry weight accumulation (Mg per hectare) in leaves (dark green), leaf sheaths (light green), stems (brown), grain (yellow) and apical tillers (blue) in the field in College Station, 2013. Under field conditions anthesis occurred at 79 days after emergence (DAE) and grain maturity at 106 DAE. The data shown are the mean ±SE of nine biological replicates.

The number of days from sorghum seedling emergence to floral initiation can vary significantly (>100 days) depending on genotype, day length, temperature, and other environmental factors. However, the number of days from floral initiation to anthesis and from anthesis to grain maturity is relatively constant in sorghum under good growing conditions (Vanderlip and Reeves 1972). Therefore, the timing of events and other data collected in this study is

expressed in days relative to anthesis (i.e., days before anthesis (A-#d) or after anthesis (A+#d)), in addition to days from plant emergence (DAE). Following germination, seedlings accumulated biomass slowly for approximately 50 days and then at an increased rate until grain maturity at ~106 DAE (A+27d) (Figure 4.1b). Leaf area per plant began increasing rapidly between 35 and 40 DAE reaching a maximum of ~2,000 cm^2 between ~70-75 DAE, a few days before anthesis (0d). Stem length, excluding the peduncle, increased rapidly following floral initiation (~50 DAE, A-29d) due to production of longer internodes. Stem growth rate decreased starting ~10-14 days before anthesis with final stem length reached by anthesis when the last internode below the peduncle was full length.

Della shoots accumulated ~18 Mg ha^{-1} of dry biomass under field conditions in 2013 by grain maturity with grain accounting for 5.6 Mg ha^{-1} (Figure 4.1b, yellow bar). Dry biomass of leaves on the main stem reached a maximum between anthesis and grain maturity (~2 Mg ha^{-1}) (Figure 4.1b, green bar). Stem dry biomass was ~4 Mg ha^{-1} at anthesis and reached ~8 Mg ha^{-1} at grain maturity (A+27d; 106 DAE) (Figure 4.1b, brown bar). Stem dry biomass declined after grain maturity coincident with an increase in apical tillering. Apical tillers, which began growing a few days before grain maturity, accumulated 4.1 Mg ha^{-1} by 133 DAE (A+54d) (Figure 4.1b, blue bar).

Interval radiation use efficiency

The analysis of interval RUE for the sweet sorghum Della revealed complex dynamics in the pattern of radiation use during growth. Early in development RUE was low and began to increase at ~35 DAE (Figure 4.2a). Between 35 DAE and 70 DAE the RUE increased rapidly and was coincident with an increase in the rate of dry biomass accumulation (Figure 4.2a and b). Between 70-80 DAE, slightly preceding but in the same interval as anthesis which occurred at 79 DAE, the RUE declined by ~60% due to to an increase in radiation interception by >2-fold relative to the prior interval. Coincident with the decline in RUE at anthesis the stem continued to accumulate dry biomass while biomass accumulation in the leaves ceased. The cessation of biomass accumulation in the leaves was due to the completion of canopy development at 79 DAE (Figure 4.2d). There was a strong increase in the rate of dry biomass accumulation during the interval between anthesis and grain maturity attributed largely to NSC accumulation in the panicle and stem. This increase in Dry biomass bolstered RUE between 80-100 DAE (Figure 4.2a-c). After 100 DAE the RUE declined from ~2.75 gMJ^{-1} to ~0 gMJ^{-1}. This reduction in RUE coincided with apical tiller outgrowth, a reduction in stem dry weight, and a reduction in leaf dry weight (Figure 4.2c). The rate of dry biomass accumulation post grain maturity was ~0 at 130 DAE while the leaf area at this time had increased slightly by ~200 cm^2 after anthesis when considering additional leaf

area added by the branches (Figure 4.2d). Light interception during this interval therefore increased slightly while the dry biomass did not increase at all.

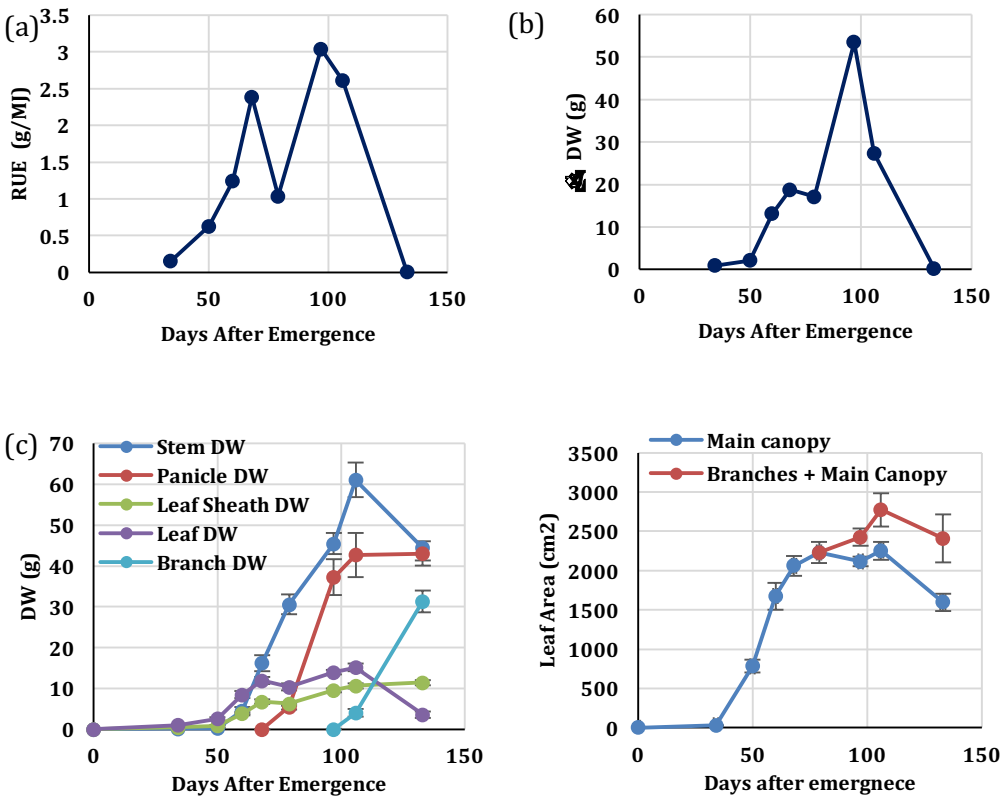

Figure 4.2. Dynamics of radiation use during the sweet sorghum phenology in the field in College Station, 2013. (a) Interval radiation use efficiency during the 2013 phenology study. (b) Change (Δ) in dry biomass during each interval. Data point is displayed on the last day of each interval shown. (c) Tissue specific dry biomass accumulation profiles. (d) Leaf area with and without branches leaf area included. Anthesis and grain maturity occurred at 79 and 106 DAE respectively. For each data point n=10.

The relationship between stem volume and carbohydrate yield

Pairwise Pearson's correlation analysis was conducted on a grain by sweet sorghum bi-parental cross RTx436*Della F_4. Figure 4.3 shows a matrix of

pairwise Pearson's correlation coefficients (PCC) of a selection of phenotypes such as days to flowering, biomass phenotypes, and stem composition. The total NSC content of the internode at grain maturity was highly correlated with internode volume (PCC = 0.85) and more highly correlated with internode length (PCC = 0.82) than internode diameter (PCC = 0.17). Total NSC content at grain maturity was also highly correlated with internode fresh and dry weight at anthesis and grain maturity. The total NSC content at grain maturity was highly correlated with sucrose and starch concentration at grain maturity (PCC = 0.88 and 0.87 respectively) and correlated with glucose and fructose content (PCC = 0.70 and 0.80). There was a low correlation between internode NSC content at grain maturity and leaf area (PCC = 0.32). The correlation between NSC content and concentration at grain maturity was PCC = 0.42. The correlation between total stem NSC content and panicle dry weight at grain maturity was PCC = -0.13. Additionally, high correlations were observed between components of the cell wall and internode dry weight and volume.

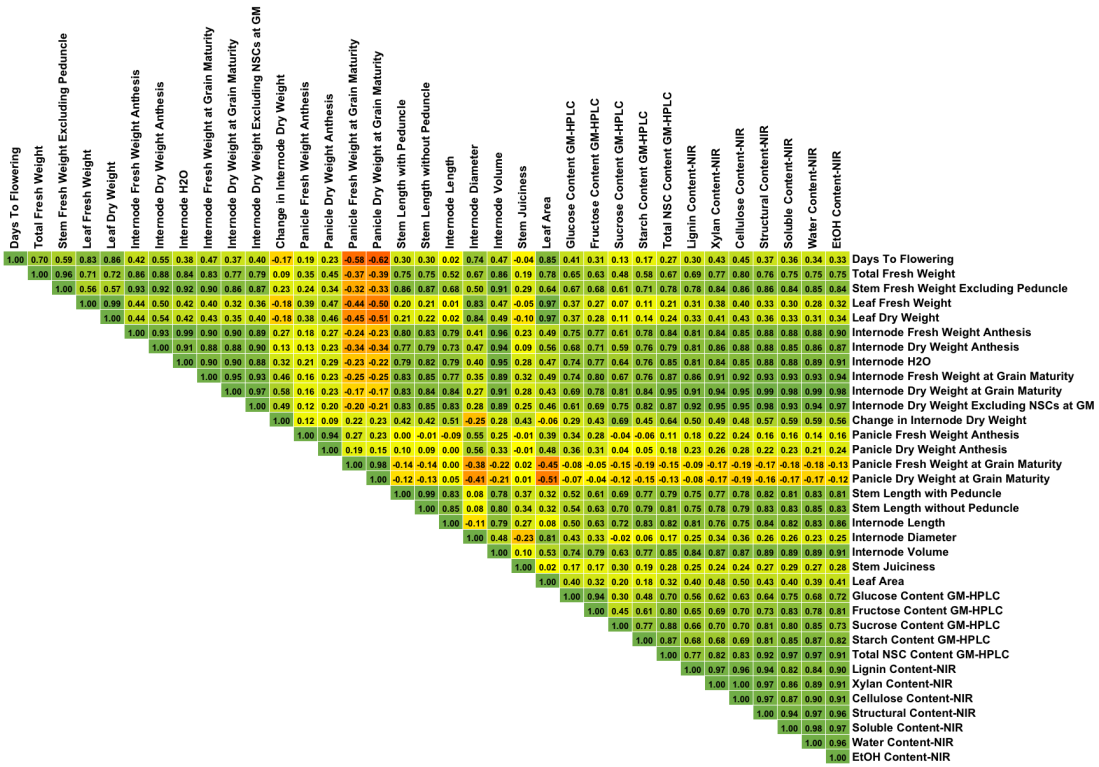

Figure 4.3. Pairwise correlation matrix of biomass and composition phenotypes for the population RTx436*Della F_4. (a) Correlation between total NSC content and internode volume. (b) Correlation between total NSC content and leaf area. Green indicates positive correlation, red indicates negative correlation and white indicates no correlation. Internode volume was calculated from internode length and stem diameter, n=3, for each RIL of the population. To assay stem NSC concentrations samples were bulked with two HPLC technical replications per RIL.

QTL analysis of RTx436*Della F_5

Using the phenotype data collected from RTx436*Della F_4 plants, composite interval mapping was performed to identify QTL and genetic relationships between the traits segregating in this population. In general, there are significant differences between the parents used to construct this population. A

comparison of flowering time shows that Della flowered later than RTx436 suggesting that segregation for flowering time alleles. Additionally, Della also exhibited a larger stem diameter and was sweeter than the grain parent RTx436.

QTL analysis of flowering time detected a QTL on linkage group 6 (LG-06) located between the physical coordinates 38-56 Mb with a maximum LOD score of 20.3. This QTL likely corresponds to the maturity loci Ma_1 that was cloned and discovered to be pseudo response regulator 37 (PRR37) a repressor of flowering under long day conditions (Murphy*, et al.* 2011). In this population, Della contributed the dominant alleles for this gene. Additionally, this QTL exhibited an impact on the total length of the stem, although it was not detected above threshold. Ma_1 was discovered to impact other traits in this population. In particular, it was identified as a QTL for stem diameter (LOD = ???). An additional flowering time QTL located on LG-01 was also detected above threshold location between 25 and 45 Mb with a maximum LOD score of 6.4.

QTL were detected for internode and stem length on LG-07 and LG-09. In both cases, the Della alleles accounted for an increase in this phenotype. These loci likely correspond to the well-known dwarfing loci Dw_3 and Dw_1 respectively (Figure 4.4 a, c, e, and g). Of the two dwarfing loci, Dw_3 had a more substantial genetic effect on internode length and internode dry weight at grain maturity than Dw_1 (Figure 4.4 C and G). All three loci (LG-06, LG-07, and LG-09) were associated with variation in internode fresh and dry weight and grain maturity.

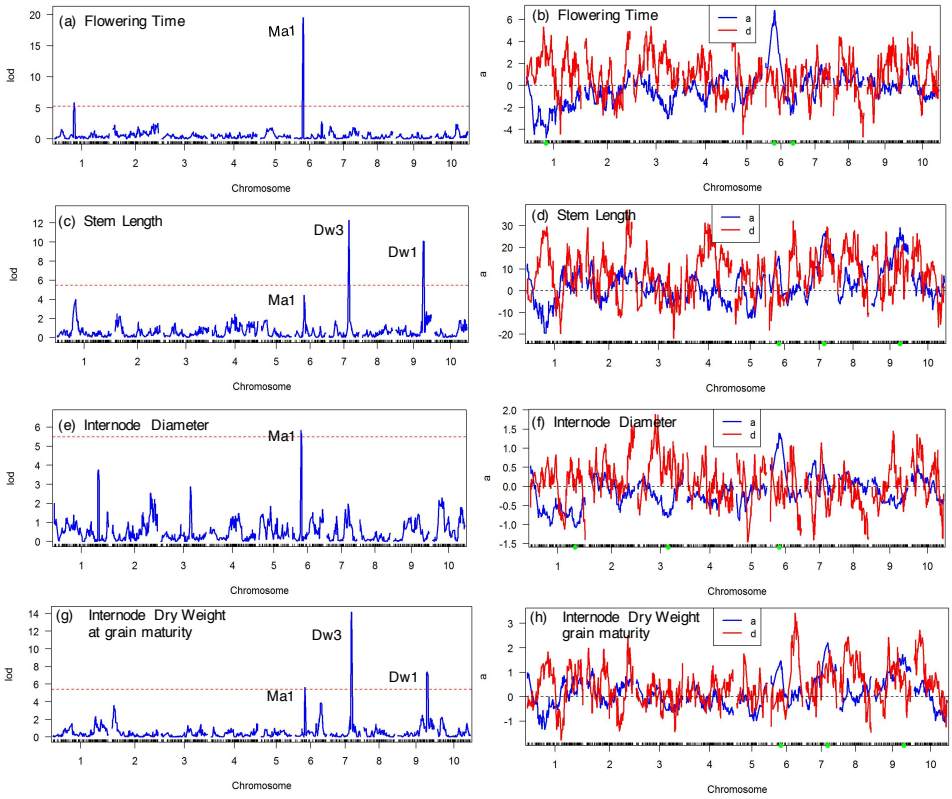

Figure 4.4. QTL maps and genetic effects for traits that govern biomass in RTx436*Della F_4. (a & b) QTL map and effect plot, (c & d) stem length and effects, (e & f) internode diameter and effects (g & h) stem dry biomass and effects. The red line in the QTL maps represent the statistical threshold of $\alpha = 0.01$, determined with 1000 permutations. The blue line in the effect plots represents the additive effects (a) and the red line represents the dominance effects (d).

Genetic analysis of sucrose, starch, glucose, and fructose content of the internode detected the contribution of QTL on LG-07 and LG-09 that were likely attributed to the dwarfing loci. The most statistically significant effects of these loci were seen for sucrose and starch content of the internode. The effects of these loci were also detected for glucose and fructose but were not statistically significant (Figure 4.5a, c, e, and g). Each locus contributed approximately

equally to the phenotype for each of the four traits. For sucrose content of the internode each QTL had an additive effect of ~ 2 g per internode (Figure 4.5b). The additive effect on starch content per internode was ~0.5 g per internode (Figure 4.5d). In all cases, the additive effect of the Della allele was greater than the RTx436 allele.

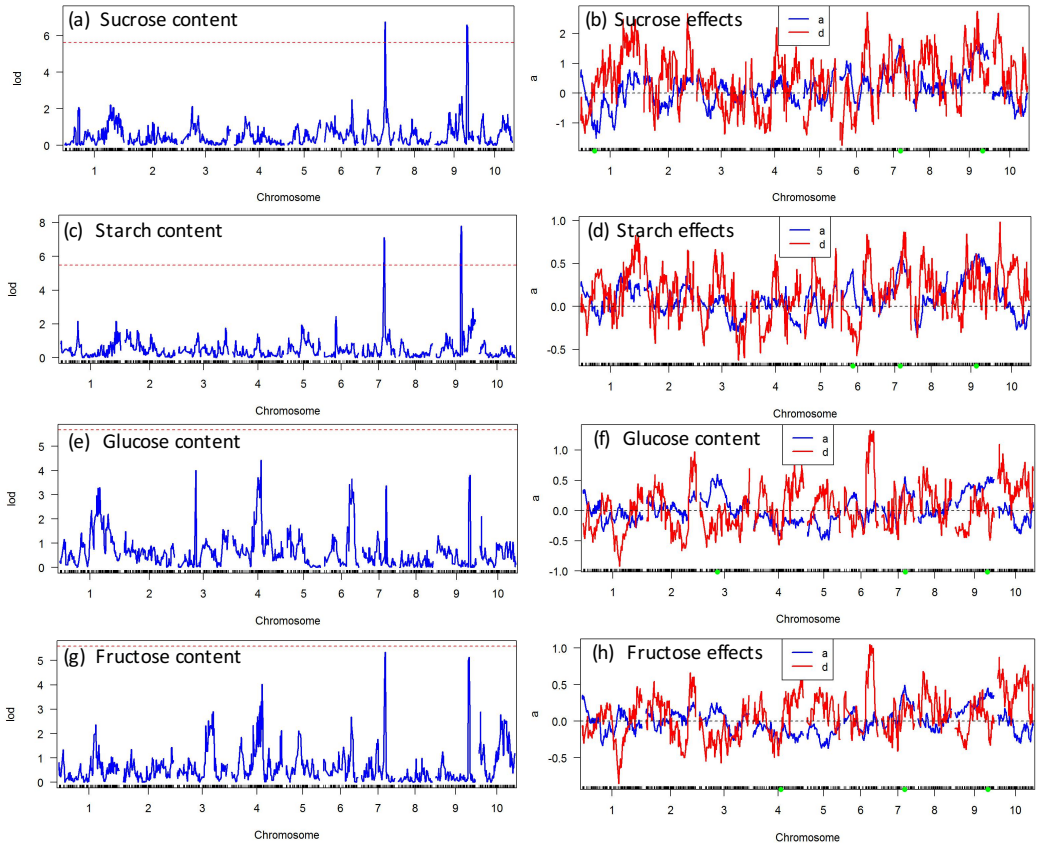

Figure 4.5. QTL maps and genetic effects for non-structural carbohydrate yield in RTx436*Della F_4. Traits are as follows (a) Sucrose content, (b) Sucrose genetic effects, (c) starch content, (d) Starch genetic effects, (e) glucose content, (f) glucose genetic effects, (g) fructose content, (h) fructose genetic effect. The dotted red line in (a, c, e, and g) represents an α=0.01 determined by 1000 permutations. The blue and red lines in (b, d, f, and h) represent the additive and dominance effects respectively.

DISCUSSION

The results of the 2013 Della phenology study show that the RUE of sorghum was biphasic with three periods of low RUE. The first period of low RUE occurred during early vegetative development. RUE was low likely because of cool temperatures in the early spring and biomass partitioning to the root system. At ~35 DAE the stem began to elongate and biomass partitioning to the canopy and stem increase substantially. The increase in RUE at ~50 DAE was driven largely by an increase in the rate of dry biomass disposition in these tissues. Slightly before anthesis the RUE peaked at ~2.5 g/MJ.

The decrease in RUE at anthesis occurred due to a significant increase in intercepted PAR, while dry biomass accumulation remained unchanged, essentially doubling the quantity of radiation intercepted per unit of dry biomass accumulated. There are two hypotheses that may explain why biomass accumulation of the plants did not increase with increased radiation interception. The first hypothesis is that once canopy, stem, and panicle growth ceased at anthesis, the stem (which was the only organ capable of accumulating carbohydrates during that harvest interval) was not strong enough of a sink to fully utilize the photoassimilate generated by the canopy. In support of this, it is known that sink limitation can cause the accumulation of photoassimilate in the vasculature and leaves resulting in feedback inhibition of photosynthesis and down regulation of photosynthetic gene expression. An alternative hypothesis is that the plant encountered a temporary drought stress which is known to occur

during this point in the season in College Station, resulting in reduced rates of photosynthesis and underutilization of intercepted radiation. After anthesis Della allocated ~1 g of dry biomass to its stem for every MJ of light intercepted. Once Della entered into grain filling (a period of time highly sensitive to drought stress), the RUE increased to ~3 g/MJ driven largely by increasing biomass deposition in the panicle. This suggests that Della was not drought stressed during the interval of grain filling during the experiment. At this time the rate of dry biomass accumulation of the stem remained virtually unchanged. Since the rate of stem biomass accumulation at anthesis (when RUE was low) was nearly equal to stem dry biomass accumulation during grain filling (when RUE was ~3 g/MJ) it is difficult to conclude from this data that the decline in RUE was due to stress. Further experimentation is needed to test these hypotheses. Additionally, future research should consider that the timing of harvest and the resulting intervals, could potentially mask declines in RUE by averaging peaks and troughs in RUE. Therefore, more harvests, resulting in smaller harvest intervals would increase resolution.

The subsequent upswing in RUE was due to the increase in biomass deposition in the filling grain. As mentioned above, the RUE was ~1 g/MJ at anthesis when biomass was only deposited in the stem and at grain maturity the total RUE was ~3 g/MJ. This data shows that Della allocated ~2 g/MJ to the grain while only ~1 g/MJ to the stem suggesting that the stem is a weaker sink than the panicle. It seems plausible that the sink strength of the stem may be

tuned such that it does not compete for sugar with the reproductive organs of the plant under normal field conditions. However, for bioenergy production in which stem biomass is the desired product, a low stem sink strength is not optimal. Additionally, the production of grain from energy sorghums is not a high priority since seed production is separate from biomass production and therefore increasing competition between the stem and the grain is not an issue. This data suggests that transgenic manipulation of the stem to increase its sink strength may have a positive impact on the total NSC yield of the plant.

RUE declined to zero after grain maturity. During this interval, the total above ground dry biomass remained unchanged while stem biomass declined and apical tiller biomass increased. This reallocation may have in part been the result of remobilization of carbohydrate reserves from the stem to the tillers. Interestingly, Della was unable to maintain a constant stem dry weight even though it retained a significant percentage of its canopy after grain maturity. Ether apical tiller growth consumed more carbohydrates than were produced by the canopy at full photosynthetic capacity, or the plants were stressed and relied heavily on stored reserves, or photosynthetic rates were depressed, possibly due to the residual effects high tissue carbohydrate concentration. Additional experimentation is needed to determine the cause of the decline in RUE post grain maturity.

Data from the sweet*grain sorghum QTL mapping population RTX436*Della also indicates that sorghum is sink limited. A comparison of loci

that affect internode volume with loci that affect internode NSC yield revealed QTL co-alignment (Dw3 and Dw1) between these traits. This result suggests that internode size/volume impacts NSC yield. Additionally, NSC yield was correlated with increased internode volume. Taken together, these results suggest that in sorghum stem volume limits carbohydrate yield and that an increase in stem volume would increase NSC yield. This may seem like an obvious conclusion and in the case of cell wall polymer content per internode is it. For instance, is the volume of the internode increases two fold then naturally the cell wall polysaccharide content of the internode will as well. However, the same cannot be said of nonstructural carbohydrate content of the internode. It is conceivable that a plant with a very large stem volume and very low leaf area would be unable to completely fill the volume of the stem during the carbohydrate accumulation phase. These results show that in this population the stem fills up to a maximum nonstructural carbohydrate concentration with internode volume ultimately limiting yield. Considering that the relationship between stem volume and carbohydrate yield held true even in RILs that had the greatest stem volume, the impact of this relationship on final carbohydrate yield may be quite large.

Further increases to the maximum sucrose concentration of the stem may not be possible as has been observed in sugarcane (Wu and Birch 2007). Increasing the sink strength of the stem through transgenic approaches geared to increase the flux of carbohydrate into polysaccharides such as the cell wall

polysaccharides, (cellulose and xylan), mixed-linkage glucans, or starch may be a viable solution. To that end, the sorghum stem accumulates starch after grain maturity indicating that the genes necessary are present and active (Chapter II, Figure 2.5). Some attempts to increase the content of starch in plant tissues have been successful (Geigenberger *et al.* 2005, Geigenberger*, et al.* 2001, Regierer *et al.* 2002, Sonnewald and Kossmann 2013, Wu and Birch 2007). Therefore, the results of this study suggests that the sorghum stem would benefit from an increase in sink strength and starch is a good choice as an additional carbohydrate partition.

CHAPTER V

CONCLUSIONS

The focus of this research was directed at developing an understanding of stem gene expression and the biochemistry/genetics of nonstructural carbohydrate accumulation in the sweet sorghum stem. An analysis of dry biomass accumulation and composition characterized the biomass accumulation profile of the sweet sorghum Della as well as changes in its composition. A follow-on stem transcriptome analysis identified the pathways that underlie the changes in composition. Genes involved in the synthesis of cellulose, GAX, and lignin were up regulated during stem growth and then sharply down regulated thereafter. These genes among others are responsible for carbohydrate consumption during the early stem growth phase. A highly expressed sucrose synthase was identified that is potentially responsible for the hydrolysis of sucrose generating UDP-glucose to support cellulose synthesis during the vegetative growth. After the cessation of stem growth at anthesis, Della accumulated excess carbohydrates in its stem mainly in the form of sucrose. It appears that down regulation of *SbVIN1,* which occurred coincident with the cessation of stem growth, may play a significant role in the accumulation of sucrose during flowering.

The Della stem also accumulated significant quantities of starch in its stem post grain maturity. The ability of the stem to accumulate starch after anthesis will be advantageous to energy sorghum improvement efforts since

starch represent an alternative partition that currently is capable of storing large amounts of reduced carbon. Analysis of the stem starch biosynthesis and degradation pathways revealed the specific gene families responsible for these processes in the stem and points to potential engineering approaches. The gene glucan-water dikinase (GWD), which has been targeted in previous composition engineering studies, was highly expressed and could be targeted with RNAi under the control of a stem specific promoter expressed after anthesis to prevent the degradation of sorghum stem starch (Weise *et al.* 2012). Targeting the beta-amylase could proved to be an alternative approach.

Genetic analysis of carbohydrate content of the sorghum stem revealed that loci that regulate stem volume appear to impact sugar content. This result suggests that the stem becomes saturated with sugar. Lending further support to this hypothesis, total sugar content was correlated with traits related to stem volume but not stem juiciness and leaf area even though these traits were segregating. An analysis of the efficiency with which radiation was utilized showed that the RUE of the stem is less than the RUE of the panicle during grain filling and suggests that the sink strength of the stem may be low to prevent competition with filling of the grain.

The information presented here will facilitate sorghum improvement by providing detailed information regarding the accumulation of components of the stem composition such as the cell wall and the nonstructural carbohydrates such as starch. Additionally, this dissertation provides the candidate genes required

to modify those components, specifically starch but also gene families involved in cell wall biosynthesis and modification.

REFERENCES

Anders, N., Wilkinson, M.D., Lovegrove, A., Freeman, J., Tryfona, T., Pellny, T.K., Weimar, T., Mortimer, J.C., Stott, K., Baker, J.M., Defoin-Platel, M., Shewry, P.R., Dupree, P. and Mitchell, R.A. (2012) Glycosyl transferases in family 61 mediate arabinofuranosyl transfer onto xylan in grasses. *Proc Natl Acad Sci U S A*, **109**, 989-993.

Arundale, R.A., Bauer, S., Haffner, F.B., Mitchell, V.D., Voigt, T.B. and Long, S.P. (2015) Environment has little effect on biomass biochemical composition of Miscanthus× giganteus across soil types, nitrogen fertilization, and times of harvest. *BioEnergy Research*, **8**, 1636-1646.

Bartley, L.E., Peck, M.L., Kim, S.R., Ebert, B., Manisseri, C., Chiniquy, D.M., Sykes, R., Gao, L., Rautengarten, C., Vega-Sanchez, M.E., Benke, P.I., Canlas, P.E., Cao, P., Brewer, S., Lin, F., Smith, W.L., Zhang, X., Keasling, J.D., Jentoff, R.E., Foster, S.B., Zhou, J., Ziebell, A., An, G., Scheller, H.V. and Ronald, P.C. (2013) Overexpression of a BAHD acyltransferase, OsAt10, alters rice cell wall hydroxycinnamic acid content and saccharification. *Plant Physiol*, **161**, 1615-1633.

Berthet, S., Demont-Caulet, N., Pollet, B., Bidzinski, P., Cezard, L., Le Bris, P., Borrega, N., Herve, J., Blondet, E., Balzergue, S., Lapierre, C. and Jouanin, L. (2011) Disruption of LACCASE4 and 17 results in tissue-specific alterations to lignification of Arabidopsis thaliana stems. *Plant Cell*, **23**, 1124-1137.

Biswal, A.K., Hao, Z.Y., Pattathil, S., Yang, X.H., Winkeler, K., Collins, C., Mohanty, S.S., Richardson, E.A., Gelineo-Albersheim, I., Hunt, K., Ryno, D., Sykes, R.W., Turner, G.B., Ziebell, A., Gjersing, E., Lukowitz, W.G., Davis, M.F., Decker, S.R., Hahn, M.G. and Mohnen, D. (2015) Downregulation of GAUT12 in Populus deltoides by RNA silencing results in reduced recalcitrance, increased growth and reduced xylan and pectin in a woody biofuel feedstock. *Biotechnology for Biofuels*, **8**, 41.

Blennow, A., Nielsen, T.H., Baunsgaard, L., Mikkelsen, R. and Engelsen, S.B. (2002) Starch phosphorylation: a new front line in starch research. *Trends in Plant Science*, **7**, 445-450.

Boyle, M.G., Boyer, J.S. and Morgan, P.W. (1991) Stem infusion of liquid culture medium prevents reproductive failure of maize at low water potential. *Crop Sci*, **31**, 1246-1252.

Bromley, J.R., Busse-Wicher, M., Tryfona, T., Mortimer, J.C., Zhang, Z., Brown, D.M. and Dupree, P. (2013) GUX1 and GUX2 glucuronyltransferases decorate distinct domains of glucuronoxylan with different substitution patterns. *Plant J*, **74**, 423-434.

Burton, R.A. and Fincher, G.B. (2012) Current challenges in cell wall biology in the cereals and grasses. *Front Plant Sci*, **3**, 130.

Burton, R.A., Wilson, S.M., Hrmova, M., Harvey, A.J., Shirley, N.J., Medhurst, A., Stone, B.A., Newbigin, E.J., Bacic, A. and Fincher, G.B. (2006) Cellulose

synthase-like CslF genes mediate the synthesis of cell wall (1,3;1,4)-beta-D-glucans. *Science*, **311**, 1940-1942.

Byrt, C.S., Grof, C.P.L. and Furbank, R.T. (2011) C-4 Plants as biofuel feedstocks: optimising biomass production and feedstock quality from a lignocellulosic perspective. *J Integr Plant Biol*, **53**, 120-135.

Calvino, M., Bruggmann, R. and Messing, J. (2008) Screen of genes linked to high-sugar content in stems by comparative genomics. *Rice*, **1**, 166-176.

Calvino, M., Bruggmann, R. and Messing, J. (2011) Characterization of the small RNA component of the transcriptome from grain and sweet sorghum stems. *BMC Genomics*, **12**, 356.

Carpita, N.C. (1996) Structure and biogenesis of the cell walls of grasses. *Annu Rev Plant Physiol Plant Mol Biol*, **47**, 445-476.

Chen, X.Y. and Kim, J.Y. (2009) Callose synthesis in higher plants. *Plant Signal Behav*, **4**, 489-492.

Chiniquy, D., Sharma, V., Schultink, A., Baidoo, E.E., Rautengarten, C., Cheng, K., Carroll, A., Ulvskov, P., Harholt, J., Keasling, J.D., Pauly, M., Scheller, H.V. and Ronald, P.C. (2012) XAX1 from glycosyltransferase family 61 mediates xylosyltransfer to rice xylan. *Proc Natl Acad Sci U S A*, **109**, 17117-17122.

Critchley, J.H., Zeeman, S.C., Takaha, T., Smith, A.M. and Smith, S.M. (2001) A critical role for disproportionating enzyme in starch breakdown is revealed by a knock-out mutation in Arabidopsis. *The Plant Journal*, **26**, 89-100.

Dal-Bianco, M., Carneiro, M.S., Hotta, C.T., Chapola, R.G., Hoffmann, H.P., Garcia, A.A.F. and Souza, G.M. (2012) Sugarcane improvement: how far can we go? *Curr Opin Biotech*, **23**, 265-270.

Denyer, K., Dunlap, F., Thorbjornsen, T., Keeling, P. and Smith, A.M. (1996) The Major Form of ADP-glucose pyrophosphorylase in maize endosperm is extra-plastidial. *Plant Physiology*, **112**, 779-785.

Dien, B.S., Sarath, G., Pedersen, J.F., Sattler, S.E., Chen, H., Funnell-Harris, D.L., Nichols, N.N. and Cotta, M.A. (2009) Improved sugar conversion and ethanol yield for forage sorghum (Sorghum bicolor L. Moench) lines with reduced lignin contents. *Bioenergy Research*, **2**, 153-164.

Dionex (2003) The determination of sugars in molasses by high-performance anion exchange with pulsed amperometric detection. Dionex Coporation. Application note 92.

Duncan, R.R., Bockholt, A.J. and Miller, F.R. (1981) Descriptive comparison of senescent and non-senescent sorghum genotypes. *Agron J*, **73**, 849-853.

Ebert, B., Rautengarten, C., Guo, X.Y., Xiong, G.Y., Stonebloom, S., Smith-Moritz, A.M., Herter, T., Chan, L.J.G., Adams, P.D., Petzold, C.J., Pauly, M., Willats, W.G.T., Heazlewood, J.L. and Scheller, H.V. (2015) Identification and characterization of a golgi-localized UDP-xylose transporter family from arabidopsis. *Plant Cell*, **27**, 1218-1227.

Ermawar, R.A., Collins, H.M., Byrt, C.S., Betts, N.S., Henderson, M., Shirley, N.J., Schwerdt, J., Lahnstein, J., Fincher, G.B. and Burton, R.A. (2015)

Distribution, structure and biosynthetic gene families of (1,3;1,4)-beta-glucan in Sorghum bicolor. *J Integr Plant Biol*, **57**, 429-445.

Felderhoff, T.J., Murray, S.C., Klein, P.E., Sharma, A., Hamblin, M.T., Kresovich, S., Vermerris, W. and Rooney, W.L. (2012) QTLs for energy-related traits in a sweet x grain sorghum [Sorghum bicolor (L.) Moench] mapping population. *Crop Sci*, **52**, 2040-2049.

Ferraris, R. (1981) Early assessment of sweet sorghum as an agro-industrial crop. 2. Maturity factors. *Animal Production Science*, **21**, 83-90.

Geigenberger, P., Regierer, B., Nunes-Nesi, A., Leisse, A., Urbanczyk-Wochniak, E., Springer, F., van Dongen, J.T., Kossmann, J. and Fernie, A.R. (2005) Inhibition of de novo pyrimidine synthesis in growing potato tubers leads to a compensatory stimulation of the pyrimidine salvage pathway and a subsequent increase in biosynthetic performance. *The Plant Cell*, **17**, 2077-2088.

Geigenberger, P., Stamme, C., Tjaden, J., Schulz, A., Quick, P.W., Betsche, T., Kersting, H.J. and Neuhaus, H.E. (2001) Tuber physiology and properties of starch from tubers of transgenic potato plants with altered plastidic adenylate transporter activity. *Plant Physiology*, **125**, 1667-1678.

Gill, J.R., Burks, P.S., Staggenborg, S.A., Odvody, G.N., Heiniger, R.W., Macoon, B., Moore, K.J., Barrett, M. and Rooney, W.L. (2014) Yield results and stability analysis from the sorghum regional biomass feedstock trial. *Bioenergy Research*, **7**, 1026-1034.

Gille, S., de Souza, A., Xiong, G., Benz, M., Cheng, K., Schultink, A., Reca, I.B. and Pauly, M. (2011) O-acetylation of Arabidopsis hemicellulose xyloglucan requires AXY4 or AXY4L, proteins with a TBL and DUF231 domain. *Plant Cell*, **23**, 4041-4053.

Gnansounou, E. and Dauriat, A. (2010) Techno-economic analysis of lignocellulosic ethanol: A review. *Bioresour Technol*, **101**, 4980-4991.

Godoy, J.G.V. and Tesso, T.T. (2013) Analysis of juice yield, sugar content, and biomass accumulation in sorghum. *Crop Sci*, **53**, 1288-1297.

Goubet, F., Barton, C.J., Mortimer, J.C., Yu, X., Zhang, Z., Miles, G.P., Richens, J., Liepman, A.H., Seffen, K. and Dupree, P. (2009) Cell wall glucomannan in Arabidopsis is synthesised by CSLA glycosyltransferases, and influences the progression of embryogenesis. *Plant J*, **60**, 527-538.

Gutjahr, S., Clement-Vidal, A., Soutiras, A., Sonderegger, N., Braconnier, S., Dingkuhn, M. and Luquet, D. (2013a) Grain, sugar and biomass accumulation in photoperiod-sensitive sorghums. II. Biochemical processes at internode level and interaction with phenology. *Funct Plant Biol*, **40**, 355-368.

Gutjahr, S., Vaksmann, M., Dingkuhn, M., Thera, K., Trouche, G., Braconnier, S. and Luquet, D. (2013b) Grain, sugar and biomass accumulation in tropical sorghums. I. Trade-offs and effects of phenological plasticity. *Funct Plant Biol*, **40**, 342-354.

Hädrich, N., Gibon, Y., Schudoma, C., Altmann, T., Lunn, J.E. and Stitt, M. (2011) Use of TILLING and robotised enzyme assays to generate an allelic series of

Arabidopsis thaliana mutants with altered ADP-glucose pyrophosphorylase activity. *Journal of Plant Physiology*, **168**, 1395-1405.

Harrison, R.L. and Miller, F.R. (1993) Registration of Della Sweet Sorghum. *Crop Sci*, **33**, 1416-1416.

Hatfield, R.D., Wilson, J.R. and Mertens, D.R. (1999) Composition of cell walls isolated from cell types of grain sorghum stems. *J Sci Food Agr*, **79**, 891-899.

Hejazi, M., Fettke, J., Haebel, S., Edner, C., Paris, O., Frohberg, C., Steup, M. and Ritte, G. (2008) Glucan, water dikinase phosphorylates crystalline maltodextrins and thereby initiates solubilization. *The Plant Journal*, **55**, 323-334.

Hill, J., Nelson, E., Tilman, D., Polasky, S. and Tiffany, D. (2006) Environmental, economic, and energetic costs and benefits of biodiesel and ethanol biofuels. *P Natl Acad Sci USA*, **103**, 11206-11210.

Hirano, K., Aya, K., Morinaka, Y., Nagamatsu, S., Sato, Y., Antonio, B.A., Namiki, N., Nagamura, Y. and Matsuoka, M. (2013) Survey of genes involved in rice secondary cell wall formation through a co-expression network. *Plant Cell Physiol*, **54**, 1803-1821.

HoffmannThoma, G., Hinkel, K., Nicolay, P. and Willenbrink, J. (1996) Sucrose accumulation in sweet sorghum stem internodes in relation to growth. *Physiol Plantarum*, **97**, 277-284.

Hrmova, M. and Fincher, G.B. (2001) Structure-function relationships of beta-D-glucan endo- and exohydrolases from higher plants. *Plant Mol Biol*, **47**, 73-91.

Jackson, D.R., Arthur, M.F., Davis, M., Kresovich, S., Lawhon, W.T., Lipinsky, E.S., Price, M. and Rudolph, A. (1980) Research report on development of sweet sorghum as an energy crop Volume I: Agricultural task. USDOE, Battele, Columbus: U.S. Department of Commerce.

Jiang, S.Y., Ma, Z., Vanitha, J. and Ramachandran, S. (2013) Genetic variation and expression diversity between grain and sweet sorghum lines. *BMC Genomics*, **14**, 18.

Jung, H.G. and Phillips, R.L. (2010) Putative Seedling Ferulate Ester (sfe) Maize Mutant: Morphology, Biomass Yield, and Stover Cell Wall Composition and Rumen Degradability. *Crop Sci*, **50**, 403-418.

Jung, H.J.G., Samac, D.A. and Sarath, G. (2012) Modifying crops to increase cell wall digestibility. *Plant Sci*, **185**, 65-77.

Knoller, A.S., Blakeslee, J.J., Richards, E.L., Peer, W.A. and Murphy, A.S. (2010) Brachytic2/ZmABCB1 functions in IAA export from intercalary meristems. *J Exp Bot*, **61**, 3689-3696.

Lee, C., Teng, Q., Zhong, R. and Ye, Z.H. (2011) The four Arabidopsis reduced wall acetylation genes are expressed in secondary wall-containing cells and required for the acetylation of xylan. *Plant Cell Physiol*, **52**, 1289-1301.

Li, M., Xiong, G., Li, R., Cui, J., Tang, D., Zhang, B., Pauly, M., Cheng, Z. and Zhou, Y. (2009) Rice cellulose synthase-like D4 is essential for normal cell-wall biosynthesis and plant growth. *Plant J*, **60**, 1055-1069.

Li, X., Weng, J.K. and Chapple, C. (2008) Improvement of biomass through lignin modification. *Plant J*, **54**, 569-581.

Liepman, A.H. and Cavalier, D.M. (2012) The cellulose synthase-like A and cellulose synthase-like C families: recent advances and future perspectives. *Front Plant Sci*, **3**, 109.

Lingle, S.E. (1987) Sucrose metabolism in the primary culm of sweet sorghum during development. *Crop Sci*, **27**, 1214-1219.

Lingle, S.E. (1989) Evidence for the uptake of sucrose intact into sugarcane internodes. *Plant Physiology*, **90**, 6-8.

Lionetti, V., Francocci, F., Ferrari, S., Volpi, C., Bellincampi, D., Galletti, R., D'Ovidio, R., De Lorenzo, G. and Cervone, F. (2010) Engineering the cell wall by reducing de-methyl-esterified homogalacturonan improves saccharification of plant tissues for bioconversion. *Proceedings of the National Academy of Sciences*, **107**, 616-621.

Lorberth, R., Ritte, G., Willmitzer, L. and Kossmann, J. (1998) Inhibition of a starch-granule-bound protein leads to modified starch and repression of cold sweetening. *Nat Biotech*, **16**, 473-477.

Martin, A.P., Palmer, W.M., Byrt, C.S., Furbank, R.T. and Grof, C.P.L. (2013) A holistic high-throughput screening framework for biofuel feedstock assessment that characterises variations in soluble sugars and cell wall composition in Sorghum bicolor. *Biotechnology for Biofuels*, **6**.

McBee, G., Waskom, R. and Creelman, R. (1983) Effect of Senescene on Carbohydrates in sorghum during late kernel maturity states. *Crop Sci*, **23**, 372-376.

Mcbee, G.G. and Miller, F.R. (1982) Carbohydrates in sorghum culms as influenced by cultivars, spacing, and maturity over a diurnal period. *Crop Sci*, **22**, 381-385.

McCann, M.C. and Carpita, N.C. (2015) Biomass recalcitrance: a multi-scale, multi-factor, and conversion-specific property. *J Exp Bot*, **66**, 4109-4118.

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

McQueen-Mason, S.J. and Cosgrove, D.J. (1995) Expansin mode of action on cell walls. Analysis of wall hydrolysis, stress relaxation, and binding. *Plant Physiol*, **107**, 87-100.

Mitchell, R.A., Dupree, P. and Shewry, P.R. (2007) A novel bioinformatics approach identifies candidate genes for the synthesis and feruloylation of arabinoxylan. *Plant Physiol*, **144**, 43-53.

Monk, R.L., Miller, F.R. and Mcbee, G.G. (1984) Sorghum improvement for energy-production. *Biomass*, **6**, 145-153.

Moore, P.H. (1995) Temporal and spatial regulation of sucrose accumulation in the sugarcane stem. *Aust J Plant Physiol*, **22**, 661-679.

Mullet, J., Morishige, D., McCormick, R., Truong, S., Hilley, J., McKinley, B., Anderson, R., Olson, S. and Rooney, W. (2014) Energy Sorghum—a genetic model for the design of C4 grass bioenergy crops. *J Exp Bot*, eru229.

Murphy, R.L., Klein, R.R., Morishige, D.T., Brady, J.A., Rooney, W.L., Miller, F.R., Dugas, D.V., Klein, P.E. and Mullet, J.E. (2011) Coincident light and clock regulation of pseudoresponse regulator protein 37 (PRR37) controls photoperiodic flowering in sorghum. *P Natl Acad Sci USA*, **108**, 16469-16474.

Murphy, R.L., Morishige, D.T., Brady, J.A., Rooney, W.L., Yang, S., Klein, P.E. and Mullet, J.E. (2014) () Represses sorghum flowering in long days: Alleles enhance biomass accumulation and grain production. *The Plant Genome*, **7**.

Murray, S.C., Rooney, W.L., Hamblin, M.T., Mitchell, S.E. and Kresovich, S. (2009) Sweet sorghum genetic diversity and association mapping for brix and height. *Plant Genome-Us*, **2**, 48-62.

Murray, S.C., Sharma, A., Rooney, W.L., Klein, P.E., Mullet, J.E., Mitchell, S.E. and Kresovich, S. (2008) Genetic improvement of sorghum as a biofuel feedstock: I. QTL for stem sugar and grain nonstructural carbohydrates. *Crop Sci*, **48**, 2165-2179.

Olson, S.N., Ritter, K., Medley, J., Wilson, T., Rooney, W.L. and Mullet, J.E. (2013) Energy sorghum hybrids: Functional dynamics of high nitrogen use efficiency. *Biomass and Bioenergy*, **56**, 307-316.

Olson, S.N., Ritter, K., Rooney, W., Kemanian, A., McCarl, B.A., Zhang, Y., Hall, S., Packer, D. and Mullet, J. (2012) High biomass yield energy sorghum:

developing a genetic model for C4 grass bioenergy crops. *Biofuels, Bioproducts and Biorefining*, **6**, 640-655.

Park, Y.B. and Cosgrove, D.J. (2015) Xyloglucan and its interactions with other components of the growing cell wall. *Plant Cell Physiol*, **56**, 180-194.

Paterson, A.H., Bowers, J.E., Bruggmann, R., Dubchak, I., Grimwood, J., Gundlach, H., Haberer, G., Hellsten, U., Mitros, T., Poliakov, A., Schmutz, J., Spannagl, M., Tang, H.B., Wang, X.Y., Wicker, T., Bharti, A.K., Chapman, J., Feltus, F.A., Gowik, U., Grigoriev, I.V., Lyons, E., Maher, C.A., Martis, M., Narechania, A., Otillar, R.P., Penning, B.W., Salamov, A.A., Wang, Y., Zhang, L.F., Carpita, N.C., Freeling, M., Gingle, A.R., Hash, C.T., Keller, B., Klein, P., Kresovich, S., McCann, M.C., Ming, R., Peterson, D.G., Mehboob-ur-Rahman, Ware, D., Westhoff, P., Mayer, K.F.X., Messing, J. and Rokhsar, D.S. (2009) The Sorghum bicolor genome and the diversification of grasses. *Nature*, **457**, 551-556.

Pauly, M. and Keegstra, K. (2008) Cell-wall carbohydrates and their modification as a resource for biofuels. *Plant Journal*, **54**, 559-568.

Penning, B.W., Hunter, C.T., 3rd, Tayengwa, R., Eveland, A.L., Dugard, C.K., Olek, A.T., Vermerris, W., Koch, K.E., McCarty, D.R., Davis, M.F., Thomas, S.R., McCann, M.C. and Carpita, N.C. (2009) Genetic resources for maize cell wall biology. *Plant Physiol*, **151**, 1703-1728.

Perlack, D. and Stokes, B.J. (2011) U.S. Billion-ton update: Biomass supply for a bioproducts indistry.

Persson, S., Paredez, A., Carroll, A., Palsdottir, H., Doblin, M., Poindexter, P., Khitrov, N., Auer, M. and Somerville, C.R. (2007) Genetic evidence for three unique components in primary cell-wall cellulose synthase complexes in Arabidopsis. *P Natl Acad Sci USA*, **104**, 15566-15571.

Qazi, H.A., Paranjpe, S. and Bhargava, S. (2012) Stem sugar accumulation in sweet sorghum - activity and expression of sucrose metabolizing enzymes and sucrose transporters. *J Plant Physiol*, **169**, 605-613.

Quinby, J.R. (1974) Sorghum improvement and the genetics of growth.

Rahall, N. (2007) H. R. 6 (100th): Energy Independence and Security Act of 2007.

Ranum, P., Peña‐Rosas, J.P. and Garcia‐Casal, M.N. (2014) Global maize production, utilization, and consumption. *Annals of the New York Academy of Sciences*, **1312**, 105-112.

Regierer, B., Fernie, A.R., Springer, F., Perez-Melis, A., Leisse, A., Koehl, K., Willmitzer, L., Geigenberger, P. and Kossmann, J. (2002) Starch content and yield increase as a result of altering adenylate pools in transgenic plants. *Nat Biotechnol*, **20**, 1256-1260.

Rennie, E.A., Hansen, S.F., Baidoo, E.E., Hadi, M.Z., Keasling, J.D. and Scheller, H.V. (2012) Three members of the Arabidopsis glycosyltransferase family 8 are xylan glucuronosyltransferases. *Plant Physiol*, **159**, 1408-1417.

Rennie, E.A. and Scheller, H.V. (2014) Xylan biosynthesis. *Curr Opin Biotechnol*, **26**, 100-107.

Ritte, G., Heydenreich, M., Mahlow, S., Haebel, S., Kötting, O. and Steup, M. (2006) Phosphorylation of C6-and C3-positions of glucosyl residues in starch is catalysed by distinct dikinases. *FEBS letters*, **580**, 4872-4876.

Ritter, K.B., Jordan, D.R., Chapman, S.C., Godwin, I.D., Mace, E.S. and McIntyre, C.L. (2008) Identification of QTL for sugar-related traits in a sweet x grain sorghum (Sorghum bicolor L. Moench) recombinant inbred population. *Mol Breeding*, **22**, 367-384.

Rooney, W.L. (2004) Sorghum improvement—integrating traditional and new technology to produce improved genotypes. *Advances in Agronomy*, **83**, 37-109.

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

Rooney, W.L., Blumenthal, J., Bean, B. and Mullet, J.E. (2007) Designing sorghum as a dedicated bioenergy feedstock. *Biofuel Bioprod Bior*, **1**, 147-157.

Roudier, F., Fernandez, A.G., Fujita, M., Himmelspach, R., Borner, G.H.H., Schindelman, G., Song, S., Baskin, T.I., Dupree, P., Wasteneys, G.O. and Benfey, P.N. (2005) COBRA, an Arabidopsis extracellular glycosyl-phosphatidyl inositol-anchored protein, specifically controls highly anisotropic expansion through its involvement in cellulose microfibril orientation. *The Plant Cell*, **17**, 1749-1763.

Ruan, Y.L., Jin, Y., Yang, Y.J., Li, G.J. and Boyer, J.S. (2010) Sugar input, metabolism, and signaling mediated by invertase: roles in development, yield potential, and response to drought and heat. *Mol Plant*, **3**, 942-955.

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

Sage, R.F. and Zhu, X.G. (2011) Exploiting the engine of C(4) photosynthesis. *J Exp Bot*, **62**, 2989-3000.

Sattler, S.E., Palmer, N.A., Saballos, A., Greene, A.M., Xin, Z.G., Sarath, G., Vermerris, W. and Pedersen, J.F. (2012) Identification and characterization of four missense mutations in Brown midrib 12 (Bmr12), the Caffeic O-Methyltranferase (COMT) of Sorghum. *Bioenergy Research*, **5**, 855-865.

Sattler, S.E., Saathoff, A.J., Haas, E.J., Palmer, N.A., Funnell-Harris, D.L., Sarath, G. and Pedersen, J.F. (2009) A nonsense mutation in a cinnamyl alcohol dehydrogenase gene is responsible for the sorghum Brown midrib 6 phenotype. *Plant Physiology*, **150**, 584-595.

Shannon, J.C., Pien, F.-M., Cao, H. and Liu, K.-C. (1998) Brittle-1, an adenylate translocator, facilitates transfer of extraplastidial synthesized ADP-glucose into amyloplasts of maize endosperms. *Plant Physiology*, **117**, 1235-1252.

Shen, H., Mazarei, M., Hisano, H., Escamilla-Trevino, L., Fu, C., Pu, Y., Rudis, M.R., Tang, Y., Xiao, X. and Jackson, L. (2013) A genomics approach to deciphering lignin biosynthesis in switchgrass. *The Plant Cell*, **25**, 4342-4361.

Shiringani, A.L. and Friedt, W. (2011) QTL for fibre-related traits in grain× sweet sorghum as a tool for the enhancement of sorghum as a biomass crop. *Theor Appl Genet*, **123**, 999-1011.

Slewinski, T.L. (2012) Non-structural carbohydrate partitioning in grass stems: a target to increase yield stability, stress tolerance, and biofuel production. *J Exp Bot*, **63**, 4647-4670.

Sluiter, A. and Sluiter, J. (2005) Determination of starch in solid biomass samples by HPLC. *NREL Laboratory Analytical Procedure.* < http://devafdc*. nrel. gov/pdfs/9360. pdf* >.

Smith, A.M. (2008) Prospects for increasing starch and sucrose yields for bioethanol production. *The Plant Journal*, **54**, 546-558.

Somerville, C., Youngs, H., Taylor, C., Davis, S.C. and Long, S.P. (2010) Feedstocks for lignocellulosic biofuels. *Science*, **329**, 790-792.

Sonnewald, U. and Kossmann, J. (2013) Starches—from current models to genetic engineering. *Plant Biotechnol J*, **11**, 223-232.

Stark, D.M., Timmerman, K.P., Barry, G.F., Preiss, J. and Kishore, G.M. (1992) Regulation of the amount of starch in plant tissues by ADP glucose pyrophosphorylase. *Science*, **258**, 287-292.

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

Stitt, M. and Zeeman, S.C. (2012) Starch turnover: pathways, regulation and role in growth. *Current Opinion in Plant Biology*, **15**, 282-292.

Sweetlove, L.J. and Burrell, M.M. (1996) Starch metabolism in tubers of transgenic potato (Solanum tuberosum) with increased ADP-glucose pyrophosphorylase. *Biochem J*, **320**, 493-498.

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

Tarpley, L., Lingle, S.E., Vietor, D.M., Andrews, D.L. and Miller, F.R. (1994a) Enzymatic control of nonstructural carbohydrate concentrations in stems and panicles of sorghum. *Crop Sci*, **34**, 446-452.

Tarpley, L. and Vietor, D.M. (2007) Compartmentation of sucrose during radial transfer in mature sorghum culm. *BMC Plant Biology*, **7**, 33.

Tarpley, L., Vietor, D.M. and Miller, F.R. (1994b) Internodal compartmentation of stem-infused [C-14] sucrose in sweet and grain-sorghum. *Crop Sci*, **34**, 1116-1120.

Taylor-Teeples, M., Lin, L., de Lucas, M., Turco, G., Toal, T.W., Gaudinier, A., Young, N.F., Trabucco, G.M., Veling, M.T., Lamothe, R., Handakumbura, P.P., Xiong, G., Wang, C., Corwin, J., Tsoukalas, A., Zhang, L., Ware, D., Pauly, M., Kliebenstein, D.J., Dehesh, K., Tagkopoulos, I., Breton, G., Pruneda-Paz, J.L., Ahnert, S.E., Kay, S.A., Hazen, S.P. and Brady, S.M.

(2015) An Arabidopsis gene regulatory network for secondary cell wall synthesis. *Nature*, **517**, 571-575.

Tetlow, I.J. and Emes, M.J. (2014) A review of starch-branching enzymes and their role in amylopectin biosynthesis. *IUBMB life*, **66**, 546-558.

Vain, T., Crowell, E.F., Timpano, H., Biot, E., Desprez, T., Mansoori, N., Trindade, L.M., Pagant, S., Robert, S., Hofte, H., Gonneau, M. and Vernhettes, S. (2014) The Cellulase KORRIGAN is part of the cellulose synthase complex. *Plant Physiology*, **165**, 1521-1532.

van der Weijde, T., Kamei, C.L.A., Torres, A.F., Vermerris, W., Dolstra, O., Visser, R.G.F. and Trindade, L.M. (2013) The potential of C4 grasses for cellulosic biofuel production. *Frontiers in Plant Science*, **4**, 107.

Vanderlip, R.L. and Reeves, H.E. (1972) Growth stages of sorghum [Sorghum bicolor,(L.) Moench.]. *Agron J*, **64**, 13-16.

Vanholme, R., Demedts, B., Morreel, K., Ralph, J. and Boerjan, W. (2010) Lignin biosynthesis and structure. *Plant Physiol*, **153**, 895-905.

Vega-Sanchez, M.E., Loque, D., Lao, J.M., Catena, M., Verhertbruggen, Y., Herter, T., Yang, F., Harholt, J., Ebert, B., Baidoo, E.E.K., Keasling, J.D., Scheller, H.V., Heazlewood, J.L. and Ronald, P.C. (2015) Engineering temporal accumulation of a low recalcitrance polysaccharide leads to increased C6 sugar content in plant cell walls. *Plant Biotechnol J*, **13**, 903-914.

Vega-Sánchez, M.E., Verhertbruggen, Y., Christensen, U., Chen, X., Sharma, V., Varanasi, P., Jobling, S.A., Talbot, M., White, R.G. and Joo, M. (2012) Loss

of Cellulose synthase-like F6 function affects mixed-linkage glucan deposition, cell wall mechanical properties, and defense responses in vegetative tissues of rice. *Plant Physiology*, **159**, 56-69.

Vega-Sanchez, M.E., Verhertbruggen, Y., Scheller, H.V. and Ronald, P.C. (2013) Abundance of mixed linkage glucan in mature tissues and secondary cell walls of grasses. *Plant Signal Behav*, **8**, e23143.

Verhertbruggen, Y., Yin, L., Oikawa, A. and Scheller, H.V. (2011) Mannan synthase activity in the CSLD family. *Plant signaling & behavior*, **6**, 1620-1623.

Vermerris, W. (2011) Survey of genomics approaches to improve bioenergy traits in maize, sorghum and sugarcane. *J Integr Plant Biol*, **53**, 105-119.

Vermerris, W., Thompson, K.J. and McIntyre, L.M. (2002) The maize Brown midrib1 locus affects cell wall composition and plant development in a dose-dependent manner. *Heredity (Edinb)*, **88**, 450-457.

Waclawovsky, A.J., Sato, P.M., Lembke, C.G., Moore, P.H. and Souza, G.M. (2010) Sugarcane for bioenergy production: an assessment of yield and regulation of sucrose content. *Plant Biotechnol J*, **8**, 263-276.

Wang, M.L., Zhu, C.S., Barkley, N.A., Chen, Z.B., Erpelding, J.E., Murray, S.C., Tuinstra, M.R., Tesso, T., Pederson, G.A. and Yu, J.M. (2009) Genetic diversity and population structure analysis of accessions in the US historic sweet sorghum collection. *Theor Appl Genet*, **120**, 13-23.

Weise, S.E., Aung, K., Jarou, Z.J., Mehrshahi, P., Li, Z.R., Hardy, A.C., Carr, D.J. and Sharkey, T.D. (2012) Engineering starch accumulation by manipulation of phosphate metabolism of starch. *Plant Biotechnol J*, **10**, 545-554.

Wilkerson, C.G., Mansfield, S.D., Lu, F., Withers, S., Park, J.Y., Karlen, S.D., Gonzales-Vigil, E., Padmakshan, D., Unda, F., Rencoret, J. and Ralph, J. (2014) Monolignol ferulate transferase introduces chemically labile linkages into the lignin backbone. *Science*, **344**, 90-93.

Withers, S., Lu, F., Kim, H., Zhu, Y., Ralph, J. and Wilkerson, C.G. (2012) Identification of grass-specific enzyme that acylates monolignols with p-coumarate. *J Biol Chem*, **287**, 8347-8355.

Wolfrum, E., Payne, C., Stefaniak, T., Rooney, W., Dighe, N., Bean, B. and Dahlberg, J. (2013) Multivariate calibration models for sorghum composition using near-infrared spectroscopy. U.S. Department of Energy, Office of Energy Efficiency & Renewable Energy: National Renewable Energy Laboratory, pp. 1-8.

Wu, L. and Birch, R.G. (2007) Doubled sugar content in sugarcane plants modified to produce a sucrose isomer. *Plant Biotechnol J*, **5**, 109-117.

Wu, X.R., Staggenborg, S., Propheter, J.L., Rooney, W.L., Yu, J.M. and Wang, D.H. (2010) Features of sweet sorghum juice and their performance in ethanol fermentation. *Ind Crop Prod*, **31**, 164-170.

Yang, S., Murphy, R.L., Morishige, D.T., Klein, P.E., Rooney, W.L. and Mullet, J.E. (2014a) Sorghum phytochrome B inhibits flowering in long days by activating

expression of SbPRR37 and SbGHD7, repressors of SbEHD1, SbCN8 and SbCN12.

Yang, S., Weers, B.D., Morishige, D.T. and Mullet, J.E. (2014b) CONSTANS is a photoperiod regulated activator of flowering in sorghum. *BMC Plant Biol*, **14**, 148.

Zhang, B., Zhao, T., Yu, W., Kuang, B., Yao, Y., Liu, T., Chen, X., Zhang, W. and Wu, A.M. (2014) Functional conservation of the glycosyltransferase gene GT47A in the monocot rice. *J Plant Res*, **127**, 423-432.

Zhu, Y.J., Komor, E. and Moore, P.H. (1997) Sucrose accumulation in the sugarcane stem is regulated by the difference between the activities of soluble acid invertase and sucrose phosphate synthase. *Plant Physiology*, **115**, 609-616.

APPENDIX

Table A.1. Sorghum genes involved in cell expansion

Family/ID	Function	Source	A-30	A-16	A-7	A	A+11	A+25	A+43	A+68	FC	FDR
α-Expansins												
Sobic.009G173700	α-expansin	[1]	145	8	8	1	9	11	0	1	>145	0.E+00
Sobic.006G191700	α-expansin	[1]	39	6	16	5	16	14	10	7	8	0.E+00
Sobic.001G033300	α-expansin	[1]	20	1	4	1	1	2	1	3	>20	0.E+00
Sobic.006G031900	α-expansin	[1]	12	0	0	0	0	0	0	0	>12	0.E+00
α-like Expansins												
Sobic.001G311000	α-like expansin	[1]	54	11	25	32	7	9	12	7	9	0.E+00
β-Expansins												
Sobic.001G300400	β-expansin	[1]	180	10	10	2	8	21	1	1	>180	0.E+00
Sobic.001G300800	β-expansin	[1]	168	6	10	0	5	4	0	1	>168	0.E+00
Sobic.001G301600	β-expansin	[1]	53	4	10	1	4	2	0	0	>53	0.E+00
Sobic.001G301500	β-expansin	[1]	37	2	10	0	3	1	0	0	>37	0.E+00
Sobic.001G300500	β-expansin	[1]	28	3	7	1	4	5	0	0	>28	0.E+00
Sobic.001G300700	β-expansin	[1]	24	1	2	0	1	3	0	0	>24	0.E+00
Sobic.002G124400	β-expansin	[1]	17	0	1	0	1	0	0	0	>17	0.E+00
Sobic.004G191600	β-expansin	[1]	15	5	5	1	3	5	3	2	13	0.E+00

Table A.2. Sorghum genes encoding cell wall proteins

Family/ID	Clade/Gene	Function	Source	A-30	A-16	A-7	A	A+11	A+25	A+43	A+68	FC	FDR
Arabinogalactan-proteins (AGPs)													
Sobic.003G062000		FASCICLIN-like arabinogalactan-protein	[1]	346	140	308	17	6	59	5	2	190	0.E+00
Sobic.003G250700		FASCICLIN-like arabinogalactan-protein	[1]	280	28	41	21	11	6	3	2	129	0.E+00
Sobic.009G055900		FASCICLIN-like arabinogalactan-protein	[1]	186	44	34	18	77	50	22	14	14	0.E+00
Sobic.009G056400		FASCICLIN-like arabinogalactan-protein	[1]	139	53	39	2	0	4	0	1	>139	0.E+00
Sobic.002G154600		FASCICLIN-like arabinogalactan-protein	[1]	44	18	34	2	14	5	1	0	>44	0.E+00
Sobic.K024700		FASCICLIN-like arabinogalactan-protein	[1]	38	19	24	5	12	11	7	4	10	0.E+00
Proline-rich repeat Extensins													
Sobic.007G096700	*OsPOEI30*	proline-rich extensin	[2]	182	84	89	54	87	101	95	83	4	0.E+00
Sobic.K008000	*OsPOEI24*	proline-rich extensin	[2]	85	36	42	9	47	31	16	7	13	0.E+00
Sobic.001G266100	*OsPOEI14*	proline-rich extensin	[2]	33	0	0	0	0	0	0	0	>33	0.E+00
Sobic.001G266300	*OsPOEI13*	proline-rich extensin	[2]	9	0	0	0	0	1	0	0	>9	0.E+00

Table A.3. Sorghum genes involved in Cellulose, MLG., mannan, and glucomannan synthesis

Family/ID	Clade/Gene	Function	Source	A-30	A-16	A-7	A	A+11	A+25	A+43	A+68	FC	FDR
CesA													
Sobic.009G063400	*OsCESA1*	cellulose synthase	[1]	172	280	186	61	248	157	69	35	9	0.E+00
Sobic.002G118700	*OsCESA3*	cellulose synthase	[1]	277	494	253	95	450	310	129	30	17	0.E+00
Sobic.002G075500	*OsCESA8*	cellulose synthase	[1]	218	310	223	45	288	187	60	29	11	0.E+00
Sobic.001G021500	*OsCESA5*	cellulose synthase	[1]	191	91	94	79	103	99	92	50	4	0.E+00
Sobic.002G094600	*OsCESA6*	cellulose synthase	[1]	154	120	146	76	26	45	16	6	26	0.E+00
Sobic.003G296400	*OsCESA4*	cellulose synthase	[1]	183	260	287	33	7	48	8	1	229	0.E+00
Sobic.001G224300	*OsCESA7*	cellulose synthase	[1]	135	193	200	20	5	20	4	1	167	0.E+00
Sobic.002G205500	*OsCESA9*	cellulose synthase	[1]	154	198	221	22	5	25	4	1	203	0.E+00
COBRA													
Sobic.001G086200		COBRA-like protein	[1]	442	479	494	340	436	321	224	118	5	0.E+00
Sobic.001G336700		COBRA-like protein	[1]	47	75	90	13	9	26	6	3	36	0.E+00
Sobic.001G399900		COBRA-like protein	[1]	58	7	44	27	12	10	12	4	16	0.E+00
Cellulose-like synthases													
Sobic.004G075900	*OsCSLA1*	mannan, glucomannan synthase	[1]	28	6	3	0	5	7	1	0	>28	0.E+00
Sobic.002G385800	*OsCSLA7*	mannan, glucomannan synthase	[1]	22	5	5	5	3	7	4	4	8	0.E+00
Sobic.004G238700	*OsCSLA6*	mannan, glucomannan synthase	[1]	7	5	3	2	1	2	0	0	>7	0.E+00
Sobic.001G490000	*OsCSLA4*	mannan, glucomannan synthase	[1]	6	3	2	1	2	4	1	2	5	2.E-02
Sobic.007G137400	*OsCSLA11*	cellulose synthase-like A11	[1]	24	123	95	106	76	38	44	36	5	2.E-02
Sobic.010G008600	*OsCSLD2*	mannan synthase	[1]	65	41	84	72	27	28	36	30	4	1.E-03
Sobic.007G050600	*OsCSLF6*	beta1,3;1,4 glucan synthase	[1]	215	188	318	241	164	146	121	62	6	0.E+00
Sobic.002G238300	*OsCSLE6*	cellulose synthase-like E6	[1]	1	1	1	1	3	3	4	6	>6	3.E-02
Sobic.003G442500	*OsCSLG2*	cellulose synthase-like G2	[1]	0	0	0	0	0	2	1	8	>8	3.E-03

Table A.4. Sorghum genes involved in callose metabolism

Family/ID	Clade/Gene	Function	Source	A-30	A-16	A-7	A	A+11	A+25	A+43	A+68	FC	FDR
Callose (beta 1-3 glucan) biosynthesis													
Sobic.001G521500		Callose Synthase	[1]	16	58	40	46	50	65	66	47	3	0.E+00
Sobic.001G542500		Callose Synthase	[1]	14	40	27	32	57	53	45	30	3	0.E+00
Sobic.001G529600		Callose Synthase	[1]	14	39	28	33	51	48	41	25	3	0.E+00
Glucan endo-1,3-beta-glucosidase (Glycoside hydrolase family 17) Callose Degradation													
Sobic.009G183400	A	glycosyl hydrolases family 17	[1]	63	17	5	6	19	10	6	28	13	0.E+00
Sobic.009G119200	A	glycosyl hydrolases family 17	[1]	27	1	0	0	0	0	0	1	>27	0.E+00
Sobic.001G061900	B	glucan endo-1,3-beta-glucosidase	[1]	75	19	36	33	21	20	25	9	9	0.E+00
Sobic.002G045800	B	glucan endo-1,3-beta-glucosidase	[1]	25	8	5	4	3	2	2	2	16	0.E+00
Sobic.001G142200	C	glucan endo-1,3-beta-glucosidase	[1]	97	20	14	18	29	23	9	10	13	0.E+00
Sobic.001G263500	C	glycosyl hydrolases family 17	[1]	23	8	8	9	4	4	6	5	7	0.E+00
Sobic.004G313200	D	glucan endo-1,3-beta-glucosidase	[1]	21	5	8	7	5	5	1	2	21	0.E+00
Sobic.002G328300	D	glucan endo-1,3-beta-glucosidase	[1]	28	20	16	9	6	3	4	3	12	0.E+00
Sobic.001G109400	C	glycosyl hydrolases family 17	[1]	36	18	30	32	28	22	13	20	3	0.E+00
Sobic.001G445700	C	glycosyl hydrolases family 17	[1]	7	7	23	12	4	3	4	3	8	0.E+00
Sobic.004G165600	C	glycosyl hydrolases family 17	[1]	42	35	117	53	9	7	9	5	23	0.E+00
Sobic.006G069200	C	glycosyl hydrolases family 17	[1]	127	29	126	198	25	19	29	8	26	0.E+00
Sobic.003G423500	A	glycosyl hydrolases family 17	[1]	0	0	0	1	1	4	6	20	>20	0.E+00
Sobic.003G421700	A	glycosyl hydrolases family 17	[1]	0	0	0	0	0	1	1	13	>13	0.E+00
Sobic.002G255600	E	glucan endo-1,3-beta-glucosidase	[1]	1	2	5	6	5	9	8	12	8	3.E-03

Table A.5. Sorghum genes involved in GAX biosynthesis

Family/ID	Clade/Gene	Function	Source	A-30	A-16	A-7	A	A+11	A+25	A+43	A+68	FC	FDR
GT43													
Sobic.001G409100	*IRX9*	Xylan backbone synthesis	[2]	66	42	110	47	19	25	17	4	28	0.E+00
Sobic.010G238800	*IRX14*	Xylan backbone synthesis	[2]	318	366	428	305	145	178	141	129	4	0.E+00
Sobic.006G242100		Nucleotide-diphospho-sugar transferase	[2]	123	107	132	87	51	49	50	32	5	0.E+00
Sobic.009G229300		Nucleotide-diphospho-sugar transferase	[2]	72	105	92	92	62	46	51	27	4	0.E+00
Sobic.009G026100		Nucleotide-diphospho-sugar transferase	[2]	44	68	72	22	29	19	10	13	8	0.E+00
Sobic.003G254700		Nucleotide-diphospho-sugar transferase	[2]	25	48	58	24	13	14	11	7	9	0.E+00
Sobic.006G002600		Nucleotide-diphospho-sugar transferase	[2]	14	17	39	34	11	11	12	7	6	0.E+00
Sobic.002G430700		Nucleotide-diphospho-sugar transferase	[2]	13	1	1	0	6	5	1	0	>13	0.E+00
GT47													
Sobic.003G410700	*IRX10-Like*	Exostosin, xylan backbone synthesis	[1]	189	107	72	64	160	133	137	93	3	1.E-03
Sobic.009G220200	*OsIRX10*	Exostosin, xylan backbone synthesis	[1]	159	117	133	82	75	65	62	53	3	0.E+00
Sobic.009G220100	*OsIRX10*	Exostosin, xylan backbone synthesis	[1]	130	116	117	23	33	25	14	17	11	0.E+00
Sobic.003G410800	*IRX10-Like*	Exostosin, xylan backbone synthesis	[1]	126	166	251	97	48	33	18	8	32	0.E+00
Sobic.003G410600	*IRX10-Like*	Exostosin, xylan backbone synthesis	[1]	47	13	20	7	17	15	11	19	7	0.E+00
Sobic.009G162700	*IRX10-Like*	Exostosin, xylan backbone synthesis	[1]	14	22	31	3	1	3	1	0	>31	0.E+00
GT37													
Sobic.004G308400	*FUT1*	xyloglucan fucosyltransferase	[1]	33	32	68	73	16	30	39	18	6	0.E+00
GT34													
Sobic.001G401600	*XXT1/2*	UDP-xylosyltransferase 2	[1]	60	9	16	9	14	9	4	2	33	0.E+00
Sobic.003G053700		O-fucosyltransferase	[1]	27	46	68	27	14	14	10	6	12	0.E+00
Sobic.K030900		O-fucosyltransferase	[1]	28	24	43	22	14	13	9	7	7	0.E+00
Sobic.010G118300		O-fucosyltransferase	[1]	16	20	22	8	16	11	6	5	5	0.E+00
Sobic.010G087400		O-fucosyltransferase	[1]	10	12	14	25	12	8	10	6	4	0.E+00
Sobic.002G217700		O-fucosyltransferase	[1]	8	17	20	30	20	20	24	16	4	1.E-03
Sobic.003G358100		O-fucosyltransferase	[1]	2	11	12	23	14	14	11	13	14	0.E+00
Sobic.009G158000		O-fucosyltransferase	[1]	1	3	2	7	9	14	11	8	>14	0.E+00
Sobic.007G173300		Rhamnogalacturonate lyase	[1]	2	9	7	6	5	5	3	2	4	7.E-03
GT31													
Sobic.004G051300		Galactosyltransferase	[1]	74	85	81	49	68	53	35	17	5	0.E+00
Sobic.001G050800		Galactosyltransferase	[1]	30	11	28	26	6	6	10	3	10	0.E+00
Sobic.002G068800		Galactosyltransferase	[1]	40	31	83	104	11	22	32	9	11	0.E+00
Sobic.008G169800		Galactosyltransferase	[1]	7	5	5	3	3	3	2	1	5	4.E-03

Table A.5. (Continued)

Family/ID	Clade/Gene	Function	Source	A-30	A-16	A-7	A	A+11	A+25	A+43	A+68	FC	FDR
GT61													
Sobic.003G390700	*XAT1*	Xylan arabinosyl transferase	[2]	182	99.8	156	193	104	94.5	89.3	74	3	0.E+00
Sobic.004G007600	*XAX1*	UDP-xylosyl substitution of arabinoxylan	[2]	123	117	146	59	42	48	36	32	5	0.E+00
Sobic.010G256600		Glycosyltransferase family 61 protein	[2]	50	37	39	19	21	18	9	8	7	0.E+00
Sobic.003G431100		Glycosyltransferase family 61 protein	[2]	0	0	0	1	2	7	6	16	>16	0.E+00
Sobic.003G108500		Glycosyltransferase family 61 protein	[2]	0	1	1	3	2	6	4	7	>7	1.E-03
GT8													
Sobic.003G376700	*GUX1/2/4*	glucuronosyl transferases	[1]	433	408	448	498	241	225	207	124	4	0.E+00
Sobic.001G479800	*GUX1/2/4*	Glucuronosyl substitution of Xylan	[1]	86	93	58	31	76	70	23	12	8	0.E+00
Sobic.009G144200	*GUX1/2/4*	Glucuronosyl substitution of Xylan	[1]	22	8	5	3	3	3	4	3	8	0.E+00
Sobic.006G157800	C	galacturonosyltransferase	[1]	35	8	11	5	9	5	2	2	23	0.E+00
Sobic.003G282600	D	galacturonosyltransferase	[1]	9	1	1	1	1	1	0	0	20	0.E+00
Sobic.001G131900	D	galacturonosyltransferase	[1]	22	12	11	3	16	11	6	3	7	0.E+00
Sobic.010G092400	D	galacturonosyltransferase	[1]	18	7	6	2	11	5	3	2	12	0.E+00
Sobic.001G460000	D	galacturonosyltransferase	[1]	24	47	49	14	39	36	27	14	4	0.E+00
Sobic.008G141800	D	galacturonosyltransferase	[1]	23	37	36	30	25	19	17	9	4	0.E+00
Sobic.004G244100	C	galacturonosyltransferase	[1]	75	37	111	83	32	18	18	8	15	0.E+00
Sobic.001G400100	C	galacturonosyltransferase	[1]	22	15	36	31	9	10	12	6	6	0.E+00
Sobic.001G138200	C	galacturonosyltransferase	[1]	31	7	41	46	26	7	6	2	23	0.E+00
Sobic.002G398400	C	galacturonosyltransferase	[1]	1	5	25	6	1	1	0	0	>25	0.E+00
DUF231 family													
Sobic.009G105000	*TBL29*	o-acetylation of xylan	[2]	26	19	28	13	10	11	10	8	4	2.E-03
Sobic.003G344200	*TBL29*	o-acetylation of xylan	[2]	7	13	23	38	9	19	13	21	5	0.E+00
Sobic.001G406700	*TBL29*	o-acetylation of xylan	[2]	33	57	71	6	1	3	1	0	>71	0.E+00
Mitchell Clade 1													
Sobic.003G043600		HXXXD-type acyl-transferase	[3]	325	85	206	131	40	14	17	13	29	0.E+00
Sobic.003G037800		HXXXD-type acyl-transferase	[3]	88	118	174	97	100	51	62	107	4	0.E+00
Sobic.003G219700		HXXXD-type acyl-transferase	[3]	19	21	27	7	25	19	13	23	4	2.E-02
Sobic.010G180100	*OsAT10*	HXXXD-type acyl-transferase (p-coumaroyl	[3]	17	14	27	4	10	9	4	4	8	0.E+00

Table A.5. (Continued)

Family/ID	Clade/Gene	Function	Source	A-30	A-16	A-7	A	A+11	A+25	A+43	A+68	FC	FDR
Transporters													
Sobic.006G149400	*UXT1*	UDP-Xylose transporter into the golgi	[2]	41	99	108	9	3	19	4	1	>108	0.E+00
Sobic.004G221300	*UXT1*	UDP-Xylose transporter into the golgi	[2]	155	130	150	37	25	21	16	12	14	0.E+00
Sobic.003G055500	*UXT2*	UDP-Xylose transporter into the golgi	[2]	104	64	59	37	70	48	44	33	3	0.E+00
Sobic.009G059500	*UXT3*	UDP-Xylose transporter into the golgi	[2]	53	70	74	15	48	40	32	27	5	0.E+00
Sobic.001G392200	*RWA1-4*	acetyl-CoA transporter	[2]	25	55	48	17	31	23	16	11	5	0.E+00
Sobic.009G246700	*RWA1-4*	acetyl-CoA transporter	[2]	72	91	101	23	44	33	20	15	7	0.E+00

Table A.6. Sorghum genes involved in nucleotide-sugar interconversion.

Family/ID	Function	Source	A-30	A-16	A-7	A	A+11	A+25	A+43	A+68	FC	FDR
UDP-XYL synthase												
Sobic.001G418100	UDP-XYL synthase 6	[1]	330	386	494	55	258	175	89	87	9	0.E+00
Sobic.003G348800	UDP-glucuronic acid decarboxylase 1	[1]	47	62	62	26	49	38	24	19	4	0.E+00
UDP-D-apiose/UDP-D-xylose synthase												
Sobic.003G442300	UDP-D-apiose/UDP-D-xylose synthase 2	[1]	33	20	14	7	22	21	18	20	5	0.E+00
UDP-D-glucuronate 4-epimerase												
Sobic.001G436100	UDP-glucuronate 4-epimerase	[1]	20	7	6	7	17	15	11	9	3	2.E-03
Sobic.002G256100	UDP-glucuronate 4-epimerase	[1]	14	5	5	6	4	2	1	2	>14	1.E-02
Sobic.004G326900	UDP-glucuronate 4-epimerase	[1]	73	17	14	17	18	10	7	8	12	0.E+00
Sobic.007G196800	UDP-glucuronate 4-epimerase	[1]	22	3	9	9	3	3	2	1	18	0.E+00
UDP-Glucose 6-dehydrogenase												
Sobic.001G084100	UDP-glucose 6-dehydrogenase	[1]	943	984	965	673	414	367	346	397	3	1.E-03
Sobic.001G459800	UDP-glucose 6-dehydrogenase	[1]	249	321	401	280	203	179	184	135	3	0.E+00
UDP-D-glucose/UDP-D-galactose 4-epimerase												
Sobic.007G114500	UDP-D-glucose/UDP-D-galactose 4-epimerase	[1]	70	48	39	21	43	38	27	21	4	0.E+00
Sobic.001G372500	UDP-D-glucose/UDP-D-galactose 4-epimerase	[1]	11	16	16	38	13	28	44	61	5	0.E+00
UDP-Xylose 4-epimerase												
Sobic.006G217300	NAD dependent epimerase/dehydratase	[1]	45	124	113	157	100	90	108	67	3	0.E+00
Sobic.002G029800	UDP-arabinose 4-epimerase 1	[1]	38	27	22	17	26	18	12	11	4	0.E+00
Sobic.001G400000	NAD dependent epimerase/dehydratase	[1]	12	19	18	4	12	9	6	4	5	0.E+00
UGPase												
Sobic.006G213100	UTP--glucose-1-phosphate uridylyltransferase	[2]	84	89	96	22	60	50	25	18	6	0.E+00

Table A.7. Sorghum genes encoding glucosyl hydrolases.

Family/ID	Function	Source	A-30	A-16	A-7	A	A+11	A+25	A+43	A+68	FC	FDR
Xyloglucan endotransglucosylase/hydrolases												
Sobic.006G070600	Xyloglucan endotransglucosylase/hydrolase	[1]	1975	262	175	48	84	30	10	9	250	0.E+00
Sobic.010G246700	xyloglucan endotransglucosylase/hydrolase 25	[1]	545	108	42	58	80	29	4	4	155	0.E+00
Sobic.002G302000	xyloglucan endotransglucosylase/hydrolase 32	[1]	75	2	3	1	5	3	0	0	>75	0.E+00
Sobic.006G071300	xyloglucan endotransglucosylase/hydrolase 13	[1]	22	3	4	2	4	2	1	1	>22	0.E+00
Sobic.002G324100	xyloglucan endotransglucosylase/hydrolase 8	[1]	19	2	2	2	2	5	5	5	11	0.E+00
Sobic.010G246400	Xyloglucan endotransglucosylase/hydrolase	[1]	825	139	350	345	126	184	154	69	13	0.E+00
Sobic.006G071400	xyloglucan endotransglucosylase/hydrolase 25	[1]	435	64	145	118	65	45	36	21	23	0.E+00
Sobic.006G205600	xyloglucan endotransglucosylase/hydrolase 24	[1]	300	51	181	168	17	22	43	14	24	0.E+00
Sobic.004G273200	xyloglucan endotransglucosylase/hydrolase 30	[1]	177	71	109	135	71	73	84	67	3	2.E-02
Sobic.010G246600	Xyloglucan endotransglucosylase/hydrolase	[1]	84	14	14	19	19	45	26	44	7	0.E+00
Sobic.005G140000	glycosyl hydrolases family 16	[1]	42	0	1	0	0	17	0	0	>42	0.E+00
Mixed linkage glucanase												
Sobic.006G070400	endo-1,3;1,4-beta-D-glucanase	[2]	17	23	20	12	13	8	7	4	6	0.E+00
Sobic.009G129800	endo-1,3;1,4-beta-D-glucanase	[2]	59	52	37	61	76	106	125	132	3	0.E+00
Sobic.009G130000	endo-1,3;1,4-beta-D-glucanase	[2]	2	2	5	8	5	22	5	15	8	4.E-03
Exo-acting glycanases B-Galactosidase family 35												
Sobic.001G493900	beta-galactosidase	[1]	22	4	4	2	3	4	2	2	12	0.E+00
Sobic.010G173800	beta-galactosidase	[1]	23	1	2	0	1	1	0	0	>23	0.E+00
Glycoside hydrolase family 9 (Endo-1,4-beta-glucanase)												
Sobic.001G099100	glycosyl hydrolase 9A1	[1]	373	378	346	109	335	251	137	81	5	0.E+00
Sobic.002G276600	glycosyl hydrolase 9B7	[1]	204	219	219	78	132	98	75	47	5	0.E+00
Sobic.002G193600	glycosyl hydrolase 9B8	[1]	21	21	18	22	27	17	9	5	5	0.E+00
Sobic.010G101900	glycosyl hydrolase 9B8	[1]	50	51	88	68	34	18	12	8	12	0.E+00
Sobic.004G244600	glycosyl hydrolase 9B8	[1]	29	27	67	53	10	10	9	4	16	0.E+00
Sobic.004G042700	glycosyl hydrolase 9B8	[1]	18	51	72	27	25	11	4	4	22	0.E+00

Table A.8. Sorghum genes encoding pectin lyases and esterases.

Family/ID	Clade/Gene	Function	Source	A-30	A-16	A-7	A	A+11	A+25	A+43	A+68	FC	FDR
Pectin lyases (Pgases), esterases													
Sobic.005G204700	A	petin lyase/polygalacturonase	[1]	31	11	6	6	11	10	6	6	7	0.E+00
Sobic.003G223100	A	petin lyase/polygalacturonase	[1]	20	3	5	3	5	5	4	3	8	0.E+00
Sobic.009G216100	D	petin lyase/polygalacturonase	[1]	19	6	4	9	11	11	5	15	5	7.E-03
Sobic.009G216300	D	petin lyase/polygalacturonase	[1]	16	4	5	1	6	6	5	2	11	0.E+00
Sobic.009G250800	A	petin lyase/polygalacturonase	[1]	8	1	1	0	1	1	0	0	>8	0.E+00
Sobic.004G319600	A	petin lyase/polygalacturonase	[1]	26	34	28	6	15	15	7	6	6	0.E+00
Sobic.010G005000	A	petin lyase/polygalacturonase	[1]	17	20	63	65	66	58	32	55	4	4.E-02
Pectate and pectin Lyases													
Sobic.006G014400	A&C	pectate lyase	[1]	30	1	3	0	0	1	0	0	>30	0.E+00
Rhamnogalacturonan I lyases													
Sobic.007G173300		rhamnogalacturonate lyase	[1]	2	9	7	6	5	5	3	2	4	7.E-03
Pectin methyl esterases													
Sobic.007G146200	D	pectinesterase	[1]	15	2	1	0	0	4	0	0	>15	0.E+00
Sobic.006G086900	A	pectinesterase	[1]	7	3	4	4	2	1	1	1	>75	1.E-03
Pectin acetylesterases													
Sobic.002G388600		pectinacetylesterase	[1]	118	87	111	110	84	61	66	41	3	0.E+00
Sobic.003G384900		pectinacetylesterase	[1]	7	6	7	5	6	2	1	1	>7	0.E+00

Table A.9. Sorghum genes involved in the phenylpropanoid pathway.

Family/ID	Clade/Gene	Function	Source	A-30	A-16	A-7	A	A+11	A+25	A+43	A+68	FC	FDR
Phenylalanine ammonia-lyase													
Sobic.004G220300	*OsPAL*	phenylalanine ammonia-lyase	[1]	4274	4883	3595	3239	1261	1267	160	62	34	0.E+00
Sobic.006G148800	*OsPAL*	phenylalanine ammonia-lyase	[1]	317	1270	658	464	173	184	26	7	54	2.E-03
Sobic.004G220700	*OsPAL*	phenylalanine ammonia-lyase	[1]	196	549	657	649	199	102	71	28	11	3.E-03
Sobic.004G220400	*OsPAL*	phenylalanine ammonia-lyase	[1]	355	364	390	375	193	229	160	139	3	2.E-02
Sobic.004G220600	*OsPAL*	phenylalanine ammonia-lyase	[1]	35	102	151	162	43	21	17	6	11	3.E-03
Cinnamate-4-hydroxylase													
Sobic.002G126600	*AtC4H*	cinnamate-4-hydroxylase	[1]	105	160	125	89	90	86	44	55	5	8.E-03
Sobic.003G337400	*AtC4H*	cinnamate-4-hydroxylase	[1]	6	15	13	7	3	3	2	1	10	0.E+00
Cinnamate-3-hydroxylase													
Sobic.009G181800	*C3H*	cytochrome P450	[1]	99	177	118	63	56	36	11	6	19	0.E+00
Sobic.003G327800	*C3H*	cytochrome P451	[1]	100	68	119	67	49	46	30	33	5	0.E+00
Caffeic acid 3-O-methyltransferase													
Sobic.007G047300	*OsCOMT*	O-methyltransferase	[1]	1270	2444	2346	1352	1157	1095	405	285	7	0.E+00
Ferulic acid 5-hydroxylase													
Sobic.001G196300	*AtF5H*	ferulic acid 5-hydroxylase 1	[1]	13	90	96	74	70	35	20	22	7	0.E+00
Cinnamoyl coa reductase													
Sobic.007G141200	*OsCCR*	cinnamoyl CoA reductase	[1]	181	652	661	355	191	51	17	5	46	0.E+00
Sobic.004G065600	*OsCCR*	cinnamoyl CoA reductase	[1]	12	38	50	66	14	15	17	10	7	5.E-03
Sobic.010G066000	*OsCCR*	cinnamoyl CoA reductase	[1]	8	45	31	31	10	2	2	1	>45	0.E+00
Sobic.001G037800	*OsCCR*	cinnamoyl CoA reductase	[1]	2	2	5	5	6	10	18	16	7	0.E+00
4-coumarate:CoA ligase													
Sobic.004G062500	*At4CL*	4-coumarate:CoA ligase 1	[1]	230	653	403	467	309	113	86	85	8	0.E+00
Sobic.010G214900	*At4CL*	4-coumarate:CoA ligase 1	[1]	79	58	70	59	56	45	41	41	3	1.E-03
Sobic.004G272700	*At4CL*	4-coumarate:CoA ligase 3	[1]	34	16	25	4	9	4	2	4	19	0.E+00
Sobic.007G145600	*At4CL*	4-coumarate:CoA ligase 1	[1]	1	6	8	9	3	4	5	10	>10	4.E-03
Sobic.001G500800	*At4CL*	4-coumarate--CoA ligase-like 7	[1]	3	17	20	43	45	69	61	59	16	0.E+00
Caffeoyl-CoA O-methyltransferase													
Sobic.010G052200	*OsCCoAOMT*	caffeoyl-CoA O-methyltransferase	[1]	632	1099	1181	1207	591	276	247	187	6	0.E+00
Sobic.007G218800	*OsCCoAOMT*	caffeoyl-CoA O-methyltransferase	[1]	442	643	494	138	88	85	61	66	12	0.E+00

Table A.9. (Continued)

Family/ID	Clade/Gene	Function	Source	A-30	A-16	A-7	A	A+11	A+25	A+43	A+68	FC	FDR
Hydroxycinnamoyl-CoA shikimate/quinate hydroxycinnamoyl transferase													
Sobic.004G212300	*AtHCT*	hydroxycinnamoyl-CoA shikimate/quinate hydroxycinnamoyl transferase	[1]	309	361	323	268	173	117	74	57	5	0.E+00
Sobic.006G136800	*AtHCT*	hydroxycinnamoyl-CoA shikimate/quinate hydroxycinnamoyl transferase	[1]	143	149	200	150	71	29	23	12	15	0.E+00
Sobic.002G175600	*AtHCT*	hydroxycinnamoyl-CoA shikimate/quinate hydroxycinnamoyl transferase	[1]	18	6	3	2	10	8	5	3	11	0.E+00
Sobic.010G066700	*AtHCT*	hydroxycinnamoyl-CoA shikimate/quinate hydroxycinnamoyl transferase	[1]	4	12	18	30	6	6	7	25	7	3.E-02
Sobic.002G205100	*AtHCT*	hydroxycinnamoyl-CoA shikimate/quinate hydroxycinnamoyl transferase	[1]	3	5	6	4	14	18	12	11	5	4.E-03
Sobic.002G205300	*AtHCT*	hydroxycinnamoyl-CoA shikimate/quinate hydroxycinnamoyl transferase	[1]	20	9	9	16	8	16	11	36	6	7.E-03
Sobic.007G142100	*AtHCT*	hydroxycinnamoyl-CoA shikimate/quinate hydroxycinnamoyl transferase	[1]	3	2	3	3	4	15	12	17	10	0.E+00
Cinnamyl alcohol dehydrogenase													
Sobic.006G014700	*OsCAD*	cinnamyl-alcohol dehydrogenase	[1]	3	23	33	59	15	27	12	16	21	2.E-03
Sobic.010G071900	*OsCAD*	cinnamyl-alcohol dehydrogenase	[1]	5	7	10	10	14	32	19	17	5	0.E+00
Sobic.010G072000	*OsCAD*	cinnamyl-alcohol dehydrogenase	[1]	2	4	4	5	6	12	7	6	4	9.E-03
p-coumaroyl-CoA:monolignol transferase													
Sobic.009G034300	*OsPMT*	HXXXD-type acyl-transferase	[4]	94	194	232	104	81	103	25	43	11	0.E+00
Sobic.010G180100	*OsPMT*	HXXXD-type acyl-transferase	[4]	17	14	27	4	10	9	4	4	5	1.E-03

Table A.10. Sorghum genes encoding laccases and peroxidases

Family/ID	Function	Source	A-30	A-16	A-7	A	A+11	A+25	A+43	A+68	FC	FDR
Laccases												
Sobic.009G162300	laccase	[2]	91	25	35	3	1	9	0	1	>91	0.E+00
Sobic.003G353200	laccase	[2]	80	11	16	0	0	2	0	0	>80	0.E+00
Sobic.003G357600	laccase	[2]	26	10	9	11	11	9	6	8	5	0.E+00
Sobic.005G005800	laccase	[2]	8	1	1	0	0	0	0	0	>8	0.E+00
Sobic.001G422300	laccase	[2]	43	82	72	14	11	18	12	4	20	0.E+00
Sobic.003G352700	laccase	[2]	39	84	77	6	1	8	2	1	>84	0.E+00
Sobic.003G352800	laccase	[2]	10	12	5	1	1	2	1	0	>12	0.E+00
Peroxidases												
Sobic.003G121100	peroxidase	[1]	224	28	12	2	3	4	1	1	>224	0.E+00
Sobic.001G528100	peroxidase	[1]	174	10	9	4	4	10	3	3	80	0.E+00
Sobic.001G444500	peroxidase	[1]	136	3	3	0	0	1	0	0	>136	0.E+00
Sobic.001G328100	peroxidase	[1]	119	25	19	3	19	14	4	1	>119	0.E+00
Sobic.004G105100	peroxidase	[1]	65	10	9	5	5	6	7	2	34	0.E+00
Sobic.010G161900	peroxidase	[1]	28	9	6	6	7	3	2	3	14	0.E+00
Sobic.008G114700	peroxidase	[1]	23	2	5	1	5	3	1	0	>23	0.E+00
Sobic.002G357000	peroxidase	[1]	22	5	6	5	3	1	1	0	>22	0.E+00
Sobic.001G277000	peroxidase	[1]	17	3	2	0	1	3	0	0	>17	0.E+00
Sobic.007G192300	peroxidase	[1]	61	21	25	40	17	14	13	22	6	0.E+00
Sobic.009G032600	peroxidase	[1]	457	337	179	61	81	93	40	31	16	0.E+00
Sobic.009G055100	peroxidase	[1]	155	136	105	131	95	84	24	47	8	5.E-02
Sobic.002G416700	peroxidase	[1]	13	10	17	3	3	1	0	1	>17	0.E+00
Sobic.005G011200	peroxidase	[1]	20	165	174	7	9	10	2	3	103	0.E+00
Sobic.003G183300	peroxidase	[1]	43	11	33	3	3	4	1	1	>43	0.E+00
Sobic.009G186600	peroxidase	[1]	16	31	23	14	17	11	6	3	9	0.E+00
Sobic.002G416600	peroxidase	[1]	26	32	4	10	24	2	1	1	32	2.E-02
Sobic.005G011300	peroxidase	[1]	150	234	206	57	154	48	26	22	11	0.E+00
Sobic.003G127100	peroxidase	[1]	224	69	215	249	72	114	112	79	4	1.E-03
Sobic.010G232500	peroxidase	[1]	61	55	23	49	35	17	8	8	9	0.E+00
Sobic.010G245800	peroxidase	[1]	86	41	155	197	45	72	54	58	5	0.E+00
Sobic.009G144900	peroxidase	[1]	1	7	19	14	1	1	1	1	>19	0.E+00
Sobic.010G245500	peroxidase	[1]	4	12	3	9	17	21	2	3	11	3.E-02
Sobic.001G235800	peroxidase	[1]	0	1	0	2	21	29	8	7	>29	0.E+00
Sobic.002G234200	peroxidase	[1]	1	5	3	13	14	11	48	97	>97	0.E+00

Table A.11. Sorghum genes involved in sucrose metabolism

Family/ID	Clade/Gene	Function	Source	A-30	A-16	A-7	A	A+11	A+25	A+43	A+68	FC	FDR
Sucrose synthase, SPS													
Sobic.001G344500	*SS4*	sucrose synthase	[2]	499	546	985	86	183	161	84	155	14	0.E+00
Sobic.001G378300	*SS3*	sucrose synthase	[2]	10	25	19	46	46	58	75	64	6	0.E+00
Sobic.005G089600	*SPS*	sucrose-phosphate synthase	[2]	1	2	2	6	6	11	12	17	>17	0.E+00
Cytosolic invertase													
Sobic.004G172700	*bCIN*	cytosolic Invertase	[2]	87	189	229	178	74	60	69	55	5	5.E-03
Sobic.003G153800	*aCIN*	cytosolic Invertase	[2]	30	10	14	19	5	5	5	5	8	0.E+00
Sobic.001G391600	*aCIN*	cytosolic Invertase	[2]	8	12	27	19	9	11	12	9	4	1.E-03
Sobic.004G255600	*bCIN*	cytosolic Invertase	[2]	5	7	16	41	5	2	8	14	24	0.E+00
Vacuolar invertase													
Sobic.004G004800	*SbVIN1*	vacuolar Invertase	[2]	358	296	193	85	31	18	2	4	225	0.E+00