CHARACTERIZATION OF STEM GROWTH AND CHEMICAL COMPOSITION IN *SORGHUM BICOLOR*

A Thesis

by

ROBERT TAYLOR ANDERSON JR

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

MASTER OF SCIENCE

Chair of Committee,	John E. Mullet
Committee Members,	William L. Rooney
	Hays Rye
	Dorothy Shippen
Head of Department,	Gregory D. Reinhart

May 2014

Major Subject: Biochemistry

ABSTRACT

Sorghum bicolor is a subtropical grass grown throughout the world for human consumption, animal feed and for the growing biofuels industry. In this thesis I characterize sorghum stem growth and chemical composition, and identify QTL and candidate genes which may regulate stem development. In addition, I attempt to correlate variation in stem composition with saccharification efficiency in two sorghum populations. Under greenhouse conditions, stem length in the vegetative phase typically accelerated starting with the 10th internode, possibly reaching a maximum elongation rate by the time the 20th internode stopped growing. The rate of stem diameter growth increased steadily and achieved a maximum diameter growth rate by the time the 11th internode stopped growing. Under these growth conditions, the plateau for the internode diameter growth rate occurred at 60 days after seedling emergence.

Variation in internode growth was driven primarily by cell division, though differences in cell size began to play a significant role as cell number increased. Internode growth was positively correlated with GRAS, bHLH and B3 transcription factor expression and was negatively correlated with AP2/EREBP, MYB and WRKY transcription factors. In addition, the outer rind of the internode had gene expression that resembled expression in young tissue while the central tissue had gene expression that resembled expression in mature tissue. Sorghum accessions displayed a wide range of internode chemical compositions and saccharification efficiencies. However, no clear

patterns were present between variation in composition from the NIR spectra and variation in saccharification efficiency on the iWALL system. Multiple linear regression analysis, however, revealed that high biomass density appeared to inhibit saccharification in these populations.

Overall, sorghum stem growth mirrors that seen in other monocots. The gene expression information presented here should be useful for future studies on the roles of various transcription factors in plant development and for the identification of transcription factor binding sites within the genome. Such information will be important for future success in molecular breeding and marker assisted selection. This information will also be invaluable for designing genotypes with novel transcripts that can be activated at specific times, in specific organs and under specific conditions.

To my wife, Jenn. Thank you for always believing in me.

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. John Mullet, for his tremendous patience during my thesis research and for his enthusiasm for the research process. I would also like to thank all of the students, staff and workers in Dr. William Rooney's lab for planting and maintaining field populations, and for allowing me to use their seed vault and other instruments. I am truly indebted to Susan Hall and the student workers of the Mullet lab, without whom I would not have been able to collect enough data to perform many of these studies. I must also acknowledge Joseph Evans and Dustin Herb for allowing me to use the NIR data that they collected as part of my thesis work. In addition, I would like to thank all the other members of the Mullet lab, but in particular Brian McKinley, for many interesting and thought-provoking conversations over the years.

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

1. INTRODUCTION

Sorghum bicolor (L) Moench is a subtropical grass that originated in the African subcontinent over 6000 years ago (Smith & Frederiksen, 2000). The species is divided into five races: Durra, Kafir, Caudatum, Guinea and Bicolor (House, 1985). These races are classified based on specific physical characteristics as well as their regions of origin and domestication. Sorghum is a photoperiod sensitive short day plant, although varieties differ in their critical photoperiod to induce floral initiation. However, all genotypes will delay flowering until a predetermined number of days (or thermal degree days), set by the maturity genes (House, 1985). All cultivated varieties of *Sorghum bicolor* are annual, though many closely related species are perennial (Price *et al*., 2005).

Sorghum is a well-established crop species that is grown for a variety of end uses. In Africa, sorghum grain is ground into meal and is a staple crop for human consumption. In the United States, sorghum has been grown historically for either animal feed or for molasses (House, 1985). Plants grown for animal fodder are bred for high digestibility and low concentration of antiquality compounds (Casler, 2001). For sorghum, this is mainly accomplished by lowering the concentration of lignin and cyanide, respectively, in the leaf biomass. Sorghum grown for molasses has a high concentration of soluble sugars in the plant stalk after grain maturity (Smith & Frederiksen, 2000). Regardless of the particular end use, sorghum is the crop of choice for high yield under semi-drought conditions. Sorghum also requires less fertilizer than

corn, and these factors make sorghum desirable as a lower-input alternative to corn for silage (House, 1985).

The sorghum germplasm collection (n=43,000) contains a relatively high amount of genetic diversity (Menz *et al*., 2004; GRIN). In addition, sorghum accessions can vary widely in many phenotypic traits of agronomic and biological interest (Menkir *et al.*, 1997). Stem morphology in particular is quite variable within the species. Figure 1 shows the variation in internode length and diameter observed within a survey of 200 photoperiod sensitive, late flowering bioenergy-type sorghum accessions. In general, sorghum genotypes can appear either 'bushy' or tree-like, depending mainly on the length, diameter and number of stalks or tillers per plant.

Stem diameter growth is critical to maintain the plant in an upright position. Very tall genotypes are especially vulnerable to stalk breakage (stalk lodging) and root lodging, especially in wet soil or after a heavy rain. I expect that breeding sorghum for increased stem diameter could improve stalk-lodging resistance, which will increase agronomic yield per field plot. In addition, the stem is a reservoir of water that can be accessed during periods of drought. In fact, small fluctuations in stem diameter have been used as a biological assay for water status in many tree species (Simonneau *et al*., 1993). Breeding for increased stem diameter might increase the reserve supply of water and prevent cellular dehydration in prolonged periods of drought, when stomata close and the plant stops fixing carbon dioxide.

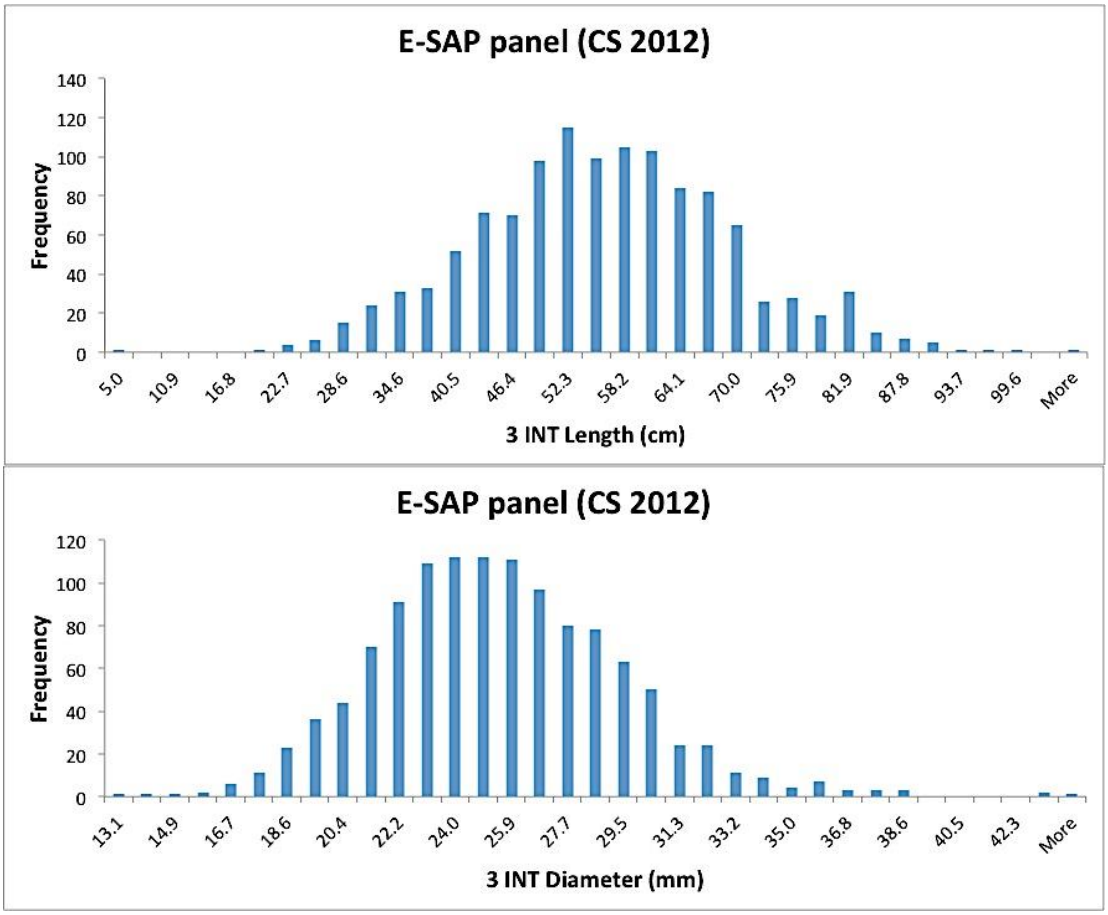

Figure 1. Histograms for stem length and diameter variation among energy sorghum accessions. Data were collected from 200 diverse lines of the E-SAP panel grown for 130 days in the College Station research field during the summer of 2012. Data of each accession represents the average of six biological replicates collected from two different field plots.

The pith parenchyma in the ground tissue of the stem can accumulate significant quantities of sucrose and low levels of starch (Tarpley & Vietor, 2007; Wilson *et al*., 1993). This energy reserve can be accessed in times of stress, during tiller outgrowth, or during ratooning (Hattori *et al*., 2008). In sweet sorghum, sucrose and other soluble

sugars accumulate to very high levels in stems at grain maturity, making up as much as 19% of the stem fresh weight in the genotype 'Rio' in some environments (Wang & Liu, 2009). In the energy sorghum collection, stem water content was more highly correlated with internode diameter than internode length (Figure 2). Therefore, if stem sugar is the economically desirable product, selecting for increased stem diameter has the potential to improve sucrose yield per plant by increasing stem sucrose sink volume.

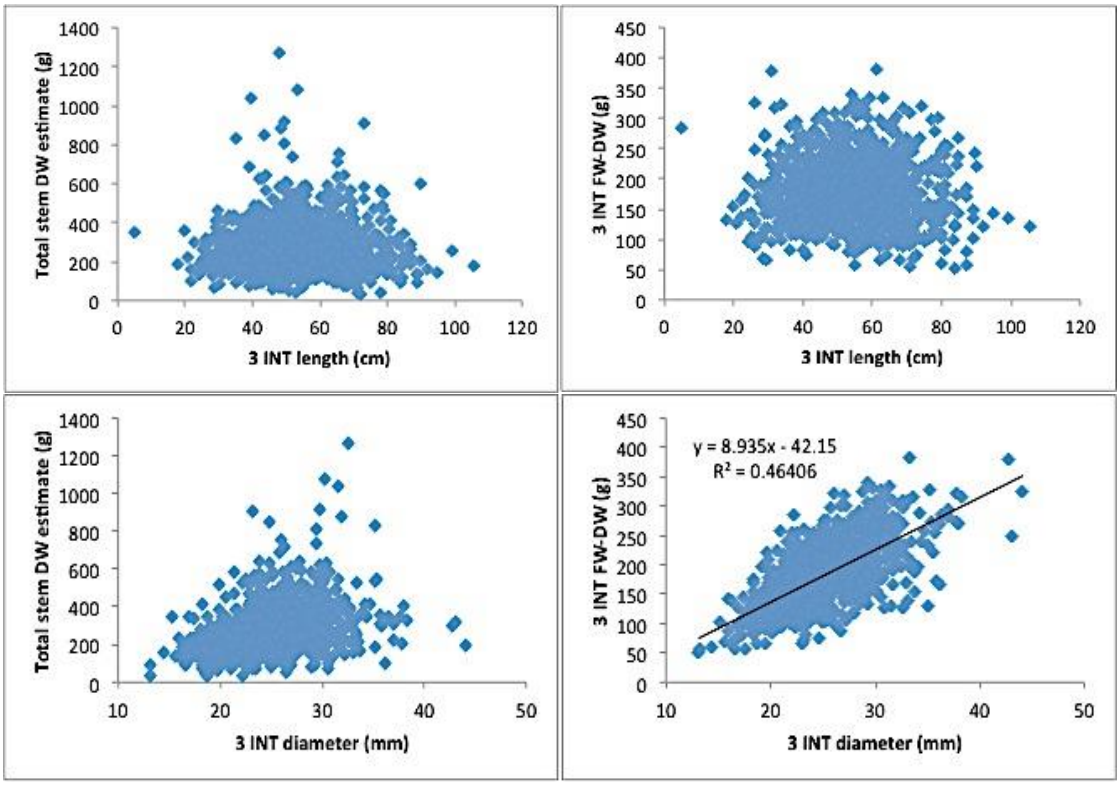

Figure 2. Correlation between stem morphology and either stem dry weight (DW) or stem water content (FW-DW). Data were collected from the E-SAP panel grown in the College Station research field during the summer of 2012.

The stem primarily evolved as a conduit for water and nutrient flow between roots and shoots, and as means of structural support for increased canopy height in land plants. Competition between plants for sunlight is one reason for the evolutionary drive to increase plant height. The shade avoidance syndrome (SAS) is an example of the response to such plant competition. The SAS response is a morphological shift that includes inducing the plant to develop longer and thinner internodes (Figure 3) under shade conditions than it would without the response (reviewed by Smith & Frederiksen, 2000).

Figure 3. Effect of plant spacing on internode size. Stem segments were collected from the base of the stalk of three plants of the energy hybrid ES5200 grown at three different effective plant spacing. The bottom segment was collected from a plant at the edge of the plot, spaced far apart from neighboring plants.

2. MATERIALS & METHODS

Plant Material and Growth Conditions

Seed of genotypes 100M, SM100, 58M, ES5200, Rio & BTx623 were obtained from the lab of Dr. William Rooney at Texas A&M University. Plants of these genotypes were grown in a 2:1 mixture of coarse vermiculite (Sun Gro Horticulture) to Brazos County silty loam soil. Approximately 30mL of Osmocote 14-14-14 fertilizer (Scotts) was then mixed into the topsoil by hand. A 6cm layer of Metro-Mix (Sun Gro Horticulture) was poured into each pot prior to adding the mixed soil in order to prevent soil loss.

Greenhouse temperature was 33^0C during the day and 23^0C at night. Average light intensity in the greenhouse at 10AM and 2PM was 450 and 670 microeinsteins respectively between January and April, during the period in which the cell size experiment was conducted. The energy sorghum hybrid ES5200 was grown in the outdoor research field in the summer of 2012 at very wide plant spacing (3 ft) under high light intensity (1500-1800 microeinsteins). All tissue harvests for gene expression and confocal microscopy studies started promptly three hours post dawn and ended no later than five hours post dawn.

Both the E-SAP association panel and the BTx642*Tx7000 mapping population were grown in the outdoor research field in the summer of 2012 and 2010, respectively, at approximately 7 inch plant spacing from stem to stem. Following harvest, plant stem

tissue was dried at 71^{0}C and ground to a powder in a Wiley Mill (Thomas Scientific). This powder was then passed through a 2mm sieve and stored at room temperature in Ziploc bags prior to composition analysis.

Microscopy

Internode cross-sections were treated with Pontamine Fast Scarlet 4B (Sigma-Aldrich) for five minutes to stain cellulose microfibrils. Cross section images were collected with a C-4040 zoom digital camera (Olympus) mounted to a MVX10 Macro Zoom Stereomicroscope (Olympus) or with a FV1000 Confocal microscope (Olympus). I estimated cell size and cell number using ImageJ image analysis software (Schneider *et al*., 2012).

RNA-seq

Fresh tissue samples were flash frozen in liquid nitrogen and stored a -80^{0}C prior to RNA extraction. Frozen samples were ground into a powder in a heat sterilized mortar and pestle, and total RNA was extracted from samples using the TRIzol protocol (Molecular Research Center). All subsequent steps of cDNA library preparation followed the Truseq protocol v2.0 for mRNA sequencing (Illumina).

cDNA libraries were sequenced on a Hiseq 2500 system (Illumina). Sequence cluster identification, quality prefiltering, base calling and uncertainty assessment were performed in real time using Illumina's HCS 1.5.15.1 and RTA 1.13.48.0 software with the default parameter settings by the Genomics and Bioinformatics Services group at

Texas A&M. The sequence was aligned to the Sorghum v2.1 genome (Gramene) and analyzed for gene expression differences using both the CLC genomics workbench (CLC bio) and the Tuxedo Suite (Trapnell *et al*., 2012). I then visualized gene expression differences using MapMan v3.5.1 (Thimm *et al*., 2004) and PageMan (Usadel *et al*., 2006).

Near-Infrared Spectroscopy

Ground stem powder from the E-SAP panel and the BTx642*Tx7000 mapping populations was analyzed for gross compositional variation in key yield components via Near-Infrared Spectroscopy (NIR) by Dustin Herb and Joseph Evans, respectively. Samples were placed in quartz sample cups and scanned on a grating-monochrometer type NIR instrument (XDS, Foss North America, Eden Prairie, MN) in reflectance mode. All samples were scanned twice and duplicate spectra were averaged prior to analysis. Quality-control spectra were also collected after every 50 scans to check instrument stability. Sample spectra were analyzed for relevant yield components using the sorghum NIR models published by Wolfrum *et al.* (2013).

iWALL High-Throughput Conversion of Biomass to Monosaccharides

Dried, ground stem tissue from the E-SAP panel and the BTx642*Tx7000 mapping populations was analyzed for enzymatic conversion efficiency by Dr. Nicholas Santoro at the Great Lakes Bioenergy Research Center (GLBRC) using the iWALL system (Santoro *et al*., 2010). In this system, solid samples are pretreated with dilute

(6.25 mM) NaOH solution in order to break crosslinking between hemicellulose and lignin, and to reduce cellulose crystallinity (Mosier *et al*., 2005). Following pretreatment, the biomass is then treated with the Accellerase 1000 enzyme, which has broad hydrolytic activity on cellulose and hemicellulosic polymers. After incubation at 50^0C for 20 hours, the digested sugars are isolated, and glucose and xylose quantity are measured in independent enzymatic assays.

Multiple Regression Analysis

Multiple linear regression analysis was performed in Rstudio v(0.98.490) using the lm( ) function. Variance-inflation factors were calculated with vif( ) using the car R package. Histograms and Q-Q plots of the residuals were performed with hist( ) and qqnorm( ), respectively. Input variable standardization was performed with the standardize( ) function in the arm R package (Gelman, 2008). Cross validation was performed with the cvFit( ) function in the cvTools package using the default settings (Alfons, 2012). All other analyses were performed in Microsoft Excel, 2010.

3. THE TIME COURSE OF INTERNODE GROWTH IN *S. BICOLOR*

During the early growth of a sorghum seedling, the stem forms an elliptical mass of tissue. After 21 days of growth, nodes are visible and the stem morphology resembles that seen in the adult plant (Figure 4, *upper right*). Stem tissue corresponding to internodes 1-5 is very small, buried underground and covered in nodal roots. Because this tissue is small and relatively inaccessible, it was not measured in the time course, though it is likely important for setting the foundations of later stem and nodal root growth. Figure 4 shows the pattern of stem growth of sorghum genotypes grown for up to 100 days in the greenhouse. Internode length remained relatively short in early development. In most genotypes the length of internodes started to increase significantly starting ~30 DAE (Figure 4, *center left*). In contrast, stem diameter steadily increased throughout early development, reaching a plateau at 50~60 DAE (Figure 4, *center right*).

Internodes of ES5200 and BTx623 displayed similar growth patterns during vegetative growth. After BTx623 became florally induced, the rate of internode growth slowed down as the plant diverted more photosynthetic resources to the developing panicle/grain. The difference in internode growth rate between ES5200 and BTx623 became apparent starting at 40 DAE, which is ~35 days prior to anthesis in BTx623. Rio began stem elongation earlier than either ES5200 or BTx623, growing rapidly starting at ~30 DAE. 58M displayed highly precocious stem elongation, and rapidly

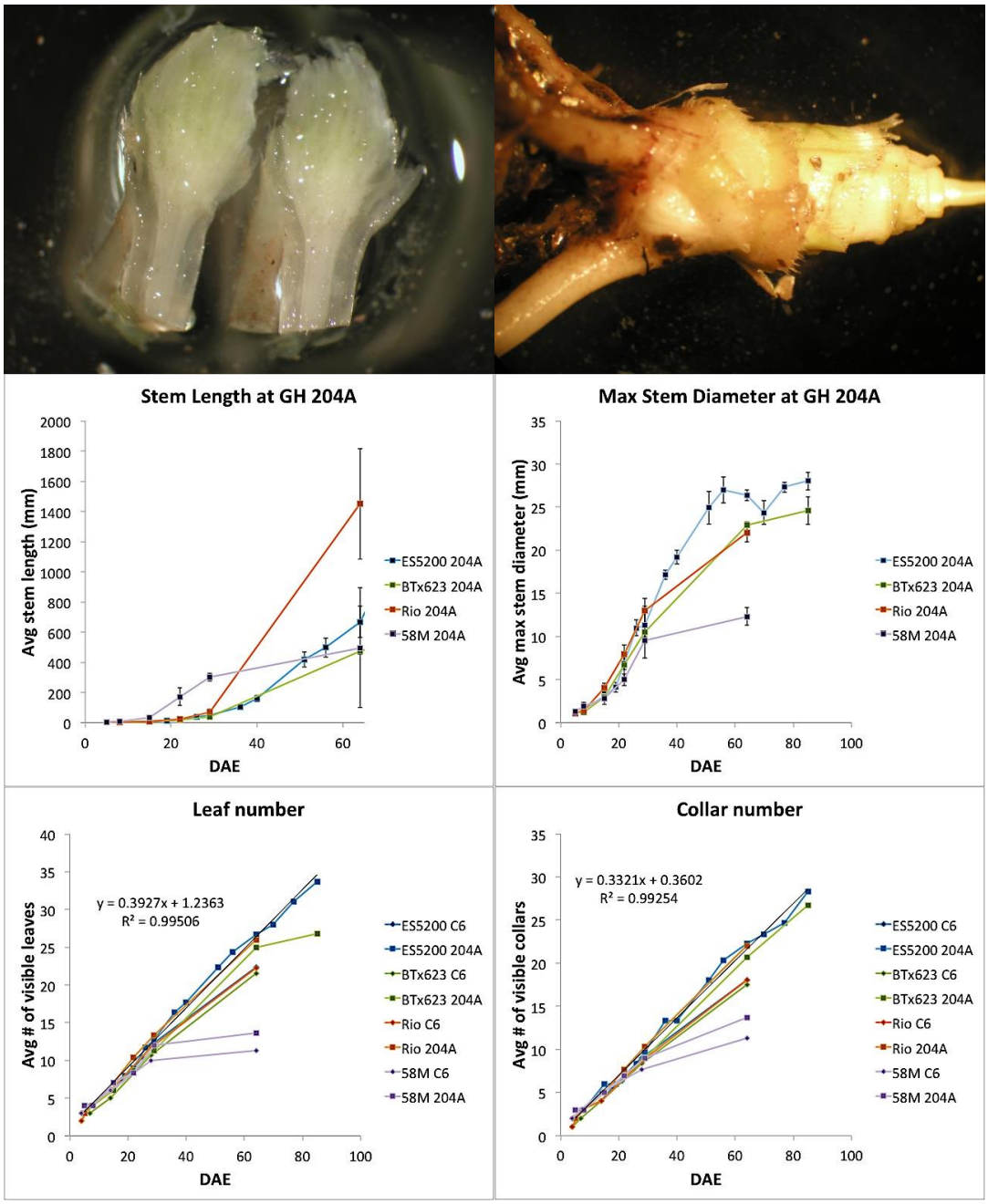

Figure 4. Juvenile stem growth in ES5200, BTx623, Rio and 58M. ES5200 stem tissue split longitudinally at 8 DAE (upper left) and stem tissue with leaf sheaths removed at 21 DAE (upper right). Time course of stem elongation (center left) and internode diameter growth (center right). Leaf (lower left) and collar (lower right) appearance rates in the greenhouse conditions surveyed.

elongated after 15 days of growth. However, stem elongation ceased in 58M once the plant transitioned to flowering.

Leaf and collar appearance rates were constant throughout plant development in the greenhouse. Plants produced a new phytomer after approximately every three days of growth (Figure 4, *lower left & right*). Using this steady rate of phytomer production, I was able to relate plant age in days to internode age. I then evaluated the relationship between internode development and plant age by plotting the growth of each internode relative to the largest internode observed at 90 DAE (Figure 5).

In ES5200 and Rio, internode diameter increased rapidly starting with internode 5 and reached a maximum diameter by internode 10~15 (Figure 5, *upper left & right*). Internode length increased rapidly starting about internode 8~9 and may have reached a maximum length by internode 20 (Figure 5, *upper left & right*). BTx623 showed a similar pattern of internode growth, however the decreased rate of internode growth at internodes eight to ten suggest that some of these plants experienced environmental conditions which inhibited the growth of these internodes (Figure 5, *lower left*). 58M displayed very sporadic internode elongation, though internode diameter growth was more consistent across internodes (Figure 5, *lower right*). This growth pattern is typical of 58M, which often has very short internodes interspersed between highly elongated internodes.

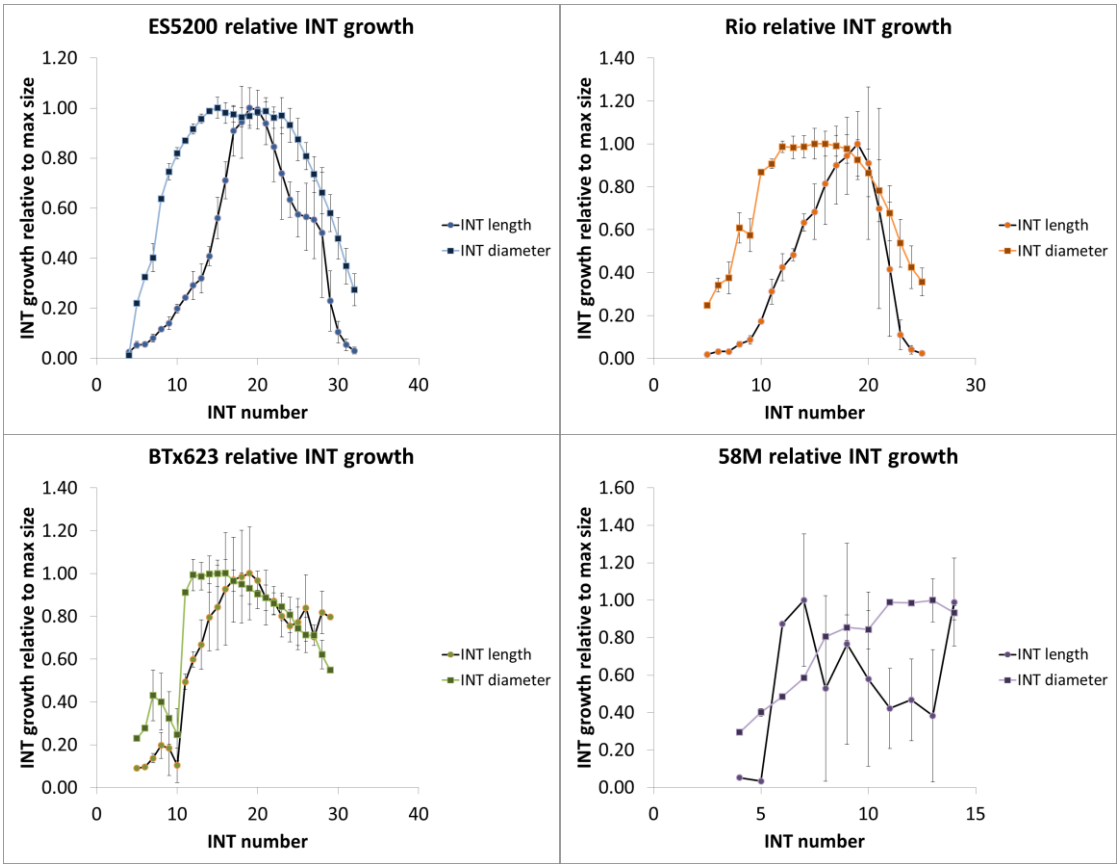

Figure 5. Time course of internode development in ES5200, BTx623, Rio and 58M. Data were obtained from plants grown in the greenhouse under 14 hour days and an average mid-day light intensity of ~1100 umol m^{-2} s^{-1}. Internode (INT) numbering is based on leaf numbering, which includes the cotyledon as leaf #1. Individual data points represent the average of at least three biological replicates. DAE = days after seedling emergence.

The genotype 58M is known to have a nonfunctioning allele of the light-sensing protein phytochrome B (Childs *et al.*, 1997). Because phytochrome B is always inactive in 58M, this genotype constitutively expresses the shade avoidance response, and this response causes the altered growth pattern in this genotype. Phytochrome B regulates the

response to red and far red light; it should be noted that the shade avoidance response can also be induced by a change in the ambient blue light, controlled by the activity of cryptochromes and phototropins (Keuskamp *et al*., 2012).

4. CELL EXPANSION & DIVISION IN SORGHUM STEM TISSUE

Figure 6 shows the shoot architecture of genotypes 58M (Ma1, ma3R), 100M (Ma1, Ma3) and SM100 (ma1, Ma3) at DAE 50 in the three genotypes surveyed. These genotypes are closely related grain type sorghums. The main source of genetic variation between these lines is the presence or absence of functional maturity (flowering time) loci Ma1 and Ma3. Ma3 was previously shown to correspond to the Sorghum phytochrome B (PhyB) gene (Childs *et al.*, 1997) and influences both flowering time and the shade avoidance response in many plant species (Lopez-Juez *et al.*, 1992). Ma1 delays flowering under long days (>14 hrs light) but has less impact on flowering time in short days (Murphy *et al*., 2011). Ma1 was shown to correspond to PRR37, a gene that is regulated by the circadian clock and light (Murphy *et al*., 2011). The consistent difference in vegetative growth between 100M and SM100 suggests that Ma1 has additional effects on plant growth apart from floral signaling (Figure 6).

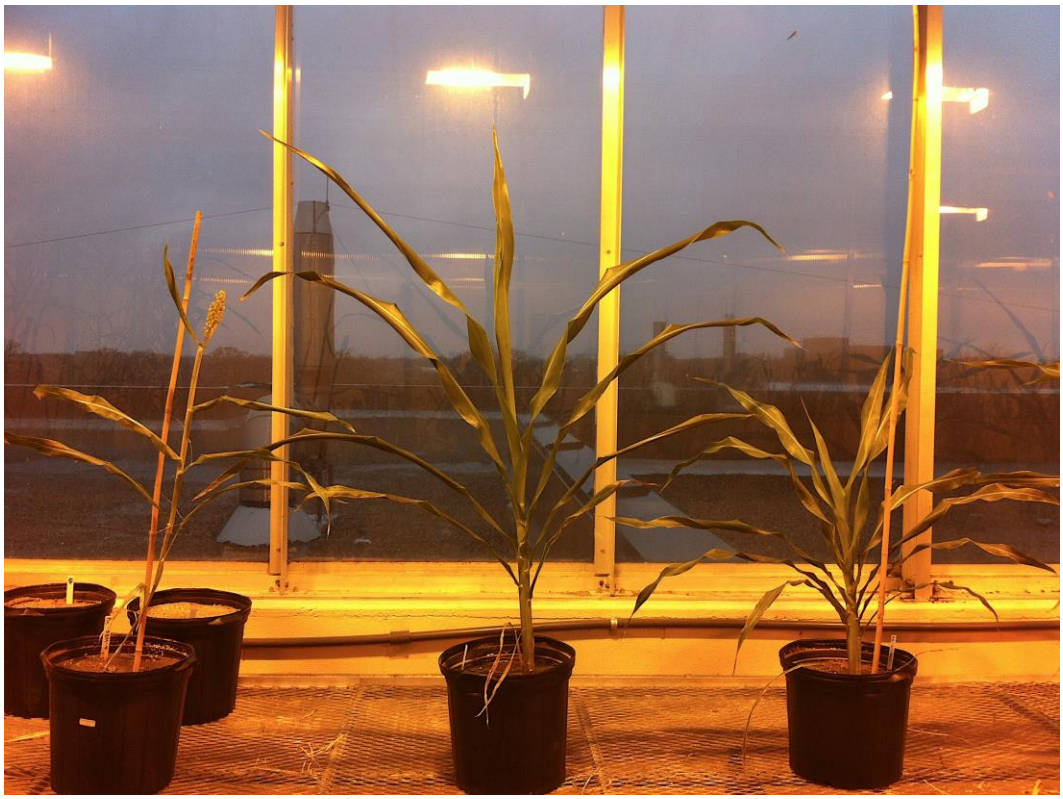

Figure 6. 58M (left), SM100 (center) and 100M (right) at 50 DAE in GH 204A.

Figure 7 shows the distribution of internode length and diameter observed within the three genotypes. As in Section 3, 58M again displayed highly elongated internodes, as long as 335mm in one internode. In addition, 58M displayed sporadic internode elongation, with adjacent internodes varying widely in length. Internode diameter was also extremely thin compared to most sorghum genotypes, and compared to SM100 and 100M in particular. Internode growth was highly similar between genotypes SM100 and 100M, and data from both lines were combined in subsequent analyses.

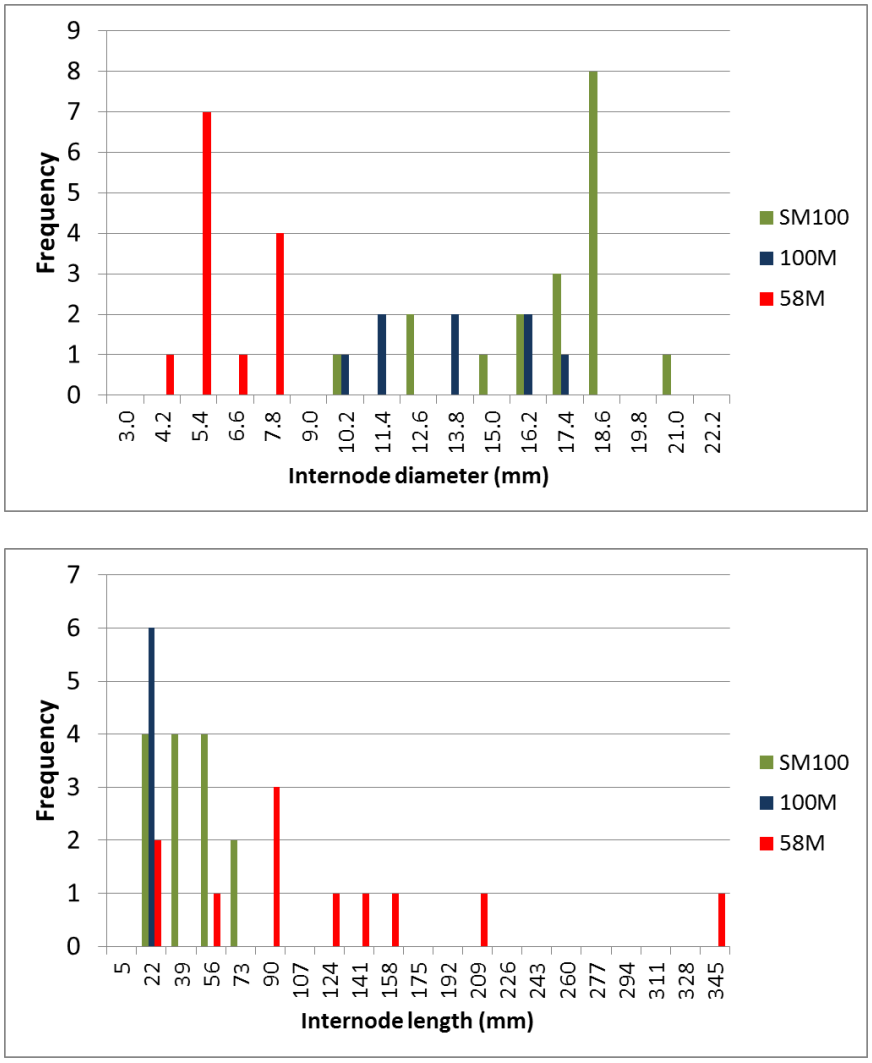

Figure 7. Distribution of internode length and diameter in genotypes SM100, 100M and 58M.

Figures 8 & 9 show confocal images of parenchyma cell size in the growing internodes of 58M, 100M and SM100. From the confocal data, I measured the average cell size and cell densities at various positions along the radial axis of the internode. Cell size was typically much smaller near the rind compared to the center of the

internode (Figure 8, *lower left* vs. *upper left*). In addition, parenchyma cells typically lost their radial symmetry during growth; likely due to variation in the pressures induced by the growth of the surrounding cells (Figure 9, *upper left* vs. *upper right*). 58M always had cells that were noticeably longer than those in internodes of SM100 and 100M (Figure 9, *bottom*). Gibberellic acid (GA) is known to stimulate cellular elongation, and so the observation that a phytochrome B mutant contains highly elongated cells fits with the known relationship between GA, DELLA and phytochrome B in sorghum and other plant species (Childs *et al*., 1997; de Lucas *et al*., 2008).

Additional studies will be necessary to establish the relationship between phyB signaling and cellular growth. However, the pattern that has emerged from this data is one in which the absence of phytochrome B directs pith parenchyma cells to rapidly elongate at the expense of radial cell division in the stem. One possibility is that the inactive form of phytochrome B directly shuts off cell division, and that the hydrostatic pressure in the stem tissue could then only be relieved through vacuolar expansion and elongation of the cells.

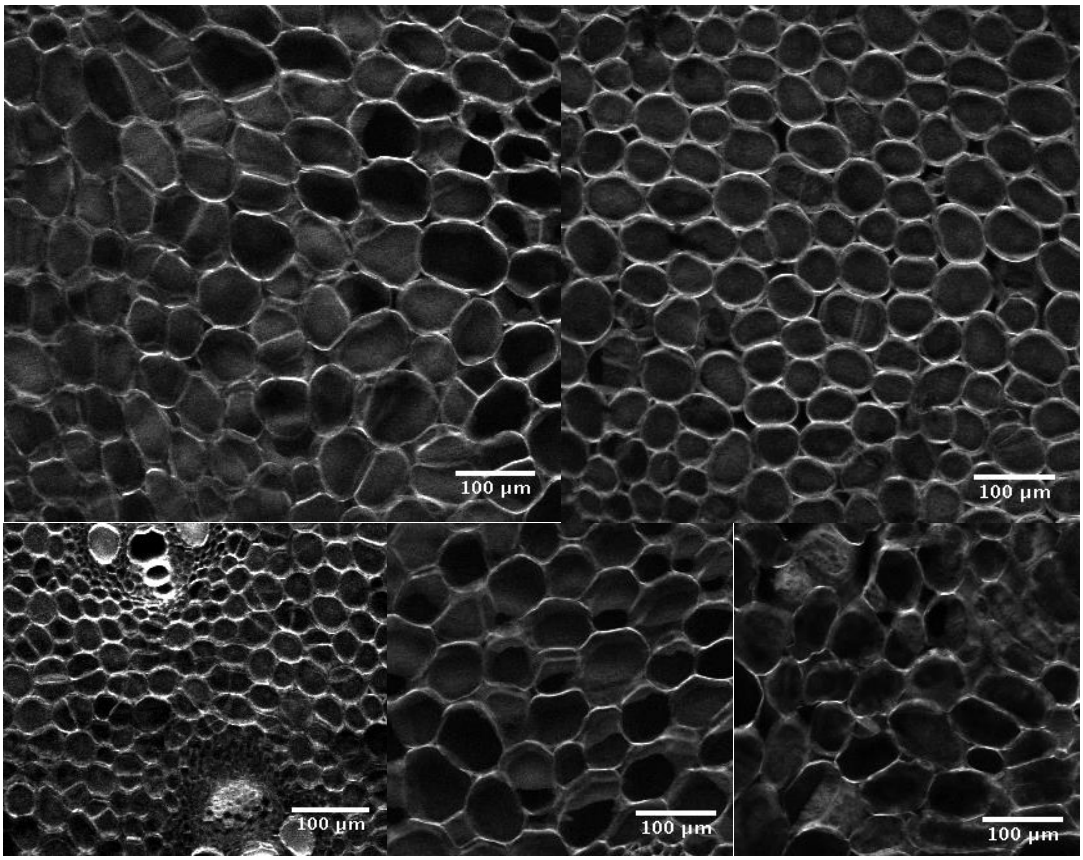

Figure 8. Confocal microscopy of internode transverse cross sections in 58M, SM100 and 100M. 100M INT 14 transverse section (upper left), 100M INT 17 transverse section (upper right), 100M INT 14 transverse section near the rind (lower left), 58M INT 10 transverse (lower center), SM100 INT 9 transverse section (lower right).

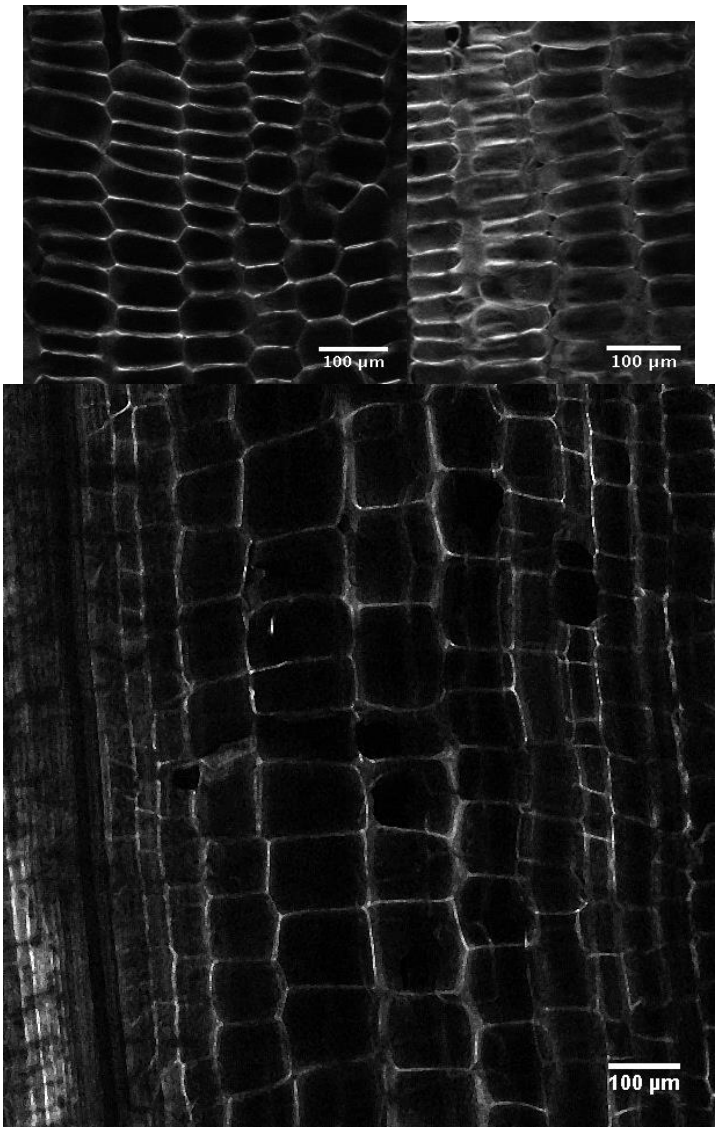

Figure 9. Confocal microscopy of internode longitudinal cross sections in 58M, SM100 and 100M. 100M internode 14 longitudinal (left), SM100 internode 9 longitudinal (right), 58M internode 10 longitudinal (bottom).

I predict that in genotypes with a functional phytochrome B, cell division proceeds primarily near the outer rind of the internode, and that many of the smaller cells near the rind will eventually expand to the size of the parenchyma cells throughout the inner pith tissue. This hypothesis is supported by the fact that the phyB mutation narrowed the variation in cell transverse area primarily by increasing the minimum size of the cells (Figure 10). It should also be noted that the cell division promoting plant hormone cytokinin has been shown previously to inhibit the reversion of phytochrome B to the inactive form in the dark through the action of ARR4 (Fankhauser, 2002). The direct connection between a cell division promoting hormone and phytochrome B activity in *Arabidopsis thaliana* provides some tentative support for a role of phytochrome B in cell cycle regulation in *Sorghum bicolor*.

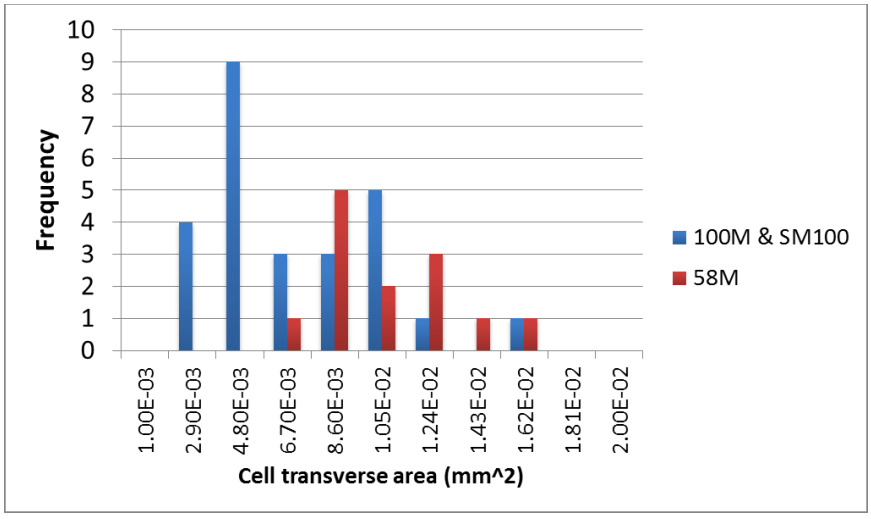

Figure 10. Cell transverse area distributions in the genotypes 100M, SM100 and 58M.

Next, I wanted to determine the relative importance of cell size and cell number in determining the size of mature internodes. However, cell size and number are not independent of each other (Table 1), so I could not simply study the bivariate relationship between each cellular trait and internode size. In order to parse out the different influences of these traits, I took a multiple regression approach. To determine internode morphology based on cell size and cell number, I used the following relationships:

$$H = cH * cL$$

$$A = cA * cT$$

$$V = H * A$$

Where H = internode length (height), cH = average cell height, cL = cell number in the longitudinal plane, A = internode transverse area, cA = average cell transverse area, cT = cell number in the transverse plane, and V = internode volume. In order to perform multiple regression analysis with this data, I first linearized the equations using a log-transformation:

$$\log(H) = \log(cH) + \log(cL)$$

$$\log(A) = \log(cA) + \log(cT)$$

$$\log(V) = \log(cH) + \log(cL) + \log(cA) + \log(cT)$$

log (cell height)	log (trans area)	log (cell # long)	log (cell # trans)	*100M & SM100 Correlations*
	0.65	0.59	-0.32	**log (cell height)**
		0.53	-0.61	**log (trans area)**
			-0.45	**log (cell # long)**
				log (cell # trans)

log (cell height)	log (trans area)	log (cell # long)	log (cell # trans)	*58M Correlations*
	0.75	0.23	-0.88	**log (cell height)**
		0.10	-0.81	**log (trans area)**
			0.14	**log (cell # long)**
				log (cell # trans)

Table 1. Correlations between cell size and cell number in the genotypes 100M, SM100 and 58M.

The relatively high bivariate correlations in Table 1 suggest that there may be multicollinearity between the independent variables in our linear model. It is important to identify these effects, since multicollinearity that is not accounted for in the model will lead to unstable estimates of the model parameters (Graham, 2003). In order to test for genuine multicollinearity between the independent variables, I calculated the variance-inflation factors for each cellular trait (Graham, 2003). Table 2 shows that all the traits in the 100 & SM100 dataset had low enough variance-inflation factors to conclude that the variables were not significantly collinear.

100M & SM100	log (cell height)	log (cell trans area)	log (cell number long)	log (cell number trans)
variance-inflation factors	2.1	2.5	1.8	1.7

58M	log (cell height)	log (cell trans area)	log (cell number long)	log (cell number trans)
variance-inflation factors	9.6	3.5	2.6	13.6

Table 2. Variance-inflation factors for the independent variables in the internode growth linear models.

Tables 3 and 4 show the results of the multiple regressions for internode morphology in genotypes 100M, SM100 and 58M. As was expected, cell size and cell division were both positively correlated with internode growth in all models. In cases where the coefficients for these trends were negative, the p-values of the relationship were not statistically significant. A negative coefficient within these models would translate to a logarithmic decay relationship in the untransformed data. A lack of negative coefficients here suggests that our model may represent biologically meaningful information.

100M+SM100 INT Volume MLR	Coefficient	Std. Error	t-stat	p-value	
(Intercept)	-4.92	1.83	-2.69	1.76E-02	*
log (Cell height)	0.65	0.42	1.57	1.39E-01	
log (Cell transverse area)	1.03	0.22	4.61	4.07E-04	***
log (Cell number long)	1.18	0.17	6.87	7.66E-06	***
log (Cell number trans)	1.27	0.28	4.48	5.20E-04	***

Multiple R-squared	0.93
Adjusted R-squared	0.91
F-statistic	45.98
p-value	6.65E-08

100M+SM100 INT Length MLR	Coefficient	Std. Error	t-stat	p-value	
(Intercept)	-2.43	1.52	-1.60	1.32E-01	
log (Cell height)	0.75	0.35	2.17	4.75E-02	*
log (Cell transverse area)	0.31	0.18	1.70	1.11E-01	
log (Cell number long)	0.95	0.14	6.69	1.02E-05	***
log (Cell number trans)	0.16	0.23	0.67	5.15E-01	

Multiple R-squared	0.92
Adjusted R-squared	0.89
F-statistic	38.77
p-value	1.98E-07

100M+SM100 INT Diameter MLR	Coefficient	Std. Error	t-stat	p-value	
(Intercept)	-1.20	0.37	-3.23	6.11E-03	**
log (Cell height)	-0.05	0.08	-0.57	5.80E-01	
log (Cell transverse area)	0.36	0.05	7.90	1.59E-06	***
log (Cell number long)	0.11	0.03	3.29	5.42E-03	**
log (Cell number trans)	0.56	0.06	9.70	1.36E-07	***

Multiple R-squared	0.91
Adjusted R-squared	0.88
F-statistic	35.24
p-value	3.62E-07

Table 3. Multiple regression of internode growth in 100M and SM100.

58M INT Volume MLR	Coefficient	Std. Error	t-stat	p-value	
(Intercept)	-0.09	0.64	-0.14	8.94E-01	
log (Cell height)	0.71	0.31	2.27	6.41E-02	.
log (Cell transverse area)	0.16	0.30	0.52	6.24E-01	
log (Cell number long)	0.95	0.17	5.54	1.46E-03	**
log (Cell number trans)	-1.61E-03	0.34	-0.01	9.96E-01	

Multiple R-squared	0.97
Adjusted R-squared	0.94
F-statistic	42.33
p-value	1.56E-04

58M INT Length MLR	Coefficient	Std. Error	t-stat	p-value	
(Intercept)	0.93	0.64	1.46	1.95E-01	
log (Cell height)	0.69	0.31	2.22	6.85E-02	.
log (Cell transverse area)	0.17	0.30	0.55	5.99E-01	
log (Cell number long)	0.99	0.17	5.79	1.16E-03	**
log (Cell number trans)	-0.73	0.34	-2.13	7.70E-02	.

Multiple R-squared	0.98
Adjusted R-squared	0.97
F-statistic	81.43
p-value	2.34E-05

58M INT Diameter MLR	Coefficient	Std. Error	t-stat	p-value	
(Intercept)	-4.55E-01	1.47E-01	-3.09	2.13E-02	*
log (Cell height)	7.98E-03	7.24E-02	0.11	9.16E-01	
log (Cell transverse area)	-5.64E-03	7.00E-02	-0.08	9.38E-01	
log (Cell number long)	-2.07E-02	3.95E-02	-0.53	6.19E-01	
log (Cell number trans)	3.66E-01	7.97E-02	4.59	3.72E-03	**

Multiple R-squared	0.98
Adjusted R-squared	0.96
F-statistic	69.48
p-value	3.72E-05

Table 4. Multiple regression of internode growth in 58M.

While independent and dependent variables need not be normally distributed in this analysis, multiple linear regression assumes that the residuals are distributed normally. Figure 11 shows the Quantile-Quantile plots for the normality of the residuals. In general, the volume regressions had the most normally distributed residuals, and regressions incorporating the genotype 58M deviated highly from normality. In particular, the Q-Q plots for length and diameter regressions in 58M both appear to resemble step functions, indicating that the trait data in this genotype is discrete (SAS Institute Inc, 2010). This may reflect both the small number of samples and the low variation within the 58M cellular trait data.

The Q-Q plots in genotypes 100M and SM100 were much closer to normality than in 58M. The slight deviations at the ends of the Q-Q plot in the volume and length regressions are due to shortened tails and deviations in the diameter regression are due to longer tails than that found in normal distributions. Shapiro-Wilk tests for non-normality in these datasets were all non-significant (Table 5). Therefore, we can conclude that the 100M & SM100 regressions have residuals that appear normal, and there is not enough evidence in these data to reject that the residuals come from a normal distribution.

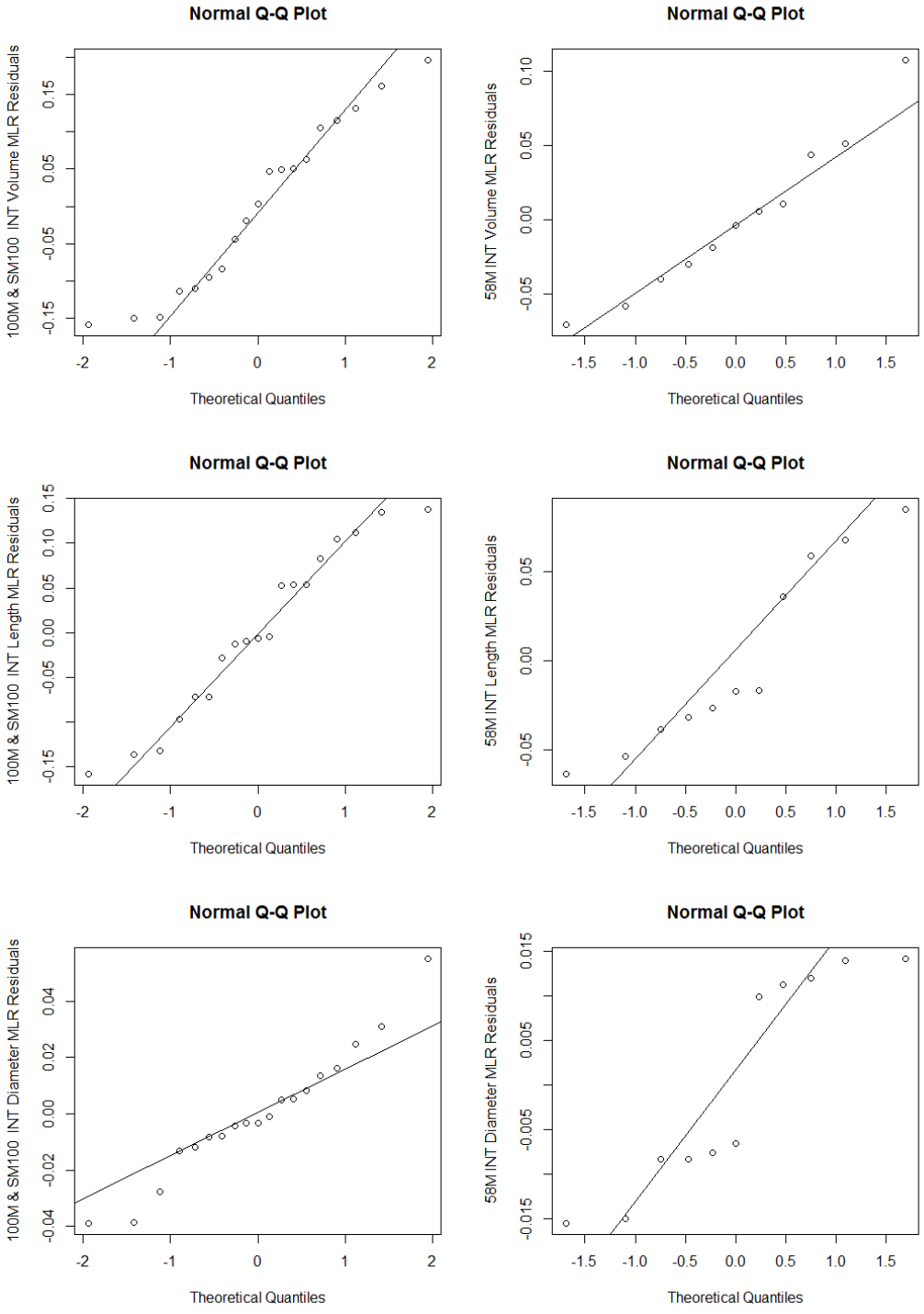

Figure 11. QQ plots for the residuals of the multiple regressions of (100M+SM100) and 58M internode growth.

	W	p-value
100M & SM100 Volume MLR	0.94	2.74E-01
100M & SM100 Length MLR	0.95	3.78E-01
100M & SM100 Diameter MLR	0.96	6.14E-01
58M Volume MLR	0.96	7.82E-01
58M Length MLR	0.89	1.45E-01
58M Diameter MLR	0.83	2.12E-02

Table 5. Shapiro-Wilk tests for deviation from normality

Because some traits within each regression had non-significant p-values, it may be possible to drop these traits from the model without significantly impacting prediction ability. In fact, dropping these potentially irrelevant traits may noticeably improve the predictive capacity of the model. Table 6 shows the results of partial F tests for variable removal in the linear models of 100M & SM100 internode growth. Based on the results in Table 6, I decided to drop variables from each model if the p-value for the ANOVA between the full and reduced model was greater than 0.05.

Regression model	variable dropped	DF	Sum of Squares	F	p-value	
100M & SM100 Volume	cell height	1	4.06E-02	2.46	1.39E-01	
100M & SM100 Volume	cell trans area	1	3.50E-01	21.23	4.07E-04	***
100M & SM100 Length	cell trans area & trans cell number	2	3.36E-02	1.48	2.60E-01	
100M & SM100 Diameter	cell height	1	2.17E-04	0.32	5.80E-01	
100M & SM100 Diameter	longitudinal cell number	1	7.29E-03	10.79	5.42E-03	**

Table 6. Partial F tests to determine whether variables can be dropped from the analysis.

From the reduced models I calculated the expected gains in internode growth based on the selection of the cellular traits. To do this, I fixed each independent variable at the median value, and then tracked the predicted internode growth as one variable approached the maximum value observed in the data. This allowed me to evaluate the influence of each cellular trait on internode growth.

Figure 12 shows the results of this analysis for the 100 & SM100 dataset. Internode volume was approximately equally influenced by the three cellular traits. However, the cellular traits clearly had different effects within the internode length and diameter models. In the internode length model, increasing longitudinal cell number from the median to the maximum value corresponded to a predicted doubling of internode length. However, increasing cell height to the maximum observed value only led to a 1.5-fold increase in internode length. The same pattern was also shown for internode diameter growth, with cell division in the transverse plane changing internode diameter more than any other factor.

Given the trends in 58M, some might find it surprising that cell height does not have a greater impact on internode elongation in 100M & SM100. Though the ma3R mutant, 58M, had highly elongated internodes and highly elongated cells, this clearly had a negative impact on internode volume and biomass. This tradeoff, and the various other pleiotropic effects of the allele, means that the ma3R phenotype should be treated as a unique case and not necessarily as representative of the species as a whole.

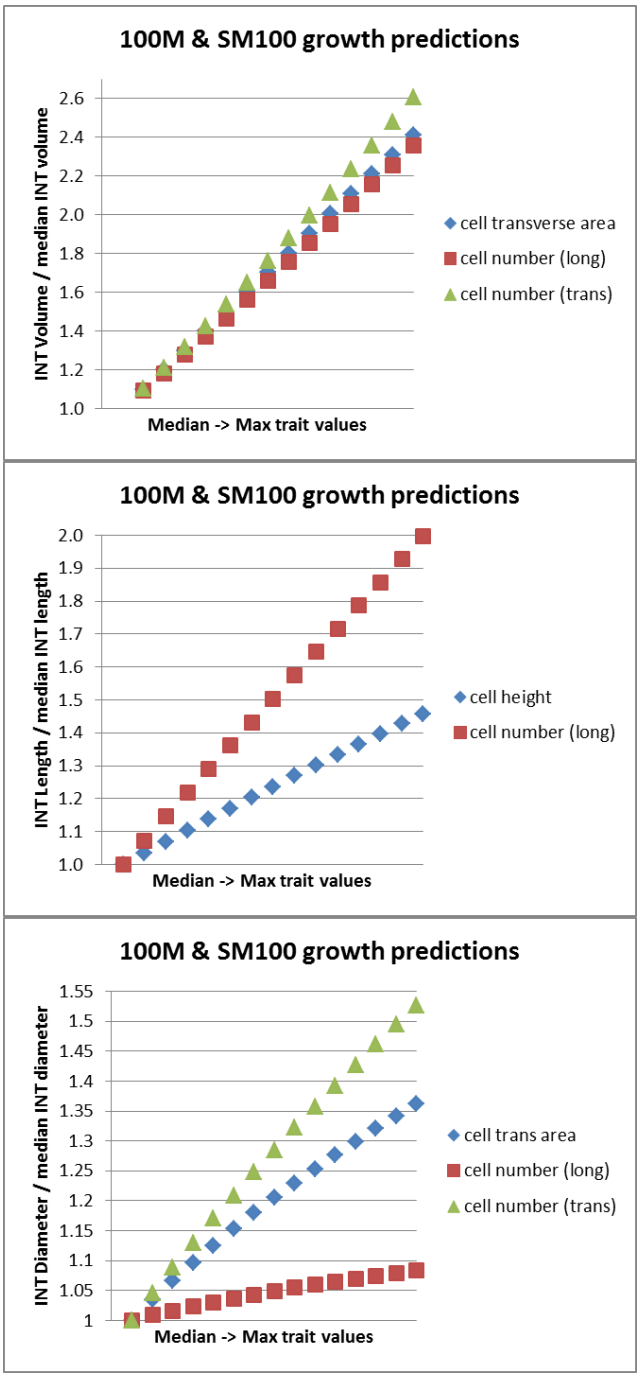

Figure 12. Graphical representation of the linear models of stem growth. Values for cell size and cell number were chosen to bracket the median and maximum values observed in the real data.

In discussions on plant organ size control, many potential factors have been proposed to play important roles. In particular, wall loosening enzymes and cell division promoting transcription factors have been labeled as highly significant to organ growth (Bögre *et al*., 2008). From the internode growth models, it is clear that changing the extent of cell division has a greater impact on internode growth than changing the degree of cellular expansion. Therefore, breeders working within sorghum or similar species should select parents to maximize the expression of cell division promoting genes in order to increase biomass yields. Examples of regions to select for allelic variation would include the WOX gene family members *SbWOX1-11* (Zhang *et al*., 2010) and potential GRAS transcription factors (Bolle *et al*., 2004; Perez-Rodriguez *et al.*, 2009).

.

5. INTERNODE GROWTH GENE EXPRESSION IN *S. BICOLOR*

The results from Sections 3 & 4 give us a sense of the extent, timing and cellular basis for stem growth in *Sorghum bicolor*. The next logical step in this analysis was to survey the expression of all genes in the stem as it developed over time. This whole transcriptome survey would provide us with a list of candidate transcripts to validate and examine their role in the stem development program. Because the sorghum stem sets down new, morphologically identical internodes iteratively, I chose to sample a series of internodes in the stem at the same time to survey gene expression in the stem across development.

For this study field grown genotype ES5200 was sampled at two different harvest dates, at approximately 100 days after planting. Internode diameter ranged from 10mm at the apical dome to 31mm in the third elongating internode and internode length ranged from 12mm in the apical dome to 73mm at the third elongating internode. Internode length in the third elongating internode represented 72% of the maximum internode length observed and diameter in this internode represented 74% of the maximum internode diameter in this experiment.

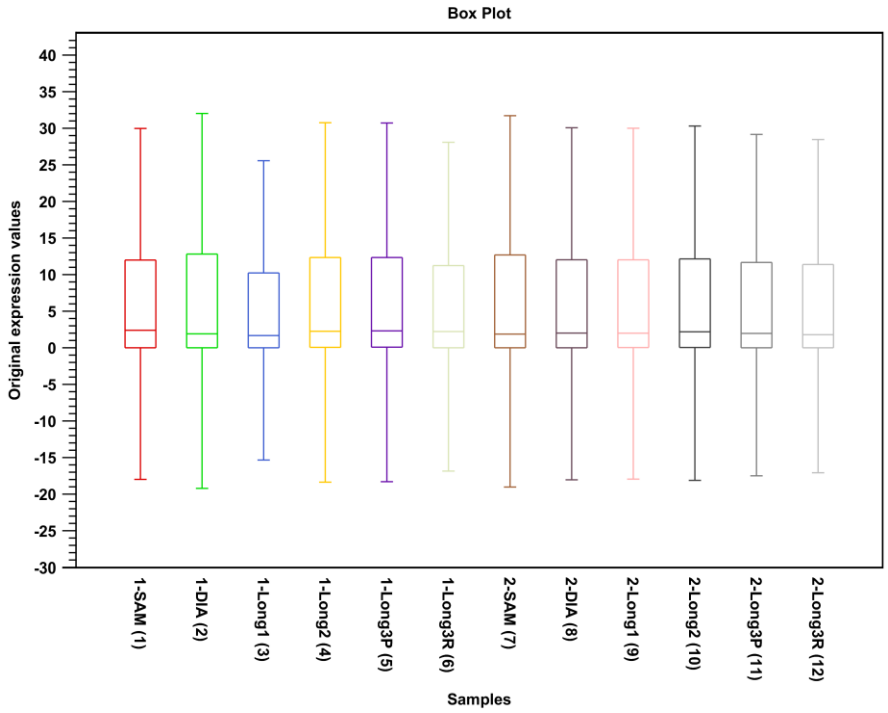

Figure 13. Total gene expression within each sample of field-grown ES5200. Sequencing data was analyzed using the CLC genomics workbench.

Figure 13 shows the total transcript levels within each sample. The boxplots show that the average level of gene expression is approximately the same across all samples, so gene expression was only normalized by the reads per kilobase of transcript per million reads (RPKM). Figure 14 shows the ordination of the first two principle coordinates in the Principle Coordinate Analysis (PCA) of gene expression variation within samples. Samples 7-12 showed good separation along both PC1 and PC2, with a pattern that agrees with the pattern of developmental age of the internodes. Samples 1, 3 and 6 showed an unexpected distribution in the ordination. In addition, Samples 1, 3 and

6 had narrow distributions of gene expression (Figure 13) and had high percentages of unmapped reads in the genome alignments. Because of their relatively low quality, these samples were removed from further analyses. In addition, samples 2, 4 and 5 were only included in analyses as validation samples.

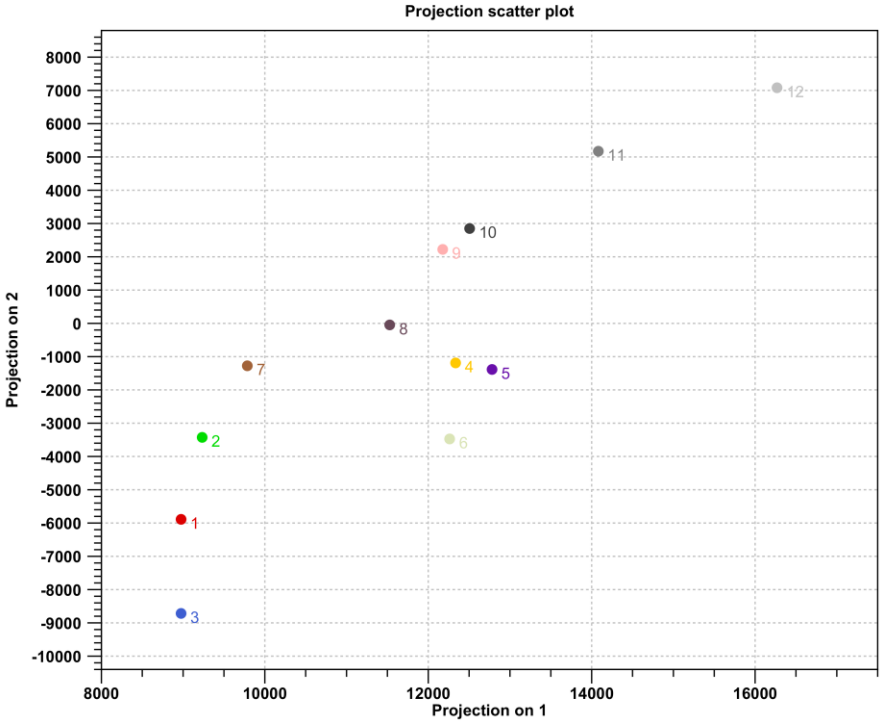

Figure 14. Ordination of gene expression principle coordinates 1 & 2 for samples of field-grown ES5200. SAM = 1 & 7, DIA = 2 & 8, LONG1 = 3 & 9, LONG2 = 4 & 10, LONG3(pith) = 5 & 11, LONG3(rind) = 6 & 12. Sequencing data was analyzed using the CLC genomics workbench.

Figures 15 & 16 show PageMan analyses of gene expression in samples across internode development and in the pith and rind tissue, respectively. MapMan ontologies were ranked by the number of expressed transcripts and compared to the median expression across all ontologies using the Wilcoxon rank-sum test. Type I error inflation from multiple testing was corrected for using the Bonferoni correction, which multiplies the calculated p-value by the number of tests conducted. In PageMan tables, the color score represents the average fold change in gene expression for all transcripts within a significantly enriched ontology (Usadel *et al.*, 2006).

In Figure 15 we see that gene expression related to photosynthesis increases during internode development. Initially, the young internode in completely enclosed in multiple layers of leaf sheaths. As the internode grows, it gradually elongates and expands past each layer of leaf sheath, until most of the internodes in this genotype are illuminated and photosynthetic. Therefore, an increase in photosynthetic gene expression over time makes sense.

The general pattern of gene expression throughout development is a relatively high level of primary nucleotide and protein synthesis during the early phases of growth and a transition to higher relative levels of secondary metabolism and cell wall synthesis later on in internode development. In addition, cell cycle and chromatin related gene expression are the highest in the apical dome and in the young internodes. Again, these observations agree with my expectations of the developmental pattern of internode growth.

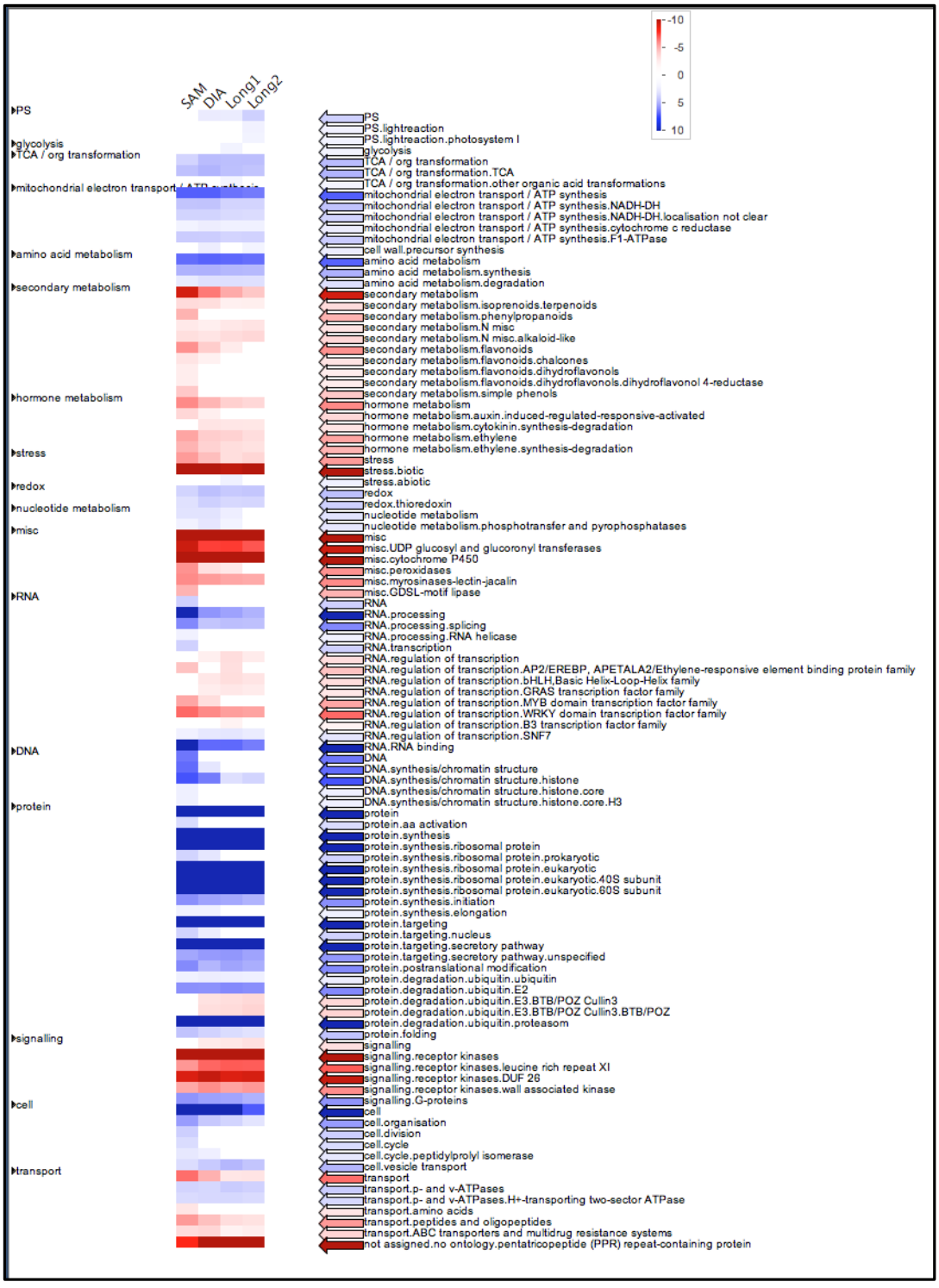

Figure 15. PageMan analysis of gene expression across internode development in field grown ES5200. Sequencing data was analyzed using the CLC genomics workbench.

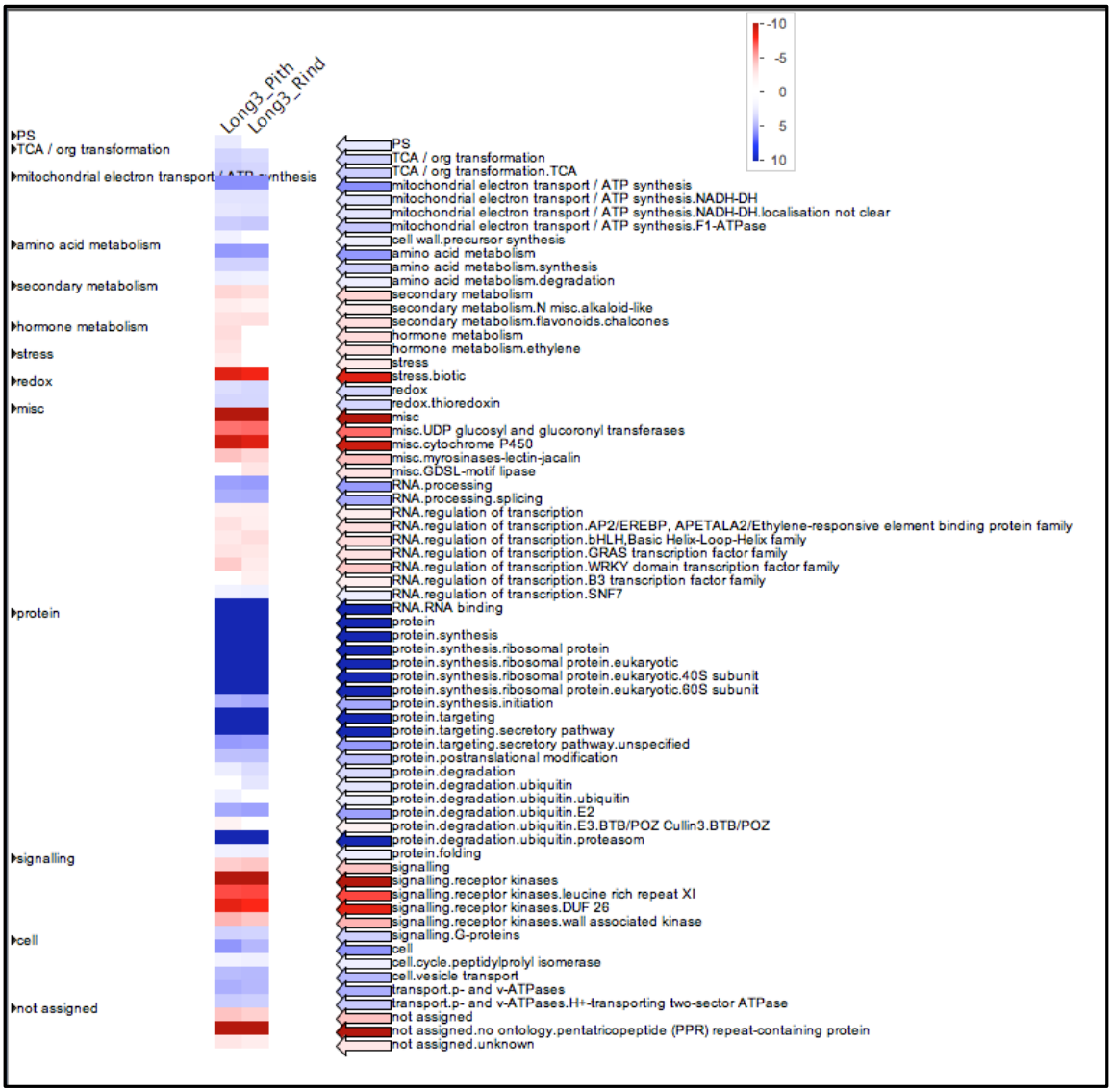

Figure 16. PageMan analysis of gene expression differences in pith and rind tissue from internodes of field grown ES5200. Sequencing data was analyzed using the CLC genomics workbench.

Interestingly, hormone metabolism in general appears to be down regulated in the young internodes, with auxin and ethylene related genes being lowly expressed. As a stress-related hormone, the ethylene down-regulation in young tissue agrees with the

general pattern of reduced stress response in early internode development in this data. The reduced auxin-related gene expression in young tissue seems surprising, given the gradient of auxin concentration in plants that is highest at the apical dome. However, it is important to note that only auxin-response genes appear to be significantly down regulated in young tissue in this dataset, indicating that this is perhaps a counterbalance to the high auxin concentration in this tissue.

There appear to be two distinct sets of transcription factors (TFs) with differential gene expression across internode development. Transcripts orthologous to bHLH, GRAS and B3 transcription factors appear to be up regulated near the apical dome, while transcripts corresponding to AP2/EREBP, MYB and WRKY transcription factors appear to be down regulated early on in internode development. The pattern of these transcripts in particular is interesting not only for the high impact of transcription factor expression on organ development, but also for future studies on transcription factor-gene promoter binding studies.

AP2/EREBP transcription factors are known to play important roles in plant hormone regulation and in response to environmental stress (Feng *et al*., 2005). MYB-related transcription factors have widely varying functions in plants, though many control the expression of genes related to secondary metabolism (Jin & Martin, 1999). The WRKY transcription factor superfamily is also a diverse group, with 100 known WRKY genes in *Arabidopsis thaliana* (Eulgem *et al*., 2000). Though this is a diverse group, most of the WRKY genes that have been observed are up regulated during senescence, biotic and abiotic stress. Given these known functions in other plant

species, it makes sense that these three transcription factor families would increase in expression as the internode ages in *Sorghum bicolor*.

GRAS transcription factors have been shown to be involved in meristem maintenance, gibberellin signal transduction and phytochrome signaling in Arabidopsis (Bolle, 2004). Therefore, it makes sense that a subset of GRAS TFs would be highly expressed near the apical dome. bHLH and B3 transcription factors have been shown to have a variety of possible functional roles in plants, and must be analyzed at the transcript level in order to evaluate their biological significance. The bHLH transcript with the highest fold change throughout internode development in this dataset corresponds to gene sb10g030470. This transcript displayed a sharp decrease in expression past the apical dome, with a 11-fold decrease in RPKM between the apical dome and the second elongating internode. The putative orthologue of this gene in Arabidopsis, At5g65640, has been shown previously to encode a bHLH TF that physically interacts with FAMA, a bHLH TF that regulates cell proliferation during stomatal development (Ohashi-Ito & Bergmann, 2006).

The B3 transcription factor with the highest fold change throughout development in this dataset corresponded to gene sb03g033260. This transcript also displayed a sharp decrease in expression outside of the apical dome, with a 9-fold decrease in expression between the apical dome and the second elongating internode. This gene is weakly similar to Arabidopsis AT3G18990, also known as Vernalization1 (*vrn1*). VRN1 was initially characterized for its role in the induction of flowering during prolonged periods of cold temperature (Levy *et al*., 2002). In addition, when the homologue of this gene in

wheat was rendered inactive, the mutant plants were unable to transition from the vegetative to the floral phase of growth (Shitsukawa *et al*., 2007). Though almost all studies on VRN1 focus exclusively on the vernalization phenotype, a recent study has highlighted the numerous other roles this transcription factor plays in *Arabidopsis thaliana* growth and development once genetic redundancy is accounted for (King *et al*., 2013).

Table 7 shows a list of transcripts that were found to have statistically significant differences in gene expression across internode development in both the CLC genomics and Tuxedo suite pipeline. Many of these transcripts show homology with genes of known function in *Arabidopsis thaliana*. One of the transcripts in Table 7, *rve1*, is known to control circadian oscillation of auxin synthesis in Arabidopsis (Rawat *et al*., 2009). Interestingly, *rve1* expression in Arabidopsis is highest in the morning, at approximately the same time that I sampled tissue for gene expression in this experiment.

Various other interesting trends can be gleaned from the significant transcripts shown in Table 7. A transcript with homology to a known ethylene zinc finger transcription factor indicates that ethylene signaling may be important to timing internode growth in *Sorghum bicolor*. This is supported by the known relationship between ethylene, the auxin transporter *aux1*, echidna, and cell elongation in *Arabidopsis thaliana* (Boutté *et al*., 2013). Table 7 also includes a transcript with sequence homology to the MYB transcription factor family, further indicating that various MYB transcription factors may be important in Sorghum stem tissue growth.

Gene ID	MapMan ontology	Description
sb01g002300	lipid metabolism.lipid degradation.lipases.triacylglycerol lipase	moderately similar to (282) AT4G16070 | Symbols: | lipase class 3 family pr
sb01g004600	not assigned.unknown	highly similar to (813) AT5G49830 | Symbols: | EXPRESSED IN: 22 plant stru
sb01g013700	signalling.G-proteins	moderately similar to (452) AT4G05410 | Symbols: | transducin family prot
sb01g030500	metal handling.binding, chelation and storage	weakly similar to (112) AT2G18196 | Symbols: | metal ion binding | chr2:7
sb01g050300	RNA.regulation of transcription.putative transcription regulator	weakly similar to (147) loc_os03g01540 12003.m35547 protein DNA bindin
sb02g002900	OPP.electron transfer	highly similar to (594) AT4G05390 | Symbols: ATRFNR1 | ATRFNR1 (ROOT F
sb02g010700	not assigned.unknown	moderately similar to (270) loc_os01g42340 12001.m10500 protein conserv
sb02g023300	RNA.regulation of transcription.C2H2 zinc finger family	weakly similar to (149) AT3G28210 | Symbols: PMZ | PMZ; zinc ion binding
sb02g023300	RNA.regulation of transcription.unclassified	weakly similar to (149) AT3G28210 | Symbols: PMZ | PMZ; zinc ion binding
sb02g034800	protein.degradation.AAA type	very weakly similar to (88.6) AT2G18190 | Symbols: | AAA-type ATPase fam
sb02g038500	not assigned.no ontology	highly similar to (520) AT5G15550 | Symbols: | transducin family protein /
sb02g042000	RNA.regulation of transcription.PHD finger transcription factor	highly similar to (573) AT5G12400 | Symbols: | PHD finger transcription fac
sb03g005800	signalling.receptor kinases.leucine rich repeat XI	highly similar to (637) AT3G47570 | Symbols: | leucine-rich repeat transme
sb03g017600	PS.lightreaction.photosystem II.PSII polypeptide subunits	highly similar to (917) ATCG00680 | Symbols: PSBB | encodes for CP47, sub
sb03g045900	stress.abiotic.drought/salt	highly similar to (722) AT3G54510 | Symbols: | early-responsive to dehydra
sb04g028600	transport.misc	moderately similar to (351) AT3G52640 | Symbols: | nicastrin-related | chr
sb04g028600	transport.misc	moderately similar to (356) AT3G52640 | Symbols: | nicastrin-related | chr
sb04g030200	RNA.regulation of transcription.unclassified	highly similar to (571) AT1G07360 | Symbols: | zinc finger (CCCH-type) fam
sb04g036200	not assigned.no ontology	highly similar to (647) AT5G34940 | Symbols: AtGUS3 | AtGUS3 (Arabidopsi
sb04g038300	not assigned.unknown	weakly similar to (126) AT2G22640 | Symbols: HSPC300, BRK1 | BRK1 (BRIC
sb05g003200	development.unspecified	moderately similar to (262) AT2G27550 | Symbols: ATC | ATC (ARABIDOPSIS
sb05g004600	not assigned.unknown	moderately similar to (243) AT4G14590 | Symbols: emb2739 | emb2739 (e
sb05g019800	redox.thioredoxin	weakly similar to (155) AT5G51010 | Symbols: | rubredoxin family protein
sb06g018100	minor CHO metabolism.others	moderately similar to (300) AT1G59960 | Symbols: | aldo/keto reductase,
sb06g018400	not assigned.unknown	weakly similar to (103) loc_os04g37760 12004.m08765 protein expressed p
sb06g024800	misc.protease inhibitor/seed storage/lipid transfer protein (LTP) family protein	very weakly similar to (81.6) AT4G12480 | Symbols: pEARLI 1 | pEARLI 1; lipi
sb06g025600	RNA.regulation of transcription.G2-like transcription factor family, GARP	very weakly similar to (92.8) AT1G14600 | Symbols: | DNA binding / transcr
sb06g025700	minor CHO metabolism.others	weakly similar to (154) AT5G52190 | Symbols: | sugar isomerase (SIS) dom
sb06g025900	RNA.regulation of transcription.AP2/EREBP, APETALA2/Ethylene-responsive elem	weakly similar to (128) AT5G51990 | Symbols: CBF4, DREB1D | CBF4 (C- RE
sb06g026000	misc.cytochrome P450	moderately similar to (489) AT2G45510 | Symbols: CYP704A2 | CYP704A2;
sb06g026100	protein.degradation.ubiquitin	moderately similar to (394) AT2G21270 | Symbols: | ubiquitin fusion degra
sb06g026100	protein.degradation.ubiquitin	moderately similar to (394) AT2G21270 | Symbols: | ubiquitin fusion degra
sb06g026500	RNA.regulation of transcription.MYB-related transcription factor family	weakly similar to (112) AT5G37260 | Symbols: RVE2, CIR1 | RVE2 (REVEILLE
sb06g026700	development.unspecified	very weakly similar to (88.6) loc_os04g49670 12004.m09900 protein conser
sb06g033800	stress.abiotic.heat	very weakly similar to (90.1) AT2G18465 | Symbols: | DNAJ heat shock N-te
sb07g003600	OPP.non-reductive PP.transaldolase	moderately similar to (483) AT1G12230 | Symbols: | transaldolase, putativ
sb07g003700	not assigned.no ontology	moderately similar to (216) AT4G00170 | Symbols: | vesicle-associated me
sb07g003700	not assigned.no ontology	moderately similar to (204) AT4G00170 | Symbols: | vesicle-associated me
sb07g006400	not assigned.unknown	weakly similar to (147) AT1G16825 | Symbols: | FUNCTIONS IN: molecular_
sb07g025800	protein.degradation.ubiquitin.E3.SCF.FBOX	moderately similar to (249) loc_os02g06520 12002.m06000 protein F-box d
sb08g021800	RNA.regulation of transcription.C2H2 zinc finger family	very weakly similar to (92.4) loc_os12g42250 12012.m08004 protein expres
sb08g023300	not assigned.unknown	moderately similar to (270) AT1G29700 | Symbols: | unknown protein | ch
sb09g003400	mitochondrial electron transport / ATP synthesis.cytochrome c oxidase	highly similar to (566) AT1G57600 | Symbols: | membrane bound O-acyl tr
sb09g016600	signalling.G-proteins	weakly similar to (195) AT5G58590 | Symbols: RANBP1 | RANBP1 (RAN BIN
sb09g021900	protein.synthesis.ribosomal protein.eukaryotic.60S subunit.unknown	no original description
sb09g024600	RNA.regulation of transcription.bZIP transcription factor family	moderately similar to (328) AT1G05270 | Symbols: | TraB family protein |
sb09g027800	RNA.regulation of transcription.MYB domain transcription factor family	moderately similar to (219) AT5G57620 | Symbols: MYB36, AtMYB36 | MYB
sb09g028400	stress.abiotic.drought/salt	weakly similar to (140) AT3G05700 | Symbols: | INVOLVED IN: response to
sb10g007000	protein.degradation.ubiquitin.E3.RING	weakly similar to (124) AT1G60360 | Symbols: | zinc finger (C3HC4-type RI
sb10g018300	not assigned.unknown	moderately similar to (335) AT3G18170 | Symbols: | transferase, transferri
sb10g022200	not assigned.unknown	weakly similar to (200) AT2G26040 | Symbols: | Bet v I allergen family prot
sb10g028100	DNA.synthesis/chromatin structure	highly similar to (682) AT5G44750 | Symbols: REV1, ATREV1 | REV1; DNA-d
sb10g028500	misc.peroxidases	moderately similar to (335) AT2G18980 | Symbols: | peroxidase, putative |

Table 7. Transcripts with significant differences in expression across internode development in both the CLC genomics and Tuxedo suite pipelines.

Recently, researchers discovered that wax precursor very long chain fatty acids directly inhibit cytokinin production and cell division in *Arabidopsis thaliana* (Nobusawa *et al.*, 2013). It would be interesting to see if the apparent triacylglycerol lipase in Table 7 has a similar effect on cytokinin levels and cell division in Sorghum

internodes. It would also be interesting to perturb the levels of the apparent ubiquitin E3 ligases shown in Table 7. If changing the expression of these genes in Sorghum has a significant effect on stem growth it would indicate that protein turnover plays a significant role in Sorghum stem growth and morphology. Clearly, there is a great deal of work remaining to be done within this system.

Figure 16 shows the differences in gene expression between pith and rind tissue in the third elongating internode sampled from field-grown ES5200. This dataset had some surprising trends, such as increased gene expression in the pith related to photosynthesis, stress, ethylene production and cell wall precursor synthesis. These observations were the opposite of what I had expected, given the biology of the internode. Part of the explanation for this odd pattern of expression may be that rind tissue is at an earlier stage of development than the pith tissue. Though trends in this dataset are weaker in general, the rind tissue appears to be more closely related to the apical dome and young internodes, while the pith tissue is more similar in gene expression to the older internodes. Examples of such trends include down-regulation of ubiquitin-related genes in older internodes and in the pith tissue and increased secondary metabolism in both old internodes and pith tissue. In addition, the same transcription factors and receptor kinases appear to be down regulated in both older internodes and pith tissue. Though differences in cell division were not observed between pith and rind tissue, given these results I would expect that active cell division is more likely to occur near the outer rind rather than within the center of the internode.

One group of genes that has been consistently shown to change expression throughout stem development in this data is the cytochrome P450 Monooxygenases (Figures 15 & 16, Table 7). Enzymes of this group are involved in a multitude of cellular processes, including the hydroxylation of phenylpropaniods, fatty acids and plant hormones (Nelson & Werck-Reichhart, 2011). Seven transcripts with differential gene expression across development that appear homologous to known P450 Monooxygenases in other plant species are shown in Figure 17. If these transcripts have true functional equivalency in sorghum, they would be very interesting to follow up on with additional molecular studies. In particular, it would be interesting to see whether perturbing CYP86 & CYP94 gene expression changes wax deposition and stem diameter growth, as was observed for very-long-chain fatty acid synthesis in Arabidopsis (Nobusawa *et al*., 2013). The importance of p-coumaric acid and ferulic acid to cell wall biochemistry is discussed in Section 6.

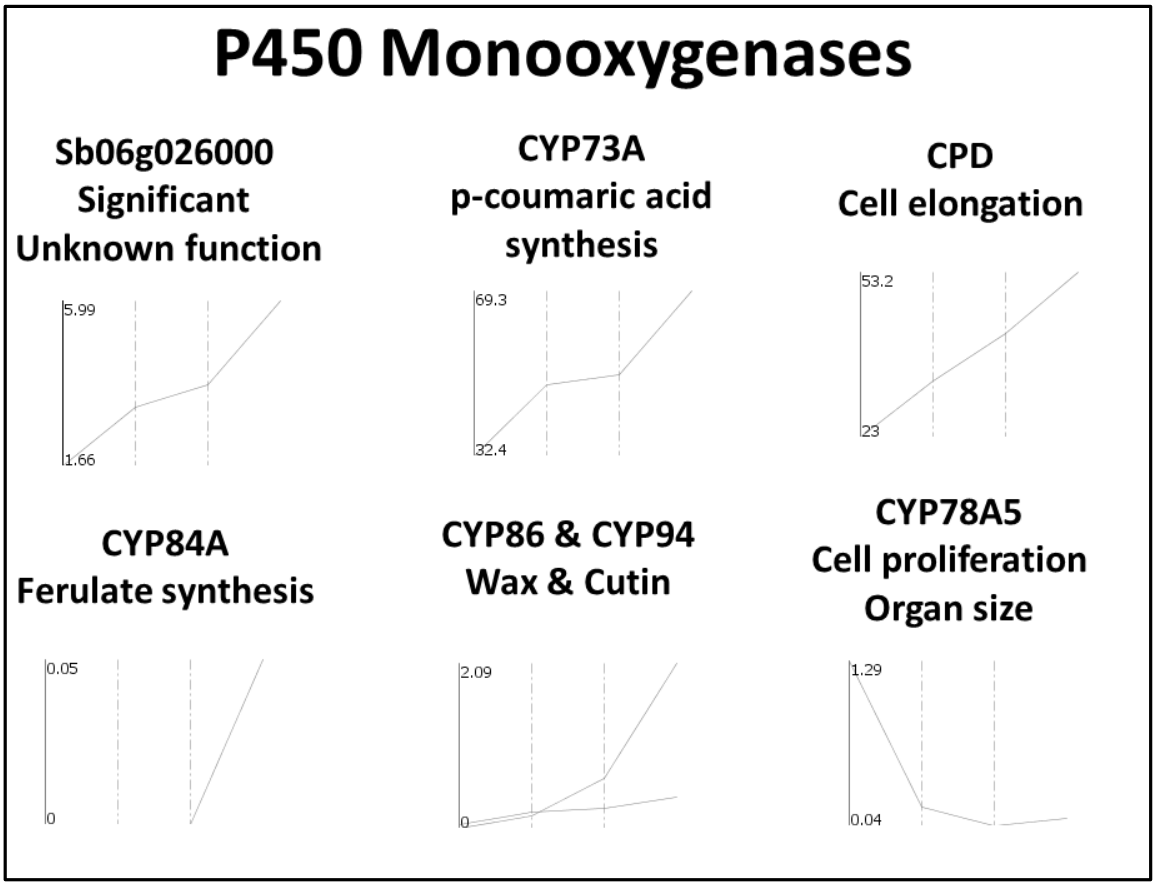

Figure 17. Seven cytochrome p450 monooxygenase homologues in Sorghum bicolor that are differently expressed across internode development. In each graph, internode age increases from left to right. Sequencing data was analyzed using the CLC genomics workbench.

Figure 18 shows a hypothetical model of the various stem growth pathways that may be operating in *Sorghum bicolor*. I created this model based upon the results of various gene expression studies in *Arabidopsis thaliana* (Powell & Lenhard, 2012; Sablowski, 2009; Fankhauser, 2002; Ikeda *et al*., 2012; Boutté *et al*., 2013; Nobusawa *et al*., 2013; Efroni *et al*., 2013). Overlaid upon this model are the gene expression profiles

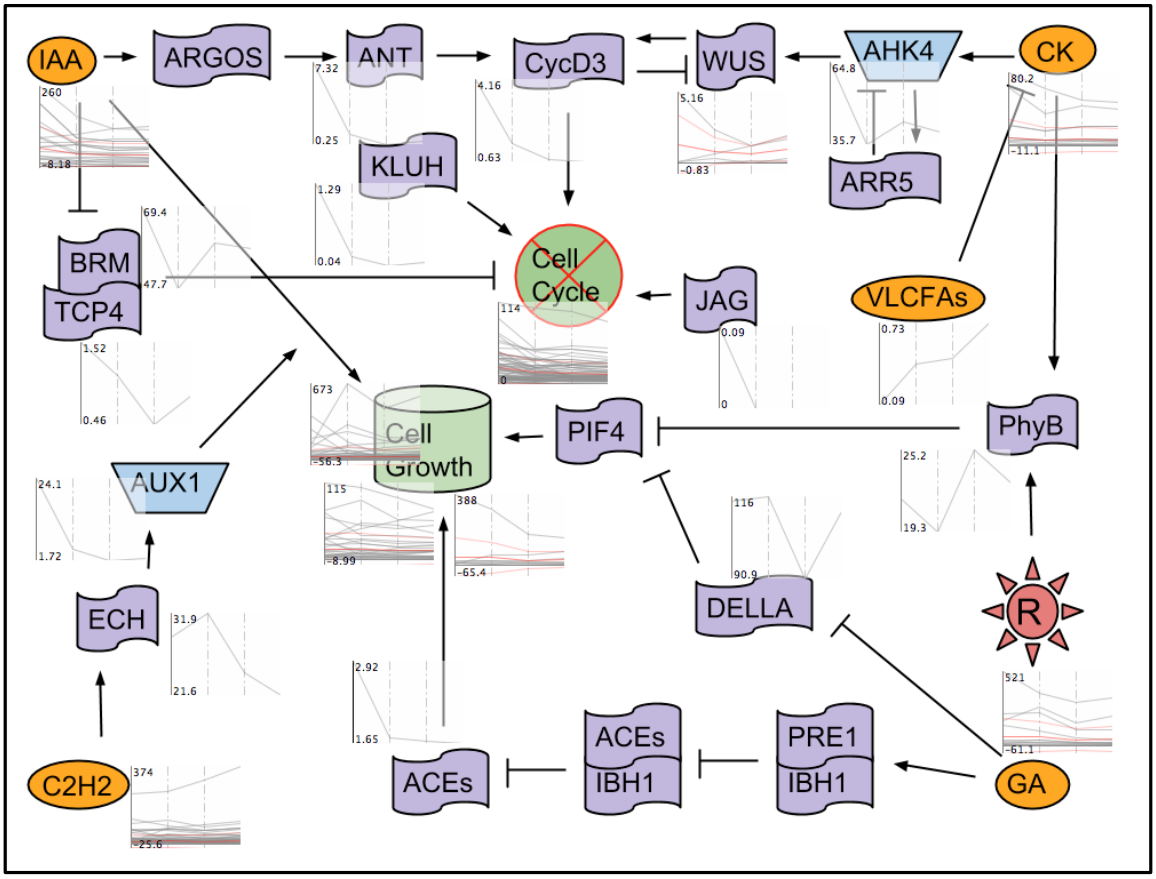

Figure 18. Hypothetical pathways influencing stem growth in Sorghum bicolor. All acronyms are from the gene names in the model plant species Arabidopsis thaliana. Transcription factors are shown in purple, receptor kinases in blue, hormones and fatty acids in orange and red light signals are shown in red. MapMan graphs of homologous gene expression across internode development are displayed next to the corresponding factor. In each graph, internode age increases from left to right. Red lines in each graph represent the average pattern of expression of all the transcripts shown. Sequencing data was analyzed using the CLC genomics workbench.

across internode development of sorghum transcripts that have homology to these genes. As I expected, transcripts related to auxin, cytokinin and gibberellin metabolism are highest in expression near the shoot apex, and their expression levels decrease as tissue growth decreases. In addition, the relationships that have been observed between cytokinin, wushel and cyclin D3 in the Arabidopsis shoot apical meristem appear to act in sorghum stem development in a similar manner (Sablowski, 2009).

Some transcripts show very little change in expression, and have very low abundance in general. These transcripts, such as the supposed JAG homologue, might not function in stem growth, and the ontological information may be incorrect. In other instances, such as with PhyB, changes in transcript abundance per se may be of minimal importance and the crucial control points may be post-transcriptional. As is the case with all the analyses in this section, targeted gene perturbation studies should be undertaken to test the hypothetical relationships presented here.

6. CELL WALL COMPOSITION AND SACCHARIFICATION EFFICIENCY

Two genetic mapping populations were evaluated for biochemical conversion efficiency of lignocellulosic biomass to monosaccharides. The E-SAP panel is a collection of diverse accessions of *Sorghum bicolor*, originally selected for their potential use in breeding sorghum inbreds for energy hybrid agriculture. The BTx642*Tx7000 population is a collection of recombinant inbred lines (RIL) created from a cross between the two genotypes for genetic mapping of QTL for various agronomic traits. Dried, ground stem tissue from these populations was analyzed for enzymatic conversion efficiency by Dr. Nicholas Santoro at the Great Lakes Bioenergy Research Center (GLBRC) using the iWALL system (Santoro *et al*., 2010).

Prior to conversion analysis, ground stem tissue was first analyzed nondestructively for cell wall composition using near-infrared spectroscopy (NIR) (Wolfrum *et al.*, 2013). From the NIR data, I have estimates of the percent of cellulose and xylan in the ground tissue prior to conversion. In addition, from the iWALL results I have data on the concentration of glucose and xylose released, expressed as a percent of the starting material. To calculate the percent yield of the conversion of polymer to monomer, I used the following relationships:

MW Glucose = 180.16 g/mol

(MW Glucose) – (MW endo-monomer of Cellulose) = 18 g/mol [H+OH]

Mean cellulose length in corn stover = 7000 monomer units [Kumar et al., 2009]

MW of cellulose = (180.16 [g/mol] * 7000) – (6999 * 18 [g/mol])

MW of cellulose = 1,135,138 g/mol

% Glucose product → g Glucose Product → moles Glucose Product

% Cellulose → g Cellulose → moles Cellulose

% Yield = (moles Glucose) / (Cellulose chain length * moles Cellulose)

I used the same relationships as above to calculate percent yield in the conversion of xylan to xylose units. The xylose yield calculation has various uncertainties, given that xylan chains can vary widely in chain length, branching and monosaccharide composition (Burton *et al*., 2010). I estimate here an average xylan chain length of 140 units (Gatenholm *et al*., 2003) and assume a straight chain of solely xylose units. This rough approximation was suitable for my purposes.

Figures 19 & 20 summarize the results of the iWALL digestion for the E-SAP panel and the BTx642*Tx7000 RILs, respectively. The diverse set of material in the E-SAP panel show a wide range of glucose (17-51%) and xylose (9-23%) yields (Figure 19, *center*). In general, the BTx642*Tx7000 RILs had higher digestibility compared to the E-SAP panel (32% vs. 27% glucose release; 21% vs. 15% xylose release) (Figures 19 & 20, *center left & center right*). Accounting for differences in the range of conversion efficiencies observed, the E-SAP panel and the BTx642*Tx7000 RILs displayed similar variation in glucose yield with respect to xylose yield (Figures 19 & 20, upper). This data suggests that both populations have extensive variation in cell wall

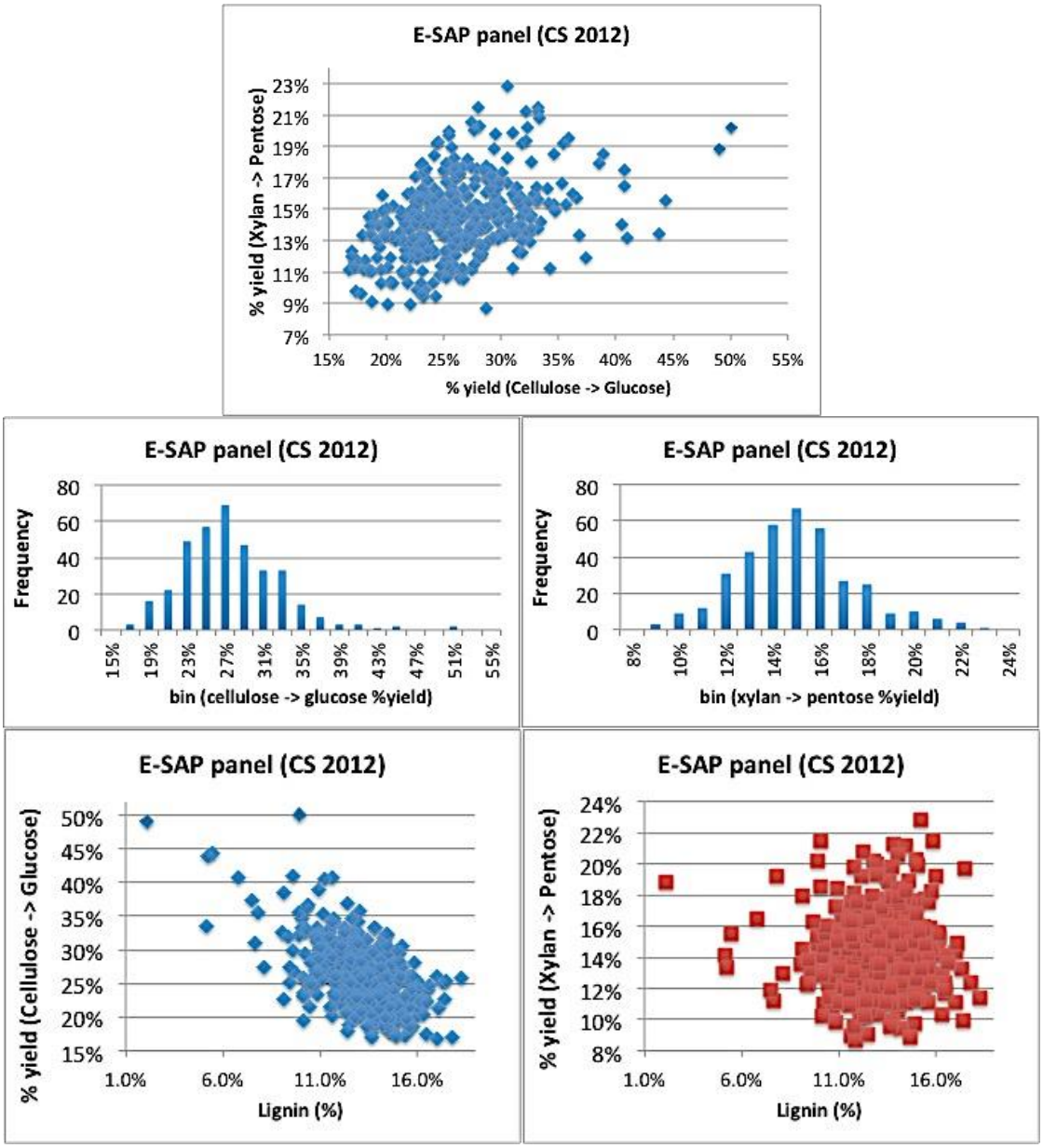

Figure 19. Variation in saccharification efficiency of the E-SAP panel via conversion on the iWALL platform.

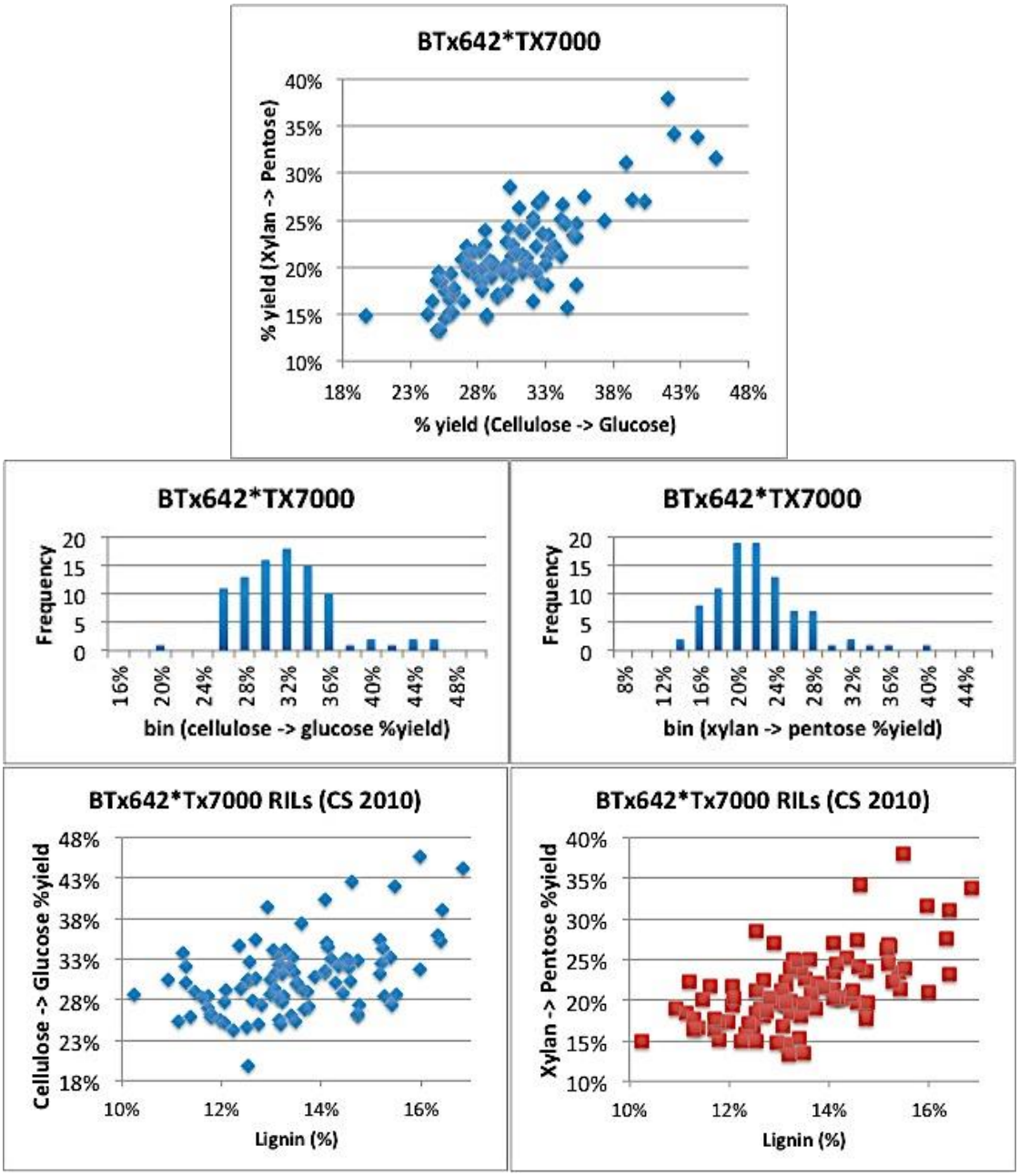

*Figure 20. Variation in saccharification efficiency of the BTx642*Tx7000 biparental mapping population via conversion on the iWALL platform.*

architecture that contributes to conversion efficiency.

Curiously, percent lignin had no statistically significant correlation to monosaccharide yield in the iWALL conversion system (Figures 19 & 20, lower left & lower right). Though the E-SAP panel appears to show an anticorrelation between glucose yield and percent lignin, the best fit linear regression had an R2=0.35. I found similar R2 values for the linear regressions of monosaccharide yield by percent lignin in the BTx642*Tx7000 population, however in these cases the trend was toward increasing yield at higher lignin concentrations.

To further dissect the cause of differential cell wall conversion, I sent samples to Dr. Laura Bartley at the University of Oklahoma for analysis of p-coumaric acid and ferulic acid biomass composition (Bartley *et al.*, 2013). P-coumaric acid is a phenolic compound that is incorporated into lignin in the secondary cell wall in plants (Ralph *et al.*, 1994), where it may serve as an attachment molecule between lignin and the substituted arabinoxylan of hemicelluloses. Ferulic acid is esterified to glucuronoarabinoxylans and can form dimers, trimers and tetramers with itself (Harris & Trethewey, 2010). Ferulic acid may also help to crosslink hemicelluloses to lignin in the secondary cell wall. Though ferulic acid and p-coumaric acid are both lowly abundant in sorghum biomass (Figure 21), they may have a relatively high impact on the accessibility of cell wall polymers to enzymatic conversion reactions. As Figure 21 shows, however, variation in ferulic acid and p-coumaric acid concentration do not correlate with variation in biomass conversion efficiency on the iWALL system.

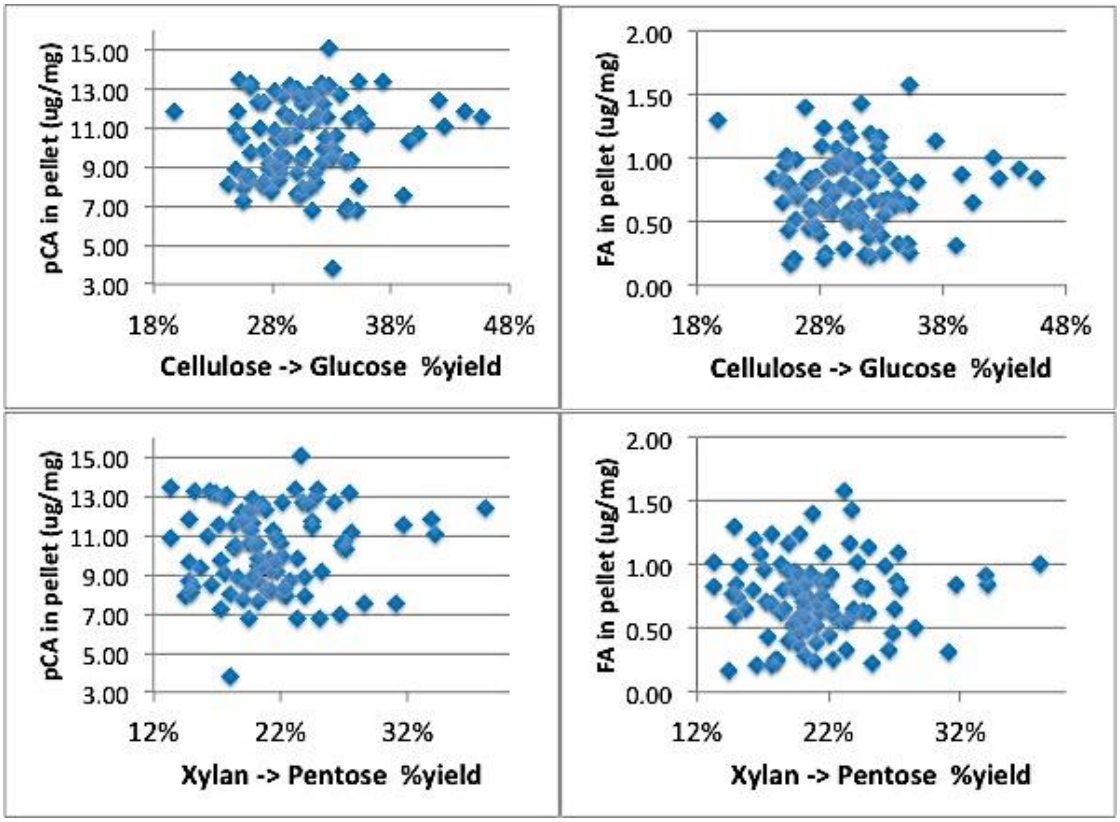

Figure 21. Lack of correlation between wall crosslinking compounds and monosaccharide yield.

Glucose Yield MLR	variance-inflation factors	*Pentose Yield MLR*	variance-inflation factors
Wall density (g/cm^3)	1.1	Wall density (g/cm^3)	1.1
%Lignin in the cell wall	1.4	%Lignin in the cell wall	1.6
Galactan+Arabinan / Xylan	1.4	Galactan+Arabinan / Xylan	1.3
Sample weight (g)	1.0	Sample weight (g)	1.0
%Cellulose in the sample	1.2	%Xylan in the sample	1.3
INT Diameter (mm)	1.9	INT Diameter (mm)	1.9
INT Length (cm)	1.9	INT Length (cm)	1.9

Table 8. *Variance-inflation factors for the independent variables in the multiple regression.*

Given the lack of clear trends in the bivariate data, I once again took a multiple regression approach for this analysis. The E-SAP panel represents a wide sampling of the genetic diversity across the species, and was chosen to model conversion efficiency in bioenergy-type sorghums. Fortunately the variance-inflation factors within this dataset were all quite low (Table 8), suggesting that the traits had minimal collinearity (Graham, 2003). The linear models for both glucose and pentose yield following digestion had low root-mean-squared error of prediction (Table 9), suggesting that the models had not over-fit the data (Hawkins *et al*., 2003). In addition, the residuals from each fit were approximately normally distributed (Figure 22) (Kratz & Resnick, 1995), showing that a linear model was appropriate for this data structure.

Model	RMSE of %20 test	RMSE of full model	SE of full model
Glucose yield	8.73001	8.72991	0.044
Pentose yield	2.58612	2.58686	0.018

Table 9. Root-mean-squared error of cross validation for the glucose and pentose yield multiple regressions. Cross validation was performed at 5-fold partitioning (80% training, 20% test) with ten repetitions. RMSE = root-mean-squared error of prediction, SE = Standard error of the dataset

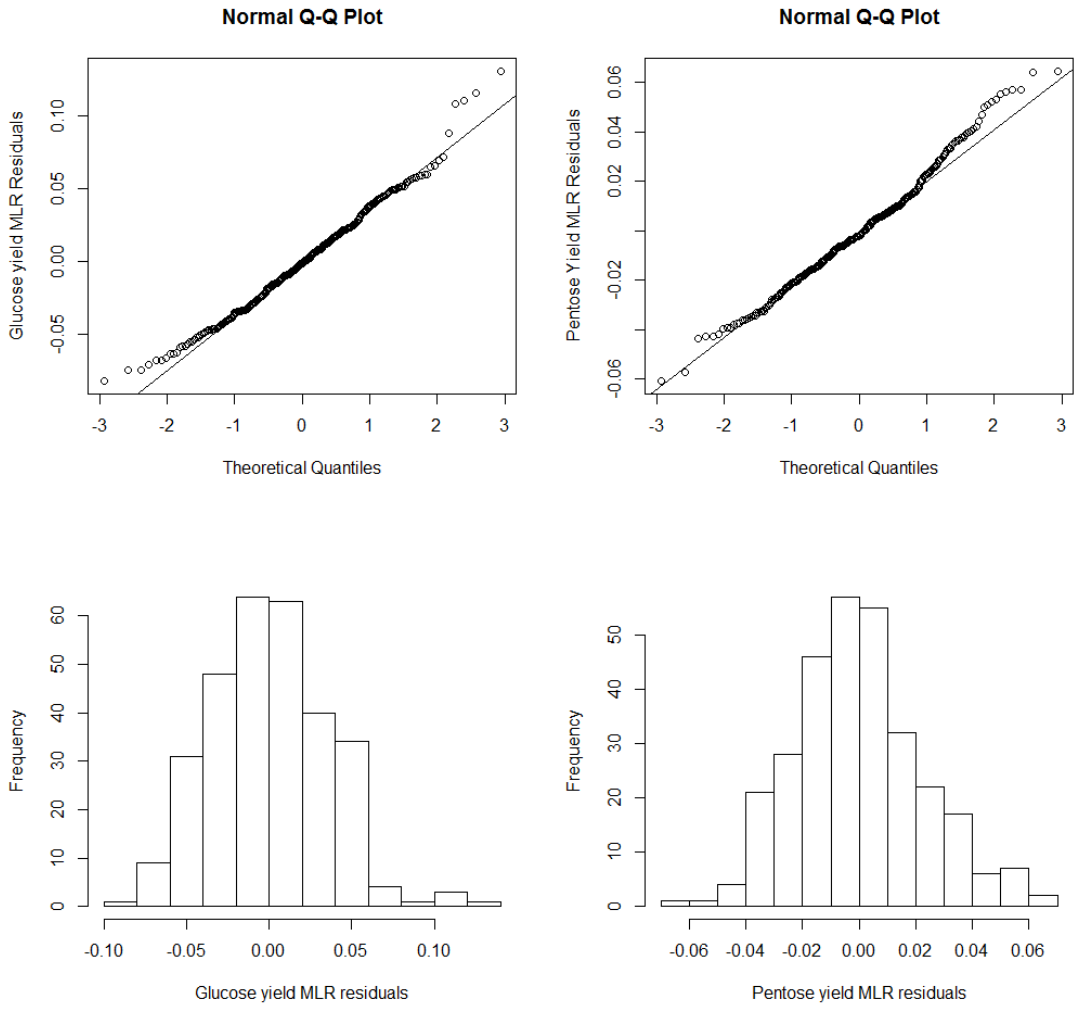

Figure 22. Quantile-Quantile plots and histograms of the residuals of the glucose yield and pentose yield multiple regressions.

Table 10 shows the regression models for glucose and pentose yield following digestion on the iWALL platform. Neither linear model explained over 50% of the variation in yield (R2 = 0.39 & 0.17, Table 10). The positive correlation between yield and internode size and the negative correlation between cell wall density and

Glucose Yield MLR	Coefficient	Std. Error	t-stat	p-value	
(Intercept)	2.66E-01	2.11E-03	126.01	2.00E-16	***
Wall density (g/cm^3)	-2.64E-02	4.44E-03	-5.96	7.13E-09	***
%Lignin in the cell wall	-3.40E-02	4.98E-03	-6.84	4.74E-11	***
Galactan+Arabinan / Xylan	2.16E-02	4.91E-03	4.40	1.53E-05	***
Sample weight (g)	-1.04E-02	4.31E-03	-2.42	1.60E-02	*
%Cellulose in the sample	-1.74E-02	4.61E-03	-3.77	1.99E-04	***
INT Diameter (mm)	1.78E-02	5.83E-03	3.05	2.50E-03	**
INT Length (cm)	1.50E-02	5.83E-03	2.57	1.07E-02	*

Multiple R-squared	0.40
Adjusted R-squared	0.39
F-statistic	27.94
p-value	2.20E-16

Pentose Yield MLR	Coefficient	Std. Error	t-stat	p-value	
(Intercept)	1.45E-01	1.32E-03	109.41	2.00E-16	***
Wall density (g/cm^3)	-8.67E-03	2.78E-03	-3.12	2.02E-03	**
%Lignin in the cell wall	-1.25E-02	3.33E-03	-3.75	2.11E-04	***
Galactan+Arabinan / Xylan	3.84E-03	3.07E-03	1.25	2.13E-01	
Sample weight (g)	-1.40E-02	2.70E-03	-5.17	4.44E-07	***
%Xylan in the sample	1.20E-02	3.07E-03	3.90	1.21E-04	***
INT Diameter (mm)	1.68E-03	3.67E-03	0.46	6.46E-01	
INT Length (cm)	1.09E-02	3.63E-03	2.99	3.02E-03	**

Multiple R-squared	0.19
Adjusted R-squared	0.17
F-statistic	9.90
p-value	4.34E-11

Table 10. *Multiple linear regression of internode morphology and compositional traits on the yield of glucose and pentoses following conversion on the iWALL platform.*

saccharification both suggest that biomass that is less densely packed may have better digestion on the iWALL system than biomass with dense cell walls. Though there are various interesting features of the multiple linear regression of pentose yield, this regression has a very low R2. This shows that the model does not take into account the

major driving forces in the dynamics of conversion of hemicelluloses to free pentose sugars in this dataset.

Interestingly, both the glucose and pentose yield multiple regressions reveal that lignin is indeed negatively influencing monosaccharide yield. In fact, lignin is actually the most influential factor within the glucose model following variable standardization. This relationship was not evident in the bivariate correlations, where various other factors influenced the analysis. This demonstrates the utility of the multiple regression approach for analyzing the saccharification efficiency of a material as complex as plant biomass.

7. CONCLUSIONS

Given the current lack of knowledge on the genetic control of stem growth in monocots, there is clearly much more work remaining to be done in this system. However, the data presented in this thesis should provide an important stepping-stone towards more in-depth analyses of sorghum developmental genetics in the future. The stem morphology and tissue-level data shown in Sections 3 & 4 provide a quantitative foundation for understanding stem growth, and could easily be incorporated into whole plant growth modeling designs. Likewise, the cell wall compositional information presented in Section 6 shows that structural features such as polymer packing and porosity may be more important determinants of saccharification efficiency than compositional features like percent lignin vs percent cellulose. The widely observed phenomenon of lignin inhibition on biochemical conversion of plant biomass to sugars is only weakly evident within the E-SAP association panel, and actually correlates positively with digestion efficiency in the BTx6428Tx7000 mapping population. If sorghum conversion is less dependent on lignin concentration than other plant species, this would further demonstrate sorghum's position as an ideal crop for biofuels production.

The gene expression data shown in Section 5 has many interesting trends to follow up on with targeted gene knockout or overexpression studies. In particular, there appears to be a correlation between the timing of cell division and the expression of

select bHLH, GRAS and B3 transcription factors. Since cell division appears to be the primary driving force in stem diameter growth, changing the expression patterns of these transcripts could have a significant impact on biomass or sucrose yield in sorghum. The majority of plant molecular biology research is focused on the model plant species *Arabidopsis thaliana*, and much of my hypothetical model of the gene network for stem growth will need to be revised in light of sorghum specific pathways and the neofunctionalization of genes following the divergence of the two plant species. However, this model should be an acceptable starting point to refine further in the light of new evidence. It is my sincere hope that this data will be carried forward in the future, and put to good use elsewhere.

REFERENCES

Alfons A (2012) cvTools: Cross-validation tools for regression models. R package version 0.3.2. http://CRAN.R-project.org/package=cvTools (20 March 2014)

Bartley LE, Peck ML, Kim SR, Ebert B, Manisseri C, Chiniquy DM, Sykes R, Gao L, Rautengarten C, Vega-Sánchez ME, Benke PI, Canlas PE, Cao P, Brewer S, Lin F, Smith WL, Zhang X, Keasling JD, Jentoff RE, Foster SB, Zhou J, Ziebell A, An G, Scheller HV, Ronald PC (2013) Overexpression of a BAHD acyltransferase, *OsAt10*, alters rice cell wall hydroxycinnamic acid content and saccharification. Plant Physiology **161:** 1615-1633

Bolle C (2004) The role of GRAS proteins in plant signal transduction and development. Planta **218:** 683-692

Boutté Y, Jonsson K, McFarlane HE, Johnson E, Gendre D, Swarup R, Friml J, Samuels L, Robert S, Bhalerao RP (2013) ECHIDNA-mediated post-Golgi trafficking of auxin carriers for differential cell elongation. PNAS **110:** 16259-16264

Burton RA, Gidley MJ, Fincher GB (2010) Heterogeneity in the chemistry, structure and function of plant cell walls. Nature Chemical Biology **6:** 724-732

Casler MD (2001) Breeding forage crops for increased nutritional value. Advances in Agronomy **71:** 51-107

Childs KL, Miller FR, Cordonnier-Pratt MM, Pratt LH, Morgan PW, Mullet JE (1997) The sorghum photoperiod sensitivity gene, Ma3, encodes a phytochrome B. Plant Physiology **113:** 611-619

de Lucas M, Davière JM, Rodríguez-Falcón M, Pontin M, Iglesias-Pedraz JM, Lorrain S, Fankhauser C, Blázquez MA, Titarenko E, Prat S (2008) A molecular framework for light and gibberellin control of cell elongation. Nature **451:** 480-484

Efroni I, Han SK, Kim HJ, Wu MF, Steiner E, Birnbaum KD, Hong JC, Eshed Y, Wagner D (2013) Regulation of leaf maturation by chromatin-mediated modulation of cytokinin responses. Developmental Cell **24:** 438-445

Eulgem T, Rushton PJ, Robatzek S, Somssich IE (2000) The WRKY superfamily of plant transcription factors. Trends in Plant Science **5:** 199-206

Fankhauser C (2002) Light perception in plants: cytokinins and red light join forces to keep phytochrome B active. Trends in Plant Science **7:** 143-145

Feng JX, Liu D, Pan Y, Gong W, Ma LG, Luo JC, Deng XW, Zhu YX (2005) An annotation update via cDNA sequence analysis and comprehensive profiling of developmental, hormonal or environmental responsiveness of the arabidopsis AP2/EREBP transcription factor gene family. Plant Molecular Biology **59:** 853-868

Gelman A (2008) Scaling regression inputs by dividing by two standard deviations. Statistics in Medicine **27:** 2865-2873

Graham MH (2003) Confronting multicollinearity in ecological multiple regression. Ecology **84:** 2809-2815

Harris PJ, Trethewey JAK (2010) The distribution of ester-linked ferulic acid in the cell walls of angiosperms. Phytochemistry Reviews **9:** 19-33

Hattori E, Uchida H, Harada N, Ohta M, Tsukada H, Hara Y, Suzuki T (2008) Incorporation and translocation of 2-deoxy-2-[(18)F]fluoro-D-glucose in *Sorghum bicolor* (L.) Moench monitored using a planar positron imaging system. Planta **227:** 1181-1186

Hawkins DM, Basak SC, Mills D (2003) Assessing model fit by cross-validation. Journal of Chemical Information and Computer Sciences **43:** 579-586

House LR (1985) A guide to sorghum breeding 2nd ed. Patancheru, India: International Crops Research Institute for the Semi-Arid Tropics

Ikeda M, Fujiwara S, Mitsuda N, Ohme-Takagi M (2012) A triantagonistic basic helix-loop-helix system regulates cell elongation in Arabidopsis. The Plant Cell **24:** 4483-4497

Jin H, Martin C (1999) Multifunctionality and diversity within the plant MYB-gene family. Plant Molecular Biology **41:** 577-585

Keuskamp DH, Keller MM, Ballaré CL, Pierik R (2012) Blue light regulated shade avoidance. Plant Signaling & Behavior **7:** 514-517

King GJ, Chanson AH, McCallum EJ, Ohme-Takagi M, Byriel K, Hill JM, Martin JL, Mylne JS (2013) The arabidopsis B3 domain protein VERNALIZATION1 (VRN1) is involved in processes essential for development, with structural and mutational studies revealing its DNA-binding surface. Journal of Biological Chemistry **288:** 3198-3207

Kratz MF, Resnick SI (1995) The QQ-Estimator and heavy tails. Stochastic Models **12:** 699-724

Levy YY, Mesnage S, Mylne JS, Gendall AR, Dean C (2002) Multiple roles of arabidopsis VRN1 in vernalization and flowering time control. Science **297:** 243-246

Lopez-Juez E, Nagatani A, Tomizawa KI, Deak M, Kern R, Kendrick RE, Furuya M (1992) The cucumber long hypocotyl mutant lacks a light-stable PHYB-like phytochrome. The Plant Cell **4:** 241-251

Menkir A, Goldsbrough P, Ejeta G (1997) RAPD based assessment of genetic diversity in cultivated races of sorghum. Crop Science **37:** 564-569

Menz MA, Klein RR, Unruh NC, Rooney WL, Klein PE, Mullet JE (2004) Genetic diversity of public inbreds of sorghum determined by mapped AFLP and SSR markers. Crop Science **44:** 1236-1244

Murphy RL, Klein RR, Morishige DT, Brady JA, Rooney WL, Miller FR, Dugas DV, Klein PE, Mullet JE (2011) Coincident light and clock regulation of psuedoresponse regulator protein 37 (PRR37) controls photoperiodic flowering in sorghum. PNAS **108:** 16469-16474

Nelson D & Werck-Reichhart D (2011) A P450-centric view of plant evolution. The Plant Journal **66:** 194-211

Nobusawa T, Okushima Y, Nagata N, Kojima M, Sakakibara H, Umeda M (2013) Synthesis of very-long-chain fatty acids in the epidermis controls plant organ growth by restricting cell proliferation. PLoS Biology **11:** 1-14

Ohashi-Ito K, Bergmann DC (2006) Arabidopsis FAMA controls the final proliferation/differentiation switch during stomatal development. The Plant Cell **18:** 2493-2505

Powell AE, Lenhard M (2012) Control of organ size in plants. Current Biology **22:** 360-367

Price HJ, Dillon SL, Hodnett G, Rooney WL, Ross L, Johnston JS (2005) Genome evolution in the genus *Sorghum* (Poaceae). Annals of Botany **95:** 219-227

Ralph J, Hatfield RD, Quideau S, Helm RF, Grabber JH, Jung HJG (1994) Pathway of p-coumaric acid incorporation into maize lignin as revealed by NMR. JACS **116:** 9448-9456

Rawat R, Schwartz J, Jones MA, Sairanen I, Cheng Y, Andersson CR, Zhao Y, Ljung K, Harmer SL (2009) REVEILLE1, a myb-like transcription factor, integrates the circadian clock and auxin pathways. PNAS **106:** 16883-16888

Sablowski R (2009) Cytokinin and WUSCHEL tie the knot around plant stem cell niches. PNAS **106:** 16016-16017

Santoro N, Cantu SL, Tornqvist CE, Falbel TG, Bolivar JL, Patterson SE, Pauly M, Walton JD (2010) A high-throughput platform for screening milligram quantities of plant biomass for lignocellulose digestibility. BioEnergy Research **3:** 93-102

Schneider CA, Rasband WS, Eliceiri KW (2012) NIH Image to ImageJ: 25 years of image analysis. Nature Methods **9:** 671-675

Shitsukawa N, Ikari C, Shimada S, Kitagawa S, Sakamoto K, Saito H, Ryuto H, Fukunishi N, Abe T, Takumi S, Nasuda S, Murai K (2007) The einkorn wheat (*Triticum monococcum*) mutant, *maintained vegetative phase*, is caused by a deletion in the *VRN1* gene. Genes & Genetic Systems **82:** 167-170

Simonneau T, Habib R, Goutouly JP, Huguet JG (1993) Diurnal changes in stem diameter depend upon variations in water content: direct evidence in peach trees. Journal of Experimental Botany **44:** 615-621

Smith CW, Frederiksen RA (2000) Sorghum: origin, history, technology, and production vol. 2. John Wiley & Sons New York, NY. p. 840

Tarpley L, Vietor DM (2007) Compartmentation of sucrose during radial transfer in mature sorghum culm. BMC Plant Biology **7:** 1-10

Thimm O, Bläsing O, Gibon Y, Nagel A, Meyer S, Krüger P, Selbig J, Müller LA, Rhee SY, StittM (2004) MapMan: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. Plant Journal **37:** 914-939

Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nature Protocols **7:** 562-578

Usadel B, Nagel A, Steinhauser D, Gibon Y, Bläsing OE, Redestig H, Sreenivasulu N, Krall L, Hannah MA, Poree F, Fernie AR, Stitt M (2006) PageMan: An interactive ontology tool to generate, display, and annotate overview graphs for profiling experiments. BMC Bioinformatics **7:** 1-8

USDA, ARS, National Genetic Resources Program Germplasm Resources Information Network - (GRIN) [Online Database]. National Germplasm Resources Laboratory, Beltsville, Maryland. http://www.ars-grin.gov/cgi-bin/npgs/html/taxon.pl?35092 (21 April 2013)

Wang F, Liu CZ (2009) Development of an economic refining strategy of sweet sorghum in the inner mongolia region of china. Energy & Fuels **23:** 4137-4142

Wilson JR, Mertens DR, Hatfield RD (1993) Isolates of cell types from sorghum stems: digestion, cell wall and anatomical characteristics. Journal of the Science of Food and Agriculture **63:** 407-417

Wolfrum E, Payna C, Stefaniak T, Rooney W, Dighe N, Bean B, Dahlberg J (2013) Multivariate calibration models for sorghum composition using Near-Infrared Spectroscopy. Contract **303:** 1-8