

EXPRESSION ANALYSIS AND IDENTIFYING CANDIDATE GENES FOR WOOD
FORMATION IN DESERT SUNFLOWER, *HELIANTHUS ARGOPHYLLUS*

by

YUAN-SHENG YU

(Under the Direction of Jeffrey F. D. Dean)

ABSTRACT

Biofuel production is an increasingly important area of research as fossil fuels are becoming scarcer and climate change caused by the burning of fossil fuels threatens the environmental status quo. The common sunflower (*Helianthus annuus*) is a source of seed oil that may be used in biodiesel fuels. A relative, the silver-leaf sunflower (*Helianthus argophyllus*), is a wood-producing desert species that has shown potential for biomass production. Gene expression profiling with microarray and the use of laser microscopy dissection to specifically isolate xylem and phloem have shown have pointed to particular proteins playing important regulatory roles in wood formation as well as cellulose, hemicellulose, and lignin biosynthesis.

Index Words: *Helianthus argophyllus*, Microarray, Laser microscopy dissection (LMD), RT-qPCR, Wood formation

EXPRESSION ANALYSIS AND IDENTIFYING CANDIDATE GENES FOR WOOD
FORMATION IN DESERT SUNFLOWER, *HELIANTHUS ARGOPHYLLUS*

by

YUAN-SHENG YU

B.S. The University of California: Berkeley, 2008

A Thesis Submitted to the Graduate Faculty of The University of Georgia in Partial
Fulfillment of the Requirements for the Degree

Master of Science

Athens, Georgia

2012

© 2012

Yuan-Sheng Yu

All Rights Reserved

EXPRESSION ANALYSIS AND IDENTIFYING CANDIDATE GENES FOR WOOD
FORMATION IN DESERT SUNFLOWER, *HELIANTHUS ARGOPHYLLUS*

by

YUAN-SHENG YU

Major Professor: Jeffrey F. D. Dean

Committee: John M. Burke
C.J. Tsai

Electronic Version Approved:

Maureen Grasso
Dean of Graduate School
The University of Georgia
August 2012

ACKNOWLEDGMENTS

I would like to thank the Dr. Jeffrey F. D. Dean for being a great advisor and mentor in my studies these past years. With his knowledgeable help and approach I was able to take full advantage of the opportunity to educate myself in the academic environment. Also, Dr. Walt W. Lorenz for giving me the technical background I needed to overcome the challenges at the lab bench and always have an answer to my questions. To the members of the lab, Michael Bordeaux and Ujwal Bagal, thank you for providing a fun and productive working environment. Michael, you were always a helping hand whenever I needed advice or just an extra set of hands to get things done. Also, I would like to thank Dr. CJ Tsai and Dr. John M. Burke for introducing me to new concepts and ideas in research. Finally, I would like to thank United States Department of Energy for the financial support in this project.

TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS	iv
LIST OF TABLES	vii
LIST OF FIGURES	viii
CHAPTER	
1 INTRODUCTION OF <i>HELIANTHUS ARGOPHLLYUS</i> AS A SOURCE OF BIOMASS FOR BIOFUEL APPLICATIONS.....	1
Biofuels	1
Sunflower as a Candidate for Biomass	3
Wood Formation	4
Gene Expression Profiling	6
Laser Microscopy Dissection.....	8
Research Objectives	9
References.....	11

2	EXPRESSION ANALYSIS OF WOOD-FORMING <i>HELIANTHUS ARGOPHYLLUS</i>	19
	Introduction.....	19
	Materials and Methods.....	24
	Results.....	28
	Discussion	31
	Conclusion	33
	References	35
3	QUANTIFICATION OF DIFFERENTIALLY EXPRESSED GENES IN XYLEM AND PHLOEM TISSUES OF <i>HELIANTHUS ARGOPHYLLUS</i>	46
	Introduction.....	46
	Materials and Methods.....	49
	Results.....	52
	Discussion	55
	Conclusion	58
	References	59
4	SUMMARY OF WORK	77
	Conclusion From Present Work.....	77
	Further Studies	78

LIST OF TABLES

	Page
Table 1. Chemical profiles of <i>H. argophyllus</i> (Arg1820) and <i>P. taeda</i> wood samples.....	18
Table 2. NanoDrop data of LCM samples.	75
Table 3. 17 genes of interest and their respective fold change values selected from microarray data.....	76

LIST OF FIGURES

	Page
Figure 1a. Paraffin embedded and stained with toluidine blue transverse section of <i>Helianthus argophyllus</i> at 8 weeks.	16
Figure 1b. Paraffin embedded and stained with toluidine blue transverse section of <i>Helianthus annuus</i> at 8 weeks..	17
Figure 2. Cellulose microfibrils are synthesized by large protein complexes (rosettes) embedded in the plasma membrane.	39
Figure 3. Schematic representation of the generalized components of plant secondary cell walls. ...	40
Figure 4. Volcano plot of 17715 differentially expressed entities.....	41
Figure 5. Scatter plot comparison of gene expression of Internode 1 between <i>H. argophyllus</i> (Arg1820) and <i>H. annuus</i> (HA4120-HO).	42
Figure 6. RT-qPCR data validating 17 selected candidate genes observed from the microarray data and 4 housekeeping genes.	43
Figure 7a. Paraffin embedded and stained with toluidine blue transverse section of <i>Helianthus argophyllus</i> at 8 weeks.	44
Figure 7b. Paraffin embedded and stained with toluidine blue transverse section of <i>Helianthus annuus</i> at 8 weeks.	45
Figure 8. Phenotype comparison of <i>H.annuus</i> and <i>H. argophyllus</i>	63
Figure 9a. 15um cross-section of the 1st node at 8-week of Arg1820 stained with toluidine blue observed at 2.5x magnification.	64

Figure 9b. 15um cross-section of the 1st node at 8-week of HA412-HO stained with toluidine blue observed at 2.5x magnification.	65
Figure 10a. 15um cross-section of the 3 rd node at 8-week of Arg1820 stained with toluidine blue observed at 2.5x magnification.	66
Figure 10b. 15um cross-section of the 3 rd node at 8-week of HA412-HO stained with toluidine blue observed at 2.5x magnification.	67
Figure 11a. 15um 15um cross-section of the base at 8-week of Arg1820 stained with toluidine blue observed at 2.5x magnification.	68
Figure 11b. 15um cross-section of the base at 8-week of HA412-HO stained with toluidine blue observed at 2.5x magnification.	69
Figure 12a. RT-qPCR expression level comparison between HA412-Xylem and Arg1820-Xylem from the base at the 8-week time point.	70
Figure 12b. RT-qPCR expression level comparison between HA412-HO-Phloem vs Arg1820-Phloem from the base at the 8-week time point.....	71
Figure 13a. RT-qPCR expression level comparison between HA412-HO-Xylem vs HA412-HO-Phloem from the base at the 8-week time point.....	72
Figure 13b. RT-qPCR expression level comparison between Arg1820-Xylem vs Arg1820-Phloem from the base at the 8-week time point.....	73
Figure 14. RT-qPCR expression level comparison between all tissue samples (HA412-Xylem, HA412-Phloem, Arg1820-Xylem, Arg1820-Phloem) of four housekeeping genes (ACT1, ACT2, PAL1, TUB2).....	74

CHAPTER 1: INTRODUCTION OF *HELIANTHUS ARGOPHLLYUS* AS A SOURCE OF BIOMASS FOR BIOFUEL APPLICATIONS

Biofuels

Biofuel production is a rapidly growing area of research as fossil fuels are becoming scarcer and climate change caused by the burning of fossil fuels threatens the environment (Dellomonaco et al. 2010). Currently biofuels are being produced from such varied feedstocks as sawdust, grass cuttings, agricultural waste, and dried manure (Briens et al. 2008). The main feedstocks for biofuel (ethanol) production have been sugars and starch from corn, wheat, barley, and sugarcane. However, the use of corn and other common food commodities has effects on food costs, and biofuel production has already caused concerns in the United States, where food prices have increased with biofuel production (Odling-Smee, 2007). Another issue for production of biofuel using sugar and starch crops is that biofuel production can actually result in increased emissions of carbon dioxide into the environment due to the use of energy-intensive fertilizers and pesticides (Hill et al. 2006). In addition to the detrimental environmental effects of first-generation biofuels, there are economic effects, particularly in poorer countries that struggle to provide enough food for their people (Gomez et al. 2008). While these “first-generation” biofuels appear to be unsustainable because they place stress on the food industry, “second-generation” biofuels are a new direction in which cheap and abundant non-food plant biomass is used in the production of biofuels (Gomez et al. 2008).

The need for a more environmentally and economically efficient source of biofuel is urgent, and the use of plant biomass may be the answer. Ideally, second-generation

biofuels are produced from dedicated biomass crops that yield significant amounts of biomass in a short period of time on with the least amount of land and with little use of fertilizers and pesticides (Gomez et al. 2008). Lignocellulosic biomass is an abundant and underutilized biological resource and is considered a primary source of material for second-generation biofuels production (US DOE, 2006). With excessive carbon dioxide emission being a drawback of first-generation biofuels, second-generation biofuels, including perennial grasses such as switchgrass, as well as woody species, such as poplar and willow, are expected to be carbon neutral or better yet, carbon negative (Gomez et al. 2008). Life cycle assessments (LCA) used to assess the sustainability of biofuels show that second-generation biofuels may offer a lower carbon footprint than their predecessors (Sheehan 2009).

Biomass can be converted to energy in a variety of processes. The most ancient form of biomass energy comes from directly burning wood for heat. The scale of this combustion ranges from space heating inside homes to generating heat and electricity in industrial plants (Briens et al. 2008). The conversion of biomass to liquid fuels is also possible through two specific pathways; thermochemical processing and biochemical processing. Thermochemical processing is the heating of biomass at various oxygen levels. Pyrolysis, heating in the absence of oxygen, results in various organic liquids that eventually can be converted to liquid fuel. At low levels of oxygen levels, gasification occurs and hydrogen and organic gases are produced, which can then be converted to desired liquid fuel (Ezeji et al. 2007). Biochemical processing is the conversion of biomass, through saccharification, to polysaccharides and fermenting them to produce alcohols, such as butanol and ethanol (Wingren et al. 2003). Both these processes have

their advantages and disadvantages. While thermochemical processing utilizes all components of biomass, biochemical processing is specific to the polysaccharide content. However the former requires a much larger feedstock, transportation costs, and high temperature processing, when the latter can operate with smaller local feedstock, but the cost-effective issue comes in the critical step of conversion of biomass to polysaccharides (Gomez et al. 2008).

Sunflower as a Candidate for Biomass Production

Perennial grasses, poplar, and willow are sources of biomass commonly considered for use in second-generation alternative fuel development. Another candidate for consideration is the common sunflower (*Helianthus annuus*). However, common sunflower has so far only been grown as a source of oilseed, which can be used for biodiesel production, but the remainder of the plant has not been considered as a source of lignocellulosic biomass for biofuel production (Beckman et al. 2008). This is primarily due to the low-density, pithy stems that characterize common sunflowers. These low-density stems contain relatively little carbon that could be converted to biofuel. However, in desert environments, there are species of sunflowers that do produce high-density woody stems, including the silverleaf sunflower (*Helianthus argophyllus*) and the Algodones dune sunflower (*Helianthus niveus*). Focusing on the woody *Helianthus argophyllus* in contrast to the pithy *Helianthus annuus* allows insight into the genetics that make them drought tolerant and lead them to produce wood.

Figure 1 displays the developmental differences of *Helianthus argophyllus* (Figure 1a) and *Helianthus annuus* (Figure 1b) at the 8-week time point. Differences in the

development of xylem between the two species are responsible for the woody characteristics of the former. At the 8-week time point the xylem of *H. argophyllus* has formed a complete ring around the stem while radial growth has diminished the presences of pith cells. *H. annuus* however does not have a complete xylem ring, instead the xylem cells are still contained in isolated bundles. In result there is a much greater presence of pith cells in the stem. The difference in proportion of xylem and pith cells between the two species is critical in the woody characteristics of *H. argophyllus* that *H. annuus* lacks. An understanding of the biosynthetic pathways responsible for this development makes it helpful in the production of hybrid sunflowers with high-density biomass that can be used for the production of biofuels.

Wood Formation

Plant cell walls constrain cell expansion and provide mechanical strength to the plant structure. In general, there are two types of cell walls in plants, primary and secondary, and they are distinguished by their composition as well as their roles in plant development. All plant cells initially synthesize primary cell walls, which help determine the final cell size and shape. In a subset of cells, secondary cell wall formation occurs after growth cessation, and their primary functions are to provide mechanical support and limit loss of critical metabolites, such as water (Dai et al. 2011, Sato et al. 2010). Lignocellulosic biomass generally contains a preponderance of secondary cell walls.

Wood is particularly rich in secondary walls from xylem fiber cells, and these cell walls are composed mainly of three biopolymers -- cellulose, hemicellulose and lignin. Generally, cellulose and hemicelluloses represent 42-50% and 25-30% of bulk wood

composition, respectively. Lignin constitutes 20-25% of wood, while the remaining 5-8% of wood dry weight is comprised of extractives, which are typically low molecular weight, non-polymeric wall compounds that can be removed using solvents (Suzuki et al. 2006). The abundance and organization of the three major wood cell wall polymers, as well as the regulation of their biosynthetic pathways, greatly affect the downstream utilization of plant biomass for biofuels production (Johansen et al. 2006).

Preliminary studies of stems from the wood-producing silverleaf sunflower using pyrolysis molecular beam mass spectrometry (pyrMBMS) and high performance liquid chromatography (HPLC) showed that the stem chemistry is similar to that for quaking aspen (*Populus tremuloides*), a relative of black cottonwood (*Populus trichocarpa*) that is widely used as a model species for studies of wood biology (Knapp 2008, Sheppard et al. 2000, Tuskan et al. 2006). Chemical profile comparison (Table 1) between *H. argophyllus* and loblolly pine (*Pinus taeda*), another commonly studied species in wood biology, shows the similarity in stem chemistry supporting *H. argophyllus* as a candidate for a source of biomass. The cellulose, hemicelluloses, and lignin content in *H. argophyllus* are in the general proportion characteristic of wood; cellulose (38.4%), hemicelluloses (18.6%), and lignin (20.5%).

Cellulose content is a particularly important factor with respect to biofuel production since it serves as the main source of polysaccharides for fermentative processes. Studies of cellulose biosynthesis and the various genes involved (e.g. cellulose synthases) will be crucial for our ability to manipulate the distinct processes of primary and secondary cell wall formation in ways that make biofuel production more efficient (Johansen et al. 2006). The regulation of primary and secondary cell wall biosynthesis

will be particularly important for understanding formation of the high-density, secondary wall-rich woody stems of silverleaf and Algodones dune sunflowers versus the low-density, primary wall-rich pithy stems of the common sunflower.

Lignin, the dominant defining component of lignocellulosic secondary cell walls, is the key to development of plant vascular systems and the great stature of many woody plants (Rogers et al. 2005). Lignin serves as a thermoplastic matrix that crosslinks components of the secondary cell wall to strengthen the mechanical characteristics of plant stems. In addition to structural support, the hydrophobic properties of lignin, in contrast to the hydrophilic properties of the secondary cell wall polysaccharides (cellulose and hemicellulose), allow it to play a critical role in conducting water throughout the plant (Chabannes et al. 2001). This aspect of lignin has obvious ramifications for the increased production of lignified xylem in the silverleaf and Algodones dune sunflowers since plant water-use efficiency (WUE) affects plant fitness and acts as a natural selector in the desert environments where these plants grow (Donovan et al. 2006). The desert floor habitats occupied by the silverleaf and Algodones dune sunflowers are extreme environments that impose much stronger water stress on these plants than is experienced by the domesticated species, common sunflower. Adaptations to the harsh desert environments have likely driven the increased production of lignocellulosic vascular tissues in the desert species (Gross et al, 2004).

Gene Expression Profiling

Gene expression profiling is one technique for identifying and characterizing regulatory roles of genes in complicated developmental pathways. In particular, large-

scale genomic approaches allow us to follow the expression of many genes in parallel simultaneously, which facilitates identification of gene networks having common patterns of expression (Lai et al. 2005). Microarrays are a routine tool used in the analysis of global gene expression. Its ability to determine the expression level of a wide array of genes makes it a critical technology in investigating the genetic expression and identifying candidate genes particular to a specific phenotype or being affected by various abiotic factors (Kilian et al. 2011). With the advancement of bioinformatics and the costs of deep sequencing exponentially declining, RNA-seq, has surfaced as a prominent form of technology in gene expression profiling. At a very high sensitivity, the transcriptome of an entire sample is sequenced and easily aligned with numerous databases for identification of the abundance of a particular gene in addition to mutations, deletions, insertions, and splicing (Pinto et al. 2011).

The development of expansive expressed sequence tag (EST) databases and improved bioinformatics approaches have facilitated putative functional identification of genes in networks discovered by gene expression profiling (Fernandez et al. 2008). With technological advances in bioinformatics, custom databases can be constructed with next-generation sequencing data. The MAGIC interface allowed for the design of a sunflower database providing a key tool in predicting gene function and analysis of candidate genes for phenotypes of interest (Bachlava et al. 2012, Liang et al. 2006).

Gene expression analysis should aid us in understanding the gene expression differences leading to wood formation in desert sunflower ecotypes versus common sunflower. Since cellulose, hemicellulose, and lignin are the primary components of angiosperm wood, differential expression in genes involved in their production will help

identify candidate genes that may play critical roles in wood formation in the desert sunflowers (Somerville et al. 2006, Persson et al. 2005). Previous work has shown that secondary cell wall formation is regulated and activated by various genes and transcription factors (Zhong et al. 2010). Additionally, a large number of genes involved in cellulose, hemicellulose, and lignin biosynthesis have been identified, and the coordinated gene expression of these genes has been demonstrated in other wood-forming species (Lerouxel et al. 2006, Somerville et al. 2006, Zhong et al. 2006). Gene expression analysis should identify sunflower homologs to the genes shown to play key regulatory and biosynthetic roles in secondary cell wall formation in other wood forming species (McCarthy et al. 2010).

Laser-capture Microdissection

Laser-capture microdissection (LCM) combines the traditional microscopy and dissection techniques to produce a cutting-edge technology that allows for the isolation of specific cell types and in some cases even individual cells (Nelson et al. 2006). With the advent of high-sensitivity genomic approaches, LCM has become a prominent technology for profiling gene expression in small subsets of cells (Brandt et al. 2002). The technique allows for isolation of sufficient quantities of RNA for downstream procedures, and LCM now plays a major role in a variety of genomic and metabolomic studies (Westphal et al. 2002).

As is the case in other higher eukaryotes, plant tissue structure derives from a myriad of various cell types, each characterized by unique gene expression patterns that ultimately contribute to the overall growth and development patterns of the organism.

The non-uniformity of gene expression between various cell types creates a layer of ambiguity for gene expression profiling that can potentially be deciphered through the application of laser-capture microdissection. There are various forms of laser-capture microdissection, each with its own advantages and disadvantages for subsequent analyses (Brandt et al. 2005, Day et al. 2005). LCM has been shown with the combination of downstream technologies, to be critical in the study of expression profiling in specific cell types (Lange et al. 2005). With the need to isolate small cellular bundles in stem cross section and the ability to isolate intact RNA from cellular bundles while not compromising the physical integrity of the stem section, formalin fixed paraffin embedded (FFPE) tissue samples is the best LCM technique for the study of wood formation in sunflower.

Research Objectives

This thesis is focused on gene expression profiling of stem development in *H. argophyllus* and *H. annuus*. Two major research objectives are as follows: 1. Identifying candidate genes that play a regulatory role in wood formation and cellulose, hemicelluloses, and lignin biosynthesis in *H. argophyllus* in addition to genetic markers differentiating *H. argophyllus* from *H. annuus*; and 2. Identifying differential gene expression levels of aforementioned candidate genes in various cell types, specifically xylem and phloem cells within each species. Chapter 2 deals with identifying candidate genes with the use of 4X44K Agilent One-Color microarrays to conducted an inter-species genes expression profiling comparison. Chapter 3 deals with differential gene expression of candidate genes in xylem and phloem, isolated by LCM samples from

FFPE tissues. The overall goal of the thesis is to better understand the regulatory roles of genes in the biosynthetic pathway of wood formation in *H. argophyllus* in contrast to its non-woody counterpart *H. annus*.

References

1. Bachlava E, Taylor CA, Tang S, Bowers JE, Mandel JR, Burke JM, Knapp SJ. 2012. SNP Discovery and Development of High-Density Genotyping Array for Sunflower. *PLoS One* 7(1): e29814
2. Beckman, W. 2008. Ancient Sunflower Fuels Debate About Agriculture in the Americas. *Science Daily*.
3. Brandt SP. 2005. Microgenomics: gene expression analysis at the tissue-specific and single-cell levels. *J. Exp. Bot.* 56: 495-505
4. Brandt S, Koska S, Altmann T, Kehr J. 2002. Using array hybridization to monitor gene expression at the single cell level. *J. Exp.* 53(23): 15-23
5. Briens, C, Piskorz, J, Berruti F. 2008. Biomass Valorization for Fuel and Chemicals Production – A Review. *International Journal of Chemical Reactor Engineering* 6: R2
6. Chabannes M, Ruel K, Yoshinaga A, Chabbert B, Jauneau A, Joseleau JP, Boudet AM. 2001. In situ analysis of lignins in transgenic tobacco reveals a differential impact of individual transformations on the spatial patterns of lignin deposition at the cellular and subcellular levels. *Plant Journal* 28(3): 271-282
7. Dai X, You C, Chen G, Li X, Zhang Q, Wu C. 2011. OsBC1L4 encodes a COBRA-like protein that affects cellulose synthesis in rice. *Plant Molecular Biology* 75(4-5): 333-345
8. Day RC, Grossniklaus U, Macknight RC. 2005. Be more specific! Laser-assisted microdissection of plant cells. *Trends Plant Science* 10: 397-406

9. Dellomonaco, C. 2010. The path to next generation biofuels: successes and challenges in the era of synthetic biology. *Microbial Cell Factories* 9:3
10. Donovan LA, Dudley, SA, Rosenthal, DM, Ludwig, F. 2007. Phenotypic selection on leaf water use efficiency and related ecophysiological traits for natural populations of desert sunflowers. *Oecologia* 152: 13-25
11. Ezeji TC, Qureshi N, Blaschek HP. 2007. Bioproduction of butanol from biomass: from genes to bioreactors. *Current Opinion in Biotechnology* 18: 220-227
12. Fernandez, P, Doslo GAA. 2008. Validating housekeeping genes for quantitative plant gene expression studies in sunflower senescence process. 54^o Congresso Brasileiro de Genetica.
13. Gomez, LD, Steele-King, CG, McQueen-Mason, SJ. 2008. Sustainable liquid biofuels from biomass: the writing's on the walls. *New Phytologist* 178: 473-485
14. Gross, BL, Kane, NC, Lexer, C, Ludwig, F, Rosenthal, DM, Donovan, LA, Rieseberg, LH. 2004. *American Naturalist* 164(2): 145-256
15. Hill J, Nelson E, Tilman D, Polasky S, Tiffay D. 2006. Environmental economic, and energetic costs and benefits of biodiesel and ethanol biofuels. *Proceedings of the National Academy of Sciences, USA* 103: 11206-11210
16. Johansen JN, Vernhettes S, Hofte H. 2006. The ins and outs of plant cell walls. *Current Opinion in Plant Biology* 9: 616-620
17. Killan J, Peschke F, Berendzen KW, Harter K, Wanke D. 2011. Prerequisites, performance and profits of transcriptional profiling the abiotic stress response. *BBA – Gene Regulatory Mechanisms* 1819: 1661-75

18. Lai Z, Livingstone K, Zou Y, Church SA, Knapp SJ, Andrews J, Riesberg LH. 2005. Identification and mapping SNPs from ESTs in sunflower. *Ther Appl Genet* 111: 1532-1544
19. Lange BM. 2005. Single-cell genomics. *Current Opinion Plant Biology* 8: 236-241
20. Lerouxel O, Cavalier DM, Liepman AH, Keegstra K. 2006. Biosynthesis of plant cell wall polysaccharides - a complex process. *Current Opinion Plant Biology* 9(6): 621-630
21. Liang C, Sun F, Wang HM, Qu JF, Freeman RM. 2006. MAGIC-SPP: a database-driven DNA sequence processing package with associated management tools. *BMC Bioinformatics* 7: 115.
22. McCarthy RL, Zhong R, Fowler S, Lyskowski D, Piyasena H, Carleton K, Spicer C, Yu ZH. 2010. The poplar MYB transcription factors, PtrMYB3 and PtrMYB20, are involved in the regulation of secondary wall biosynthesis. *Plant Cell Physiology*. Epub ahead of print.
23. Nelson T, Tausta SL, Gandotra N, Liu T. 2006. Laster Microdissection of Plant Tissue: What You See Is What You Get. *Annual Review Plant Biology* 57: 181-201
24. Odling-Smee L. 2007. Biofuels bandwagon hits a rut. *Nature* 446: 483
25. Persson, S, Wei, H, Milne, J, Page, G.P. 2005 Large-scale coexpression analysis reveals novel genes involved in cellulose biosynthesis. *Proc. National Academy of Science USA*. 102: 8633-8638

26. Pinto AC, Melo-Barbosa HP, Miyoshi A, Silva A, Azevedo V. 2011. Application of RNA-seq to reveal the transcript profile in bacteria. *Genetics and Molecular Research* 10(3): 1707-1718
27. Rogers LA, Dubos C, Surman C, Willment J, Cullis IF, Mansfield SD, Campbell MM. 2005. Comparison of lignin deposition in three ectopic lignification mutants. *New Phytologist* 168: 123-140
28. Sato K, Ito S, Fujii T, Suzuki R, Takenouchi S, Nakaba S, Funada R, Sano Y, Kajita S, Kitano H, Katayama Y. 2010. The carbohydrate-binding module (CBM)-like sequence is crucial for rice CWA1/BC1 function in proper assembly of secondary cell wall materials. *Plant Signal Behavior* 5(11): 1433-1436
29. Sheehan, JL. 2009. Biofuels and the conundrum of sustainability. *Current Opinion in Biotechnology* 20: 318-324
30. Sheppard, L. 2000. A DEFICIENS Homolog from the Dioecious Tree Black Cottonwood Is Expressed in Female and Male Floral Meristems of the Two-Whorled Unisexual Flowers. *Plant Physiology* 124: 627-640
31. Somerville C.R. 2006. Cellulose synthesis in higher plants. *Cell Developmental Biology*. 22: 53-78
32. Suzuki S, Li L, Sun YH, Chiang VL. 2006. The cellulose synthase gene superfamily and biochemical functions of xylem-specific cellulose synthase-like genes in *Populus trichocarpa*. *Plant Physiology* 142(3): 1233-1245
33. Tuskan, GA, Difazio, S, Jansson, S, Bohlmann, J. 2006. The genome of black cottonwood, *Populus trichocarpa*. *Science* 15(313): 1596-1604

34. United States Department of Energy (US DOE). 2006. Breaking the biological barriers to cellulosic ethanol: a join research agenda.
35. Westphal G, Burgemeister R, Friedemann G, Wellmann A, Wernert N. 2002. Noncontact laser catapulting: a basic procedure for functional genomics and proteomics. *Methods Enzymology* 356: 80-99
36. Wingren A, Galbe M, Zacchi G. 2003. Techno-economic evaluation of producing ethanol from softwood: comparison of SSF and SHF and identification of bottlenecks. *Biotechnology Progress* 19: 1109-1117
37. Zhong R, Demura T, Ye ZH. 2006. SND1, a NAC domain transcription factor, is a key regulator of secondary wall synthesis in fibers of Arabidopsis. *Plant Cell* 18(11): 3158-3170
38. Zhong R, Yu ZH. 2010. The poplar PtrWNDs are transcriptional activators of secondary cell wall biosynthesis. *Plant Signal Behavior* 5(4)

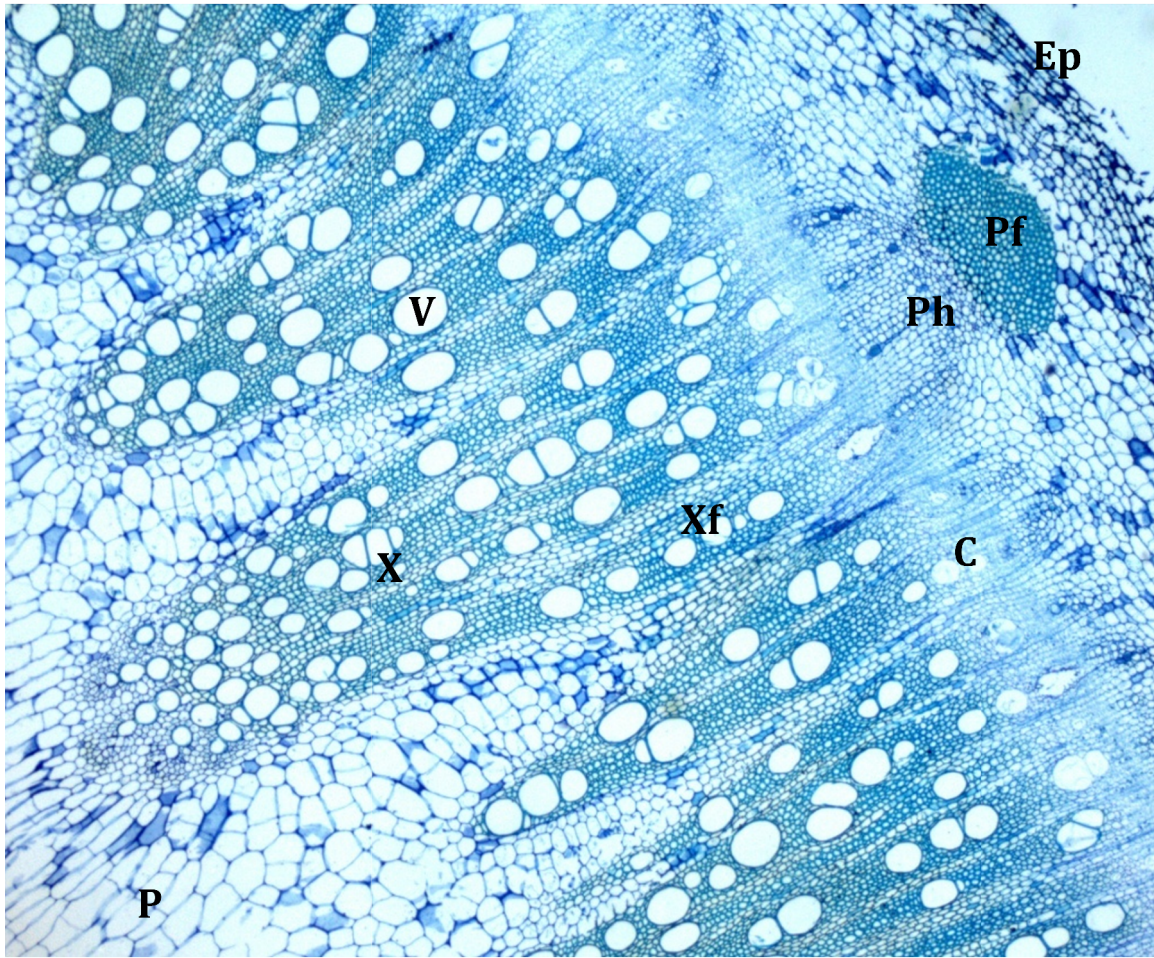


Figure 1a. Paraffin embedded and stained with toluidine blue transverse section of *Helianthus argophyllus* at 8 weeks. Pith (P), xylem (X), xylem fibers (Xf), vessels (V), vascular cambium (C), phloem (Ph), phloem fiber (Pf), and epidermis (Ep).

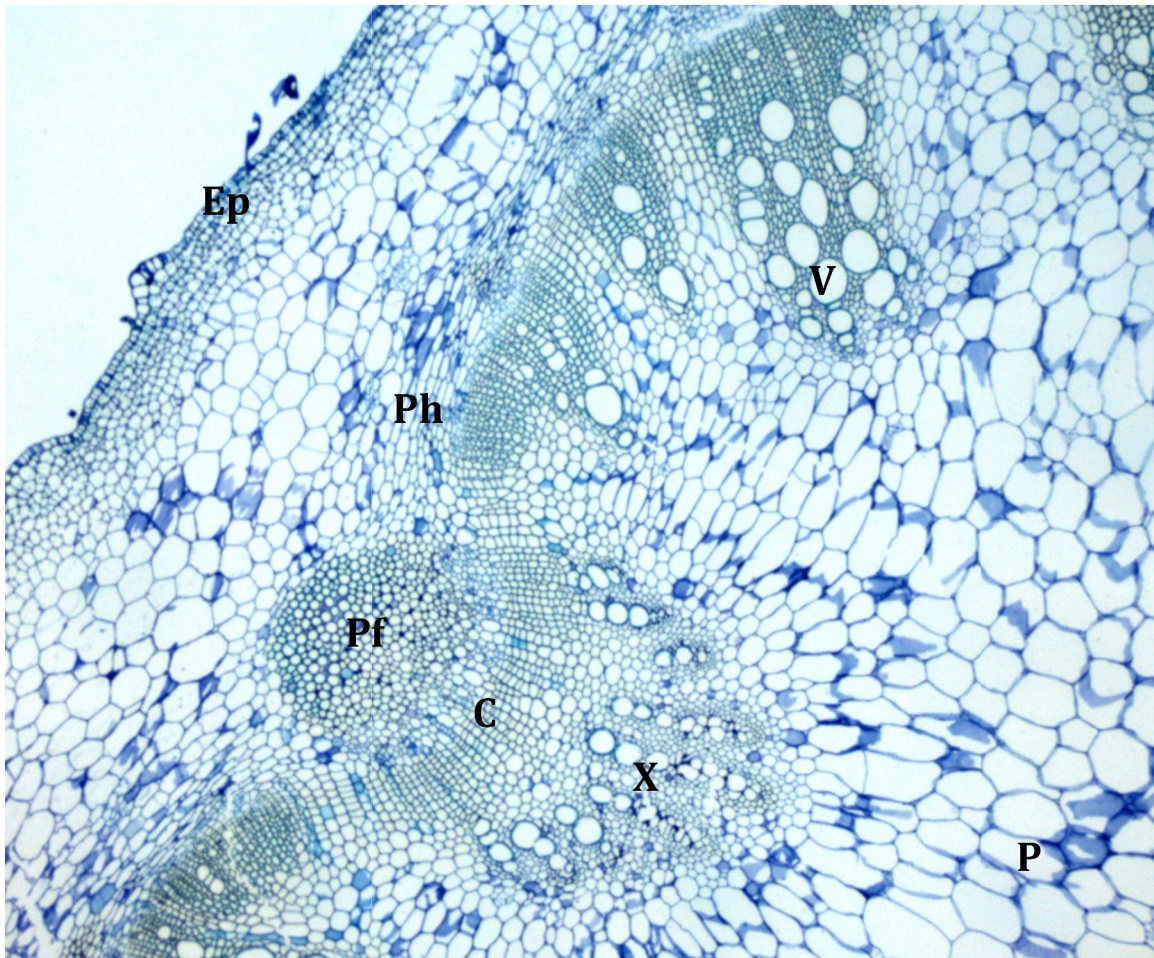


Figure 1b. Paraffin embedded and stained with toluidine blue transverse section of *Helianthus annuus* at 8 weeks. Pith (P), xylem (X), vessels (V), vascular cambium (C), phloem (Ph), phloem fiber (Pf), and epidermis (Ep).

Compound	Arg1820 (%)	Pine (%)
Arabinan	0.2	1.3
Galactan	0.9	2.1
Glucan	38.4	40.2
Xylan	18.6	6.5
Mannan	0.8	9.7
Total Sugars	58.9	59.8
Lignin	20.5	31.2
Solids	79.4	91.0

Table 1.
Chemical profiles of *H. argophyllus* (Arg1820) and *P. taeda* wood samples analyzed by HPLC (Kaar et al. 1991) at the Weyerhaeuser Analytical and Testing Services Facility (Federal Way, WA)

CHAPTER 2: EXPRESSION ANALYSIS OF WOOD-FORMING *HELIANTHUS* *ARGOPHYLLUS*

Introduction

Biofuel production is a rapidly growing area of research as fossil fuels are becoming scarcer and climate change caused by the burning of fossil fuels threatens the environment (Dellomonaco et al. 2010). Currently biofuels are being produced from such varied feedstocks as sawdust, grass cuttings, agricultural waste, and dried manure (Briens et al. 2008). The main feedstocks for biofuel (ethanol) production have been sugars and starch from corn, wheat, barley, and sugarcane. However, the use of corn and other common food commodities has effects on food costs, and biofuel production has already caused concerns in the United States, where food prices have increased with biofuel production (Odling-Smee, 2007). Another issue for production of biofuel using sugar and starch crops is that biofuel production can actually result in increased emissions of carbon dioxide into the environment due to the use of energy-intensive fertilizers and pesticides (Hill et al. 2006). In addition to the detrimental environmental effects of first-generation biofuels, there are economic effects, particularly in poorer countries that struggle to provide enough food for their people (Gomez et al. 2008). While these “first-generation” biofuels appear to be unsustainable because they place stress on the food industry, “second-generation” biofuels are a new direction in which cheap and abundant non-food plant biomass is used in the production of biofuels (Gomez et al. 2008).

The need for a more environmentally and economically efficient source of biofuel is urgent, and the use of plant biomass may be the answer. Ideally, second-generation

biofuels are produced from dedicated biomass crops that yield significant amounts of biomass in a short period of time on with the least amount of land and with little use of fertilizers and pesticides (Gomez et al. 2008). Lignocellulosic biomass is an abundant and underutilized biological resource and is considered a primary source of material for second-generation biofuels production (US DOE, 2006). With excessive carbon dioxide emission being a drawback of first-generation biofuels, second-generation biofuels, including perennial grasses such as switchgrass, as well as woody species, such as poplar and willow, are expected to be carbon neutral or better yet, carbon negative (Gomez et al. 2008). Life cycle assessments (LCA) used to assess the sustainability of biofuels show that second-generation biofuels may offer a lower carbon footprint than their predecessors (Sheehan 2009).

Perennial grasses, poplar, and willow are sources of biomass commonly considered for use in second-generation alternative fuel development. Another candidate for consideration is the common sunflower (*Helianthus annuus*). However, common sunflower has so far only been grown as a source of oilseed, which can be used for biodiesel production, but the remainder of the plant has not been considered as a source of lignocellulosic biomass for biofuel production (Beckman et al. 2008). This is primarily due to the low-density, pithy stems that characterize common sunflowers. These low-density stems contain relatively little carbon that could be converted to biofuel. However, in desert environments, there are species of sunflowers that do produce high-density woody stems, including the silverleaf sunflower (*Helianthus argophyllus*) and the Algodones dune sunflower (*Helianthus niveus*). Focusing on the woody *Helianthus*

argophyllus in contrast to the pithy *Helianthus annuus* allows insight into the genetics that make them drought tolerant and lead them to produce wood.

Plant biomass, or lignocellulose, is composed of cellulose, hemicellulose, and lignin. All three are involved in the development of plant structure. Specifically, secondary plant cell walls make up the greater composition of lignocellulosic biomass (Gomez et al. 2008). Cellulose is a linear chain of glucose sugars that are synthesized by the cellulose synthase (CesA) protein. Six CesA proteins are arranged in what is called a cellulose synthase complex, and six complexes are arranged in a rosette, which is a plasma membrane responsible for the production of cellulose microfibrils (Figure 2.) Hemicelluloses are linear polymers that bind to cellulose microfibrils and prevent them from touching each other and in turn arranging a complex network of cellulose microfibrils (Figure 2.). The third component of lignocellulosic biomass is lignin, which is critical in both the plants physical structure and plays key roles in defense. Lignin serves as the sheath for the cellulose microfibril and hemicellulose network (Figure 3.) (Somerville et al. 2009). In order to understand the regulatory roles of plant biomass development, CesA and other genes of interest are investigated to understand their expression in plant development and eventually biomass accumulation.

The cellulose synthase (CesA) gene family is known to be regulated in the synthesis of wood and in the cell wall biosynthesis pathway (Kalluri et al 2009). The synthesis of cellulose is vital in the development of woody tissue in plants. It has been shown in *Populus trichocarpa*, the model organism for the study of wood development, that many CesA genes are xylem specific, and crucial in the understanding of the synthesis of wood (Suzuki et al 2006). In primary cell walls, cellulose is important in

maintaining cell shape and cell expansion, which is essential during plant growth. Cellulose also is essential in secondary cell walls, where it has its role in mechanical strength for the plant (Wightman et al. 2010). In specific cell types (xylem and phloem) primary and secondary cell walls differ in the quantity and quality of cellulose deposits (Kalluri et al. 2009). These differences are seen in the higher amounts of greater degree of polymerization and crystallinity type cellulose found particularly in secondary cell walls (Delmer et al. 1999). The cellulose synthase gene family regulates and controls this specified synthesis of cellulose and its integration in primary and secondary cell walls. The cellulose synthase (*CesA*) gene family contains different members that encode CesA proteins for primary and secondary cell wall biosynthesis (Nair et al. 2011).

The COBRA gene encodes a glycosylphosphatidylinositol (GPI)-anchored plant specific protein. It belongs to a gene family involved in cellulose synthesis in secondary cell walls and the development of plant strength (Dai et al. 2011). COBRA protein is a key regulator of diffuse anisotropic expansion throughout postembryonic development in *Arabidopsis thaliana*. The GPI membrane anchor is important in modifications of cell wall proteins and cellulose deposit in secondary cell wall synthesis (Roudier et al 2005). The COBRA-like protein is known to play a role in the development of secondary cell walls. There are various secondary cell wall types that distinguish themselves between woody and non-woody tissue. It has been shown in previous studies that mutants of COBRA-like proteins in rice have been found to show brittle skeletal structures, inferring that COBRA-like proteins play a crucial role in the deposition of cellulose in secondary walls (Kotake et al. 2011). The secondary wall is a coordinated arrangement of polymers

and proteins for mechanical structure and support of the plant body. With the lack of a highly ordered assembly, plant structure is jeopardized (Sato et al. 2010).

Gene expression profiling is one technique for identifying and characterizing regulatory roles of genes in complicated developmental pathways. In particular, large-scale genomic approaches allow us to follow the expression of many genes in parallel simultaneously, which facilitates identification of gene networks having common patterns of expression (Lai et al. 2005). Gene expression analysis should aid us in understanding the gene expression differences leading to wood formation in desert sunflower ecotypes versus common sunflower. Since cellulose, hemicellulose, and lignin are the primary components of angiosperm wood, differential expression in genes involved in their production will help identify candidate genes that may play critical roles in wood formation in the desert sunflowers (Somerville et al. 2006, Persson et al. 2005). Previous work has shown that secondary cell wall formation is regulated and activated by various genes and transcription factors (Zhong et al. 2010). Additionally, a large number of genes involved in cellulose, hemicellulose, and lignin biosynthesis have been identified, and the coordinated gene expression of these genes has been demonstrated in other wood-forming species (Lerouxel et al. 2006, Somerville et al. 2006, Zhong et al. 2006). Gene expression analysis should identify sunflower homologs to the genes shown to play key regulatory and biosynthetic roles in secondary cell wall formation in other wood forming species (McCarthy et al. 2010).

Materials and Methods

Plant Materials

Two species of the sunflower were grown; the silverleaf sunflower, *Helianthus argophyllus*, and the common sunflower, *Helianthus annuus*. Seeds were acquired through the Crop and Soil Science Department at the University of Georgia (Athens, GA, USA) and grown on site. The silverleaf sunflower (*Helianthus argophyllus*) is accession Arg1820 and the common sunflower (*Helianthus annuus*) is accession HA412-HO.

Arg1820 seeds must be germinated prior to being placed in soil. Seeds are placed on water-saturated filter paper and enclosed in a plastic box for one night (Felitti et al. 1997). After the seed has absorbed moisture overnight, the top of the seed is cut off with a razor blade and placed back into the plastic box. Seed germination then occurs within 2-3 days and seedlings are transferred to soil. Seedlings are transplanted to flats and grown in a growth chamber for four weeks at 25°C/20°C, 16 hour/8hour light/dark cycles. Plants are then transferred to pots and moved to the greenhouse at 25°C with 16-hour light cycles in addition to natural sunlight.

The HA412-HO sunflower however requires no preparatory work and is directly placed into the soil for germination. Plants are grown in flats in a growth chamber for four weeks at 26°C/20°C, 14 hour/10hour light/dark cycles. Plants are then transferred to pots and moved to the greenhouse at 25°C with 16-hour light cycles in addition to natural sunlight.

Fertilizer is applied twice a week (Tuesday and Thursday) and watered three times a week (Monday, Wednesday, Friday). HA412-HO and Arg1820 plants were

harvested for light microscopy and laser microscopy dissection analysis at “4-week”, “8-week”, and “12-week” time points.

Due to the different maturation speed of HA412-HO and Arg1820, where Arg1820 has a much longer vegetative emergence stage, an arbitrary “0-week” point is set with the appearance of the HA412-HO from the dirt. From this time point, samples are collected at the “4-week”, “8-week”, and “12-week” time points.

Anatomical Analysis

Samples for physical comparison were fixed with 4% formaldehyde. Approximately 5mm sections are taken from the base, 3rd node, and the top node at the “8-week” time point from both Arg1820 and HA412-HO. Sections are fixed overnight at 4°C. A dehydration series is conducted over a course of 3 days starting from 50% ethanol with gradual a phase transfer into tert-butyl alcohol, with overnight steps at 70% ethanol and 100% ethanol. Samples are embedded in paraffin wax and cross-sections of 15um are made for analysis. Samples were stained with toluidine blue.

EST Sequencing

Dr. Sukhpreet Sandhu conducted the following experimental procedures. Non-normalized cDNA was prepared from RNA of Arg1820 and HA412-HO stem samples. Triplicate stem samples from first and third internode were collected from both Arg1820 and HA412-HO at week 4, 8, and 12. RNA was extracted from pooled Arg1820 and pooled HA412-HO ground stem samples using TRIzol as per manufacturer’s instructions (Invitrogen, CAT#15596-026). Following layer separation with chloroform, RNA was

pelleted with isopropanol and cleaned with 75% ethanol. RNA was resuspended in 1x TE and stored at -20C. Total RNA was quantified on the NanoDrop ND-1000 (Thermo Scientific, Waltham, MA, USA) and checked for quality on the Agilent 2100 BioAnalyzer (Agilent Technologies, Santa Clara, CA, USA). cDNA was prepared using the SMART MMLV Reverse Transcriptase kit (Clontech) as per manufacturer's instructions. Sequencing of cDNA was done using the 454-FLX Titanium platform (Roche, Penzberg, Upper Bavaria, Germany). The reads were processed and followed by assembly on a CAP3 assembler (Huang et al. 1999).

Selection of Unigene Sets

Dr. Sukhpreet Sandhu conducted the following experimental procedures. Stem 454 ESTs generated by pyrosequencing were pooled with existing Sanger and SOLEXA EST. To reduce the redundancy of unigenes, the new ESTs dataset was queried using BLAST with the older data sets. The resulting set of non-duplicated sequences comprised the initial anchor Unigene set. The following strategy was used to identify non-redundant unigenes.

Non-redundant ESTs between combined Arg1820 ESTs (454 and Sanger) and combined HA412-HO ESTs were identified by BLASTn, named Unigene Set 1. Out of the redundant set, 1300 unigenes were selected to be represented twice on the array to allow comparison between Arg1820 and HA412-HO gene complements. The 1300 unigenes represented 800 highly expressed unigenes and 500 randomly selected for probe design. All remaining Sanger ESTs were pooled and filtered to obtain non-redundant Unigene Set 2. Likewise, all SOLEXA ESTs were pooled and filtered, yielding Unigene

Set 3. Finally, Unigene set 1, 2 and 3 were pooled and subjected BLASTn to reduce redundancy. Thus yielding the Anchor Unigene Set. This Anchor Set and 1300 redundant unigenes was then subjected to BLASTX with six reference genomes namely, *Arabidopsis thaliana*, *Populus trichocarpa*, *Vitis vinifera*, *Mimulus guttatus*, *Glycine max*, and *Oryza sativa*. The ESTs that showed at least one match with one or more of the reference genomes were used for probe design. Remaining ESTs with no match with the reference genome were sorted by size of the estimated protein and only >400bp ORF were selected for probe design. The selected ESTs were submitted to Agilent e-array platform for the probe design. Probe selection for 4x44K Agilent array was done by Agilent and 10 One-Color Arrays (for 40 hybridizations) were constructed. For all the non-redundant ESTs, a single probe was designed and for the redundant 1300 ESTs, and two probes were designed.

Microarray Hybridizations

Dr. Sukhpreet Sandhu conducted the following experimental procedures. Fresh RNA was extracted from each individual sample including 3 replicates from Arg1820 plants, and two internode samples from each plant; likewise for HA412-HO plants. All samples were checked for quality with the Agilent 2100 BioAnalyzer (Agilent Technologies, Santa Clara, CA, USA). Samples with RIN below 7.0 were discarded and RNA was re-extracted. Prior to hybridization, RNA was quantified on the NanoDrop ND-1000 (Thermo Scientific, Waltham, MA, USA). Agilent's One-Color Quick Amp Labeling Kit was used for RNA target labeling. Further, Agilent's Gene Expression Kit was used for array hybridization. Labeling and hybridization procedures were conducted

as per manufacturer's instructions. The arrays were scanned with an Agilent High-Resolution Microarray Scanner (Agilent Technologies, CAT#G2505, Santa Clara, CA, USA) and images were extracted with Feature Extraction program (v10.5).

Quantitative PCR

Prior to cDNA synthesis step, 1 μ g RNA from sample was treated with TURBO DNA-free Kit (Ambion, Cat. No.AM1907) as per manufacturer's instructions. cDNA was synthesized from cDNA was synthesized using SuperScriptIII First Strand Synthesis for RT-PCR (Invitrogen, Cat. No. 18080-51) as per manufacturer's instructions. Primer pairs from 18 candidate genes were designed using Primer3. RT-qPCR was conducted using the Bio-Rad SYBR Green Super Mix on the Bio-Rad iCycler following manufacturer's protocol (Bio-Rad, Hercules, CA, Provenzano et al. 2007, Zhou et al 2006). Control gene GAPDH was used for RT-qPCR and RT-qPCR results were interpreted for gene quantification (Fernandez et al. 2008, Roche et al. 2009). Samples were run in triplicate for each primer pair. Threshold cycle (Ct) values were deduced from the iQ5 iCycler (Bio-Rad, Hercules, CA, USA) software and normalized data was calculated using GAPDH and ACT2 as the control gene.

Results

Sequencing

RNA from stems of *Helianthus annuus* (Ann) and *Helianthus argophyllus* (Arg) were sequenced on a 454-Titanium FLX platform. A total of 305900 and 597654 raw read reads were produced for Ann and Arg respectively. The raw reads were processed

with CAP3 assembler, thus yielding 22171 ESTs for Ann and 33393 ESTs for Arg. This EST selection and microarray features EST selection for microarray probe design was done by pooling ESTs from all available sources and then eliminating redundancy through BLAST. First we pooled all Arg ESTs including 35721 Sanger ESTs from NCBI and 33393 454 ESTs (produced as above). Sanger ESTs were run through a BLAST with 454 ESTs. BLAST results identified 17,384 ESTs that were non-redundant (<95% similarity) in the Sanger ESTs. This analysis confirmed (33,392 454 ESTs+ 17,384 Sanger ESTs) 50,776 non-redundant Arg ESTs. Similar approach was used for HA412 specific Sanger and 454 ESTs . This confirmed (22171 454 ESTs+16502 Sanger ESTs) 38674 non-redundant Ann 412 ESTs. Thirdly, we pooled all available Ann ESTs (excluding HA412-HO ESTs). We obtained a total of 62,945 Sanger ESTs and 42,823 Solexa ESTs. These three unigene pools were sequentially run through BLAST to detect further redundancy and produced 84,953 unique ESTs (redundant were excluded >95% similarity for >100 bp).

Further, these 84953 ESTs were sorted by BLASTx using reference peptide database from six sequenced genomes. The reference genomes used for this analysis were *Arabidopsis thaliana*, *Populus trichocarpa*, *Vitis vinifera*, *Mimulus guttatus*, *Glycine max*, *Oryza sativa*. Out of 84,953 ESTS, we found 57866 that matched to at least one of the peptide sequences in one or more of the reference genomes. The remaining 27087 ESTs that did not match to any known peptide sequence were sorted by length of the predicted protein. Based on size distribution, we selected 1161 ESTs with ORF >400bp. A total of 58397 +1161 non-redundant unigenes, (referred to as the anchor unigene set)/

was obtained. As a control, we also included 1300 ESTs that were redundant between Arg1820 and HA412-HO ESTs.

EST sequences from the anchor unigene set and redundant unigene set were submitted to the Agilent e-array online tool. One probe per EST was designed for the anchor unigene set, and 2 probes per EST were designed from the redundant unigenes. The format of microarray was 4x44 and included 473 controls with 172 negative controls and remaining positive controls. A total of 10 arrays (for 40 samples) were constructed.

Transcriptome Analysis with Microarray Data

RNA from individual Arg1820 and HA412-HO samples were hybridized. The feature extraction software (Agilent) yielded raw intensity signals. GeneSpring GX (Agilent, CAT#G3784AA, Santa Clara, CA, USA) was used for interpretation of intensity signal values. Filtering of all entities for expression in at least 60% of replicates and background value 50 yielded 28935 entities. The filtering process was done to eliminate intensity signals that represented actual data rather than background noise and the 60% cutoff to ensure consistent values across replicates. Based on t-ttest ($p < 0.05$), 17715 entities showed significant difference in expression value in a 2x2 model (2 genotype x 2 tissue types). Out of these, 7533 entities differentially expressed between Arg1820 and HA412-HO by fold change (FC) > 2 as shown in the volcano plot in Figure 4. Figure 5 shows a plot of differentially expressed genes between Arg1820 and HA412-HO at Internode 1 in Week 8.

Validation of Candidate Genes RT-qPCR

Candidate genes based on fold change and known genes to play regulatory roles in secondary xylem biosynthesis were selected. The selected genes were chosen from the 7533 differentially expressed genes that showed a fold change greater or equal to 2. To verify further the microarray data, 17 genes were selected with different levels of expression both up and down in comparison in addition to 4 housekeeping genes (ACT1, ACT2, PAL1, TUB2). RT-qPCR results are reflective of the observations made from the microarray data in regards to their fold-change in comparison between the HA412-HO pithy stem and Arg1820 woody tissue (Figure 6). RT-qPCR validation shows an approximate 5 fold change in cellulose synthase (0.24 microarray), a 2 fold difference in COBRA-LIKE4 protein (0.40 microarray), and 5 fold change in NAC2 (7.32 microarray). Even though some candidate genes showed a variation in expression level in comparison between the microarray and RT-qPCR data, the selected genes were in good agreement with those of the microarray analysis in terms of up or down expression.

Discussion

Helianthus argophyllus has shown phenotypically that its stem structure differs from that of the common sunflower, *Helianthus annuus* (Figure 7) While the common sunflower develops a pithy stem, *argophyllus* produces a woody structure much stronger than its species counterpart. The differential expression profiling conducted with the microarray produces data showing this explicit difference in structure as numerous genes with regulatory roles in the development of secondary xylem are expressed at different levels.

The cellulose synthase (CesA) gene family is known to be regulated in the synthesis of wood and in the cell wall biosynthesis pathway (Kalluri et al 2009). Microarray and RT-qPCR data however shows a negative fold change in regards to cellulose synthase expression levels in Arg1820 compared to HA412-HO. This could be due to the role that cellulose synthase plays in general cell wall biosynthesis. A total of six CesA proteins have been shown to be required for cell wall biosynthesis in vascular tissues of *Arabidopsis thaliana*, with three associated with primary cell wall development (CesA1, CesA3, CesA6) and three associated with secondary cell wall development (CesA4, CesA7, CesA8) (Dai et al. 2011, Gardiner et al. 2003, Tanaka et al. 2003). BLAST annotations from the microarray design show that the cellulose synthase in question showing the negative fold change in comparison of Arg1820 and HA412-HO is design based off CesA3, a known cellulose synthase in primary cell wall development. This explains the negative fold change as secondary cell walls are the critical components of wood development not primary cell walls (Gardiner et al. 2003).

The COBRA-like protein is known to play a role in the development of secondary cell walls. Proteins orthologous to the COBRA-like protein in various species of plants have all shown similar phenotypes to elude the role COBRA-like protein plays in secondary wall development. Barley *brittle culm (bc)*, *Arabidopsis irregular xylem (irx)*, and maize *Brittle Stalk 2 (Bk2)* mutants have all shown phenotypes of reduced cellulose content in secondary walls, reduced stiffness of internodes, disordered secondary wall structure, and effects on cell expansion orientation (Ching et al. 2006, Sato et al. 2010). Mutations in COBRA-LIKE4 (CBL4) in *Arabidopsis*, homologous to *BC1*, resulted in weaker stems and an *irx* phenotype (Brown et al. 2005).

The microarray data showed only a slight fold change (0.40) between HA412-HO and Arg1820 tissues, as both species develop secondary cell wall structures to support their plant mass. Albeit only showing a slight fold change between HA412-HO and Arg1820 from the microarray analysis, the RT-qPCR validation resulted in a greater fold change in expression of CBL4. This is a technical issue in regards to microarray analysis due to the dynamic range of microarray systems that limits the comparative analysis due to statistical boundaries (Sharov et al. 2004).

Numerous transcription factors play a role in wood formation, however only the NAC2 domain has been shown to play a key regulation in the biosynthesis of all three key components of wood (Zhong et al. 2010). Previous studies in *Arabidopsis thaliana* has shown the downstream regulation that NAC domains play in secondary cell wall synthesis. Its upregulation alone leads to secondary wall thickening, deposition of ligning, xylan, and cellulose, resulting in an increased stem biomass (Wang et al. 2010). In compliment to the microarray data showing a upward fold change of NAC2 (7.32) and a RT-qPCR validation, it is understandable why Arg1820 would have higher expression of NAC2 in its stem tissue compared to the much less dense stem of HA412-HO.

Conclusion

Gene expression profiling with the use of the microarray is a quick overview approach in discovering genetic markers and the study of candidate genes. The microarray assessment of *Helianthus argophyllus* and *Helianthus annuss* has provided a list of over 8000 differentially expressed genes which can be further narrowed down and investigated for specific regulatory roles. In the study of wood formation, select few

candidate genes have been validated with RT-qPCR and confirmation of microarray data has been provided. However, in this experiment the study looks at whole sections of the plant stems and do not focus in on specific cell types, which are critical in the study of wood formation, as certain genes are only expressed in specific cell types. As is the case in other higher eukaryotes, plant tissue structure derives from a myriad of various cell types, each characterized by unique gene expression patterns that ultimately contribute to the overall growth and development patterns of the organism. The non-uniformity of gene expression between various cell types creates a layer of ambiguity for gene expression profiling. In addition to specific cell types, another fallacy that may fall upon the microarray data analysis is the fact that certain cell types might be more prevalent in some tissue compared to others. This could very well skew the expression level of certain genes one-way or the other. In fact, the use of an entire stem also could have diluting effects with the addition of the pith and epidermal cells, which would not have high levels of the candidate genes that were focused on.

References

1. Beckman, W. 2008. Ancient Sunflower Fuels Debate About Agriculture in the Americas. Science Daily.
2. Briens, C, Piskorz, J, Berruti F. 2008. Biomass Valorization for Fuel and Chemicals Production – A Review. International Journal of Chemical Reactor Engineering 6: R2
3. Brown DM, Zeef LA, Ellis J, Goodrace R, Turner SR. 2005. Identification of novel genes in Arabidopsis involved in secondary cell wall formation using expression profiling and reverse genetics. Plant Cell 17(8): 2281-2295
4. Ching A, Dhugga KS, Appenzeller L, Meeley R, Bourett TM, Howard RJ, Rafalksi A. 2006. Brittle stalk 2 encodes a putative glycosylphosphatidylinositol-anchored protein that affects mechanical strength of maize tissues by altering the composition and structure of secondary cell walls. Planta 224(5): 1174-1184
5. Dai X, You C, Chen G, Li X, Zhang Q, Wu C. 2011. OsBC1L4 encodes a COBRA-like protein that affects cellulose synthesis in rice. Plant Molecular Biology 75(4-5): 333-345
6. Dellomonaco, C. 2010. The path to next generation biofuels: successes and challenges in the era of synthetic biology. Microbial Cell Factories 9:3
7. Delmer DP. 1999. Cellulose Biosynthesis: Exciting Times For A Difficult Field of Study. Annual Review Plant Physiology Plant Molecular Biology 50: 245-276
8. Felitti, SA, Chan, RL, Gago, G, Valle, EM, Gonzalez, DH. 1997. Expression of sunflower cytochrom c mRNA is tissue-specific and controlled by nitrate and light. Physiologia Plantarum. 99: 342-347

9. Gardiner JC, Taylor NG, Turner SR. 2003. Control of cellulose synthase complex localization in developing xylem. *Plant Cell* 15(8): 1740-1748
10. Gomez, LD, Steele-King, CG, McQueen-Mason, SJ. 2008. Sustainable liquid biofuels from biomass: the writing's on the walls. *New Phytologist* 178: 473-485
11. Hill J, Nelson E, Tilman D, Polasky S, Tiffay D. 2006. Environmental economic, and energetic costs and benefits of biodiesel and ethanol biofuels. *Proceedings of the National Academy of Sciences, USA* 103: 11206-11210
12. Huang X, Madan A. 1999. CAP3: A DNA sequence assembly program. *Genome Resources* 9(9): 868-877
13. Kalluri UC, Hurst GB, Lankford PK, Ranjan P, Pelletier DA. 2009. Shotgun proteome profile of *Populus* developing xylem. *Proteomics* 9(21): 4871-80
14. Kotake T, Aohara T, Hirano K, Sato A, Kaneko Y, Tsumuraya Y, Takatsuji H, Kawasaki S. 2011. Rice Brittle culm 6 encodes a dominant-negative form of CesA protein that perturbs cellulose synthesis in secondary cell walls
15. Lai Z, Livingstone K, Zou Y, Church SA, Knapp SJ, Andrews J, Riesberg LH. 2005. Identification and mapping SNPs from ESTs in sunflower. *Ther Appl Genet* 111: 1532-1544
16. Lerouxel O, Cavalier DM, Liepman AH, Keegstra K. 2006. Biosynthesis of plant cell wall polysaccharides - a complex process. *Current Opinion Plant Biology* 9(6): 621-630
17. Nair M, Debolt S. 2011. Analysing cellulose biosynthesis with confocal microscopy. *Methods in Molecular Biology* 715: 141-152
18. Odling-Smee L. 2007. Biofuels bandwagon hits a rut. *Nature* 446: 483

19. Persson, S, Wei, H, Milne, J, Page, G.P. 2005 Large-scale coexpression analysis reveals novel genes involved in cellulose biosynthesis. *Proc. National Academy of Science USA*. 102: 8633-8638
20. Provenzano M, Mocellin S. 2007. Complementary Techniques: Validation of Gene Expression Data by Quantitative Real Time PCR. *Microarray Technology and Cancer Gene Profiling*.
21. Roudier F, Fernandez AG, Fujita M, Himmelspach R, Borner GH, Schidelman G, Song S, Baskin TI, Dupree P, Wasteneys GO, Benfey PN. 2005. COBRA, an Arabidopsis extracellular glycosyl-phosphatidyl inositol-anchored protein, specifically controls highly anisotropic expansion through its involvement in cellulose microfibril orientation. *Plant Cell* 17(6): 1749-1763
22. Sato K, Ito S, Fujii T, Suzuki R, Takenouchi S, Nakaba S, Funada R, Sano Y, Kajita S, Kitano H, Katayama Y. 2010. The carbohydrate-binding module (CBM)-like sequence is crucial for rice CWA1/BC1 function in proper assembly of secondary cell wall materials. *Plant Signal Behavior* 5(11): 1433-1436
23. Sharov V, Kwong KY, Frank B, Chen E, Hasseman J, Gaspard R, Yu Y, Yang I, Quackenbush J. 2004. The limits of log-ratios. *BMC Biotechnology* 8(4):3
24. Sheehan, JL. 2009. Biofuels and the conundrum of sustainability. *Current Opinion in Biotechnology* 20: 318-324
25. Somerville C.R. 2006. Cellulose synthesis in higher plants. *Cell Developmental Biology*. 22: 53-78

26. Suzuki S, Li L, Sun YH, Chiang VL. 2006. The cellulose synthase gene superfamily and biochemical functions of xylem-specific cellulose synthase-like genes in *Populus trichocarpa*. *Plant Physiology* 142(3): 1233-1245
27. Tanaka K, Murata K, Yamazaki M, Onosato K, Miyao A, Hirochika H. 2003. Three distinct rice cellulose synthase catalytic subunit genes required for cellulose synthesis in the secondary wall. *Plant Physiology* 133(1): 73-83
28. United States Department of Energy (US DOE). 2006. Breaking the biological barriers to cellulosic ethanol: a joint research agenda.
29. Wang H, Avci U, Nakashima J, Hahn MG, Chen F, Dixon RA. 2010. Mutation of WRKY transcription factors initiates pith secondary wall formation and increases stem biomass in dicotyledonous plants. *Proc National Academy of Science USA* 107(51): 22338-22343
30. Wightman R, Turner S. 2010. Trafficking of the plant cellulose synthase complex. *Plant Physiology* 153(2): 427-432
31. Zhou G, Richardson R, Morrison E, Nairn J. 2006. The Poplar glycosyltransferase GT47C is functionally conserved with Arabidopsis Fragile Fiber8. *Plant Cell Physiology* 47:1229-1240
32. Zhong R, Demura T, Ye ZH. 2006. SND1, a NAC domain transcription factor, is a key regulator of secondary wall synthesis in fibers of Arabidopsis. *Plant Cell* 18(11): 3158-3170
33. Zhong R, Yu ZH. 2010. The poplar PtrWNDs are transcriptional activators of secondary cell wall biosynthesis. *Plant Signal Behavior* 5(4)

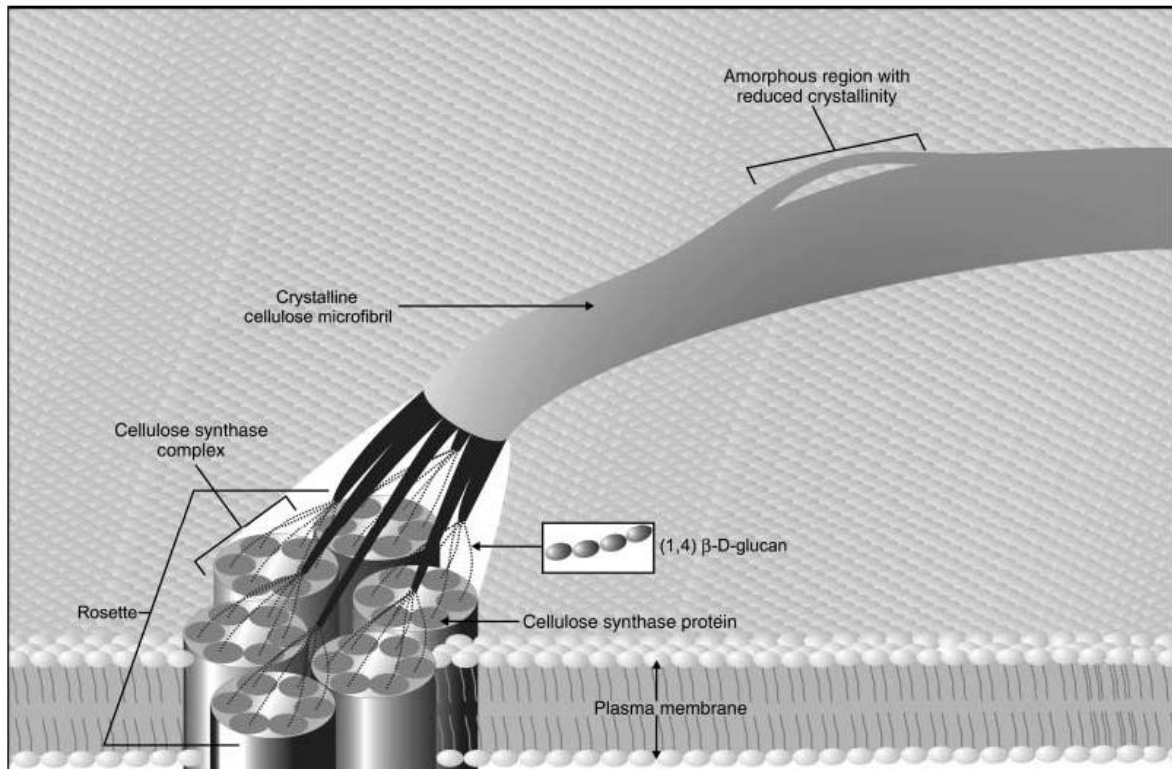


Figure 2. Cellulose microfibrils are synthesized by large protein complexes (rosettes) embedded in the plasma membrane. Each lobe of the rosette is thought to contain six cellulose synthase catalytic sites producing cellulose chains in a coordinated fashion, giving rise to subfibrils, which come together to form crystalline microfibrils, with occasional areas of discontinuity, or amorphous regions (Gomez et al. 2008).

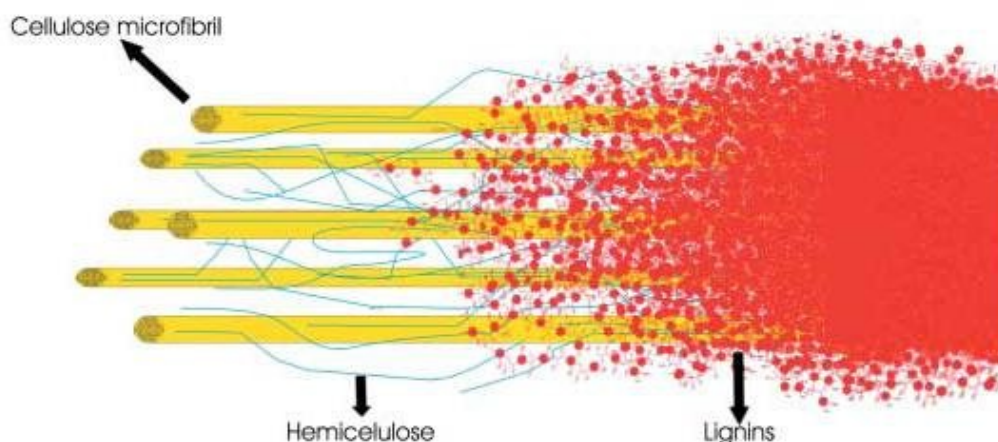


Figure 3. Schematic representation of the generalized components of plant secondary cell walls. Cellulose microfibrils provide the structural framework of the wall and these are associated with a coating of hemicellulosic polysaccharides that hydrogen-bond to the microfibril surface and span the distance between fibrils, effectively tethering them to one another. This polysaccharide complex is effectively interpenetrated and encased by lignin, a polyphenolic polymer. In the diagram, the lignin is artificially thinned towards the left hand side of the image, simply to enable the polysaccharide components to be seen; in lignocellulosic biomass, the polysaccharides are mostly sealed up in the lignin matrix (Gomez et al. 2008).

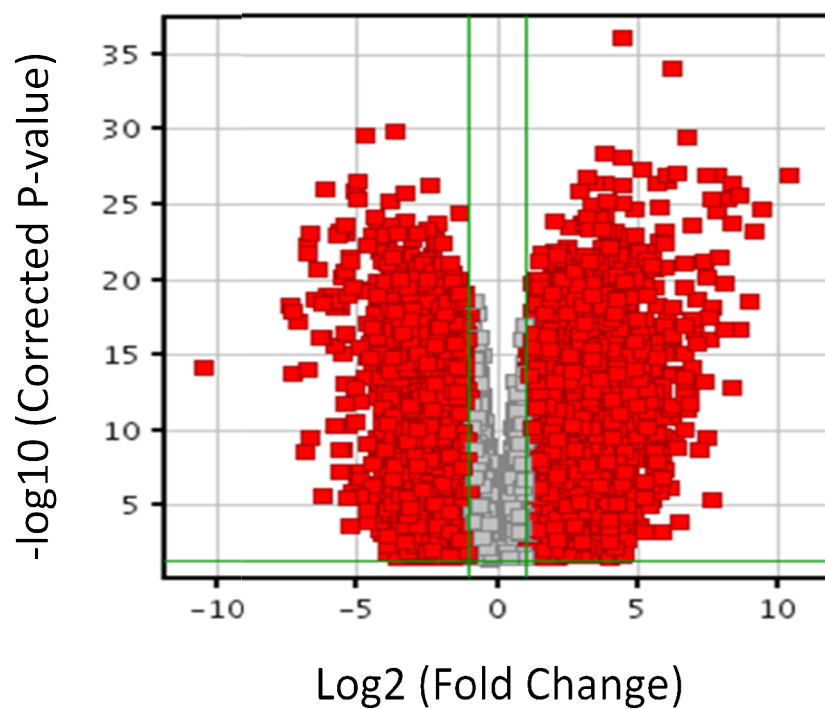
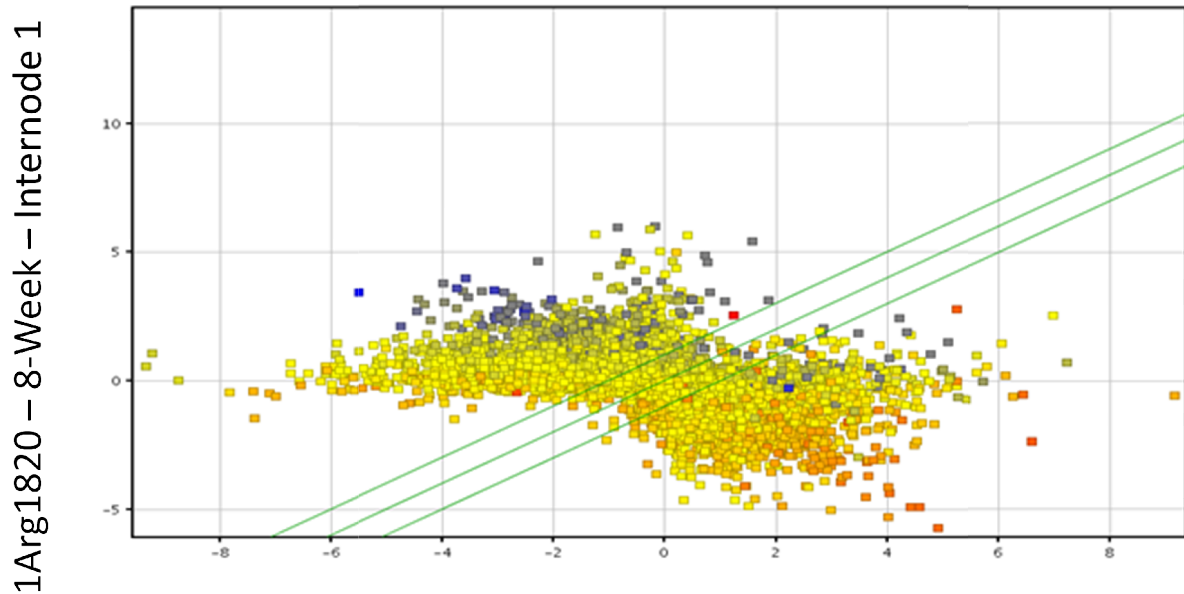


Figure 4. Volcano plot of 17715 differentially expressed entities in a 2x2 (2 genotype – Arg1820, HA412-HO, 2 tissue types – Internode 1 and Internod 3) comparison with 7533 (show in red) entities that have a fold-change greater or equal to 2.



HA4120-HO, 8-Week, Internode 1

Figure 5. Scatter plot comparison of gene expression of Internode 1 between *H. argophyllus* (Arg1820) and *H. annus* (HA4120-HO). The derivation the majority of unique genes from the best fit line characterizes the level of differential expression between the two species at the 8-week timepoint, reflective of the phenotypical differences that are concurrently observed.

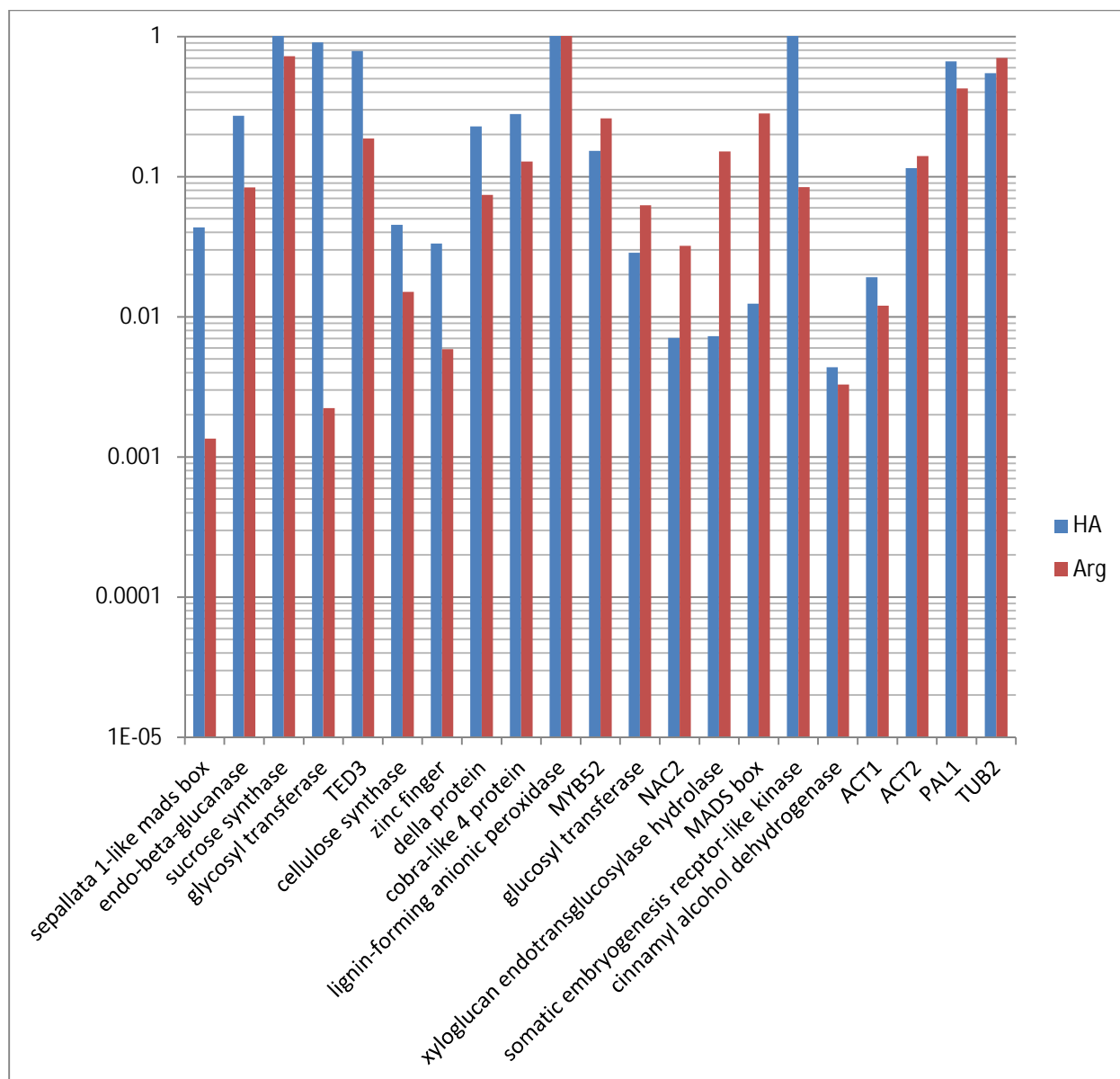


Figure 6. RT-qPCR data validating 17 selected candidate genes observed from the microarray data and 4 housekeeping genes. These candidate genes show the up and down-regulation between HA412-HO and Arg1820 at Internode 1 during Week 8.

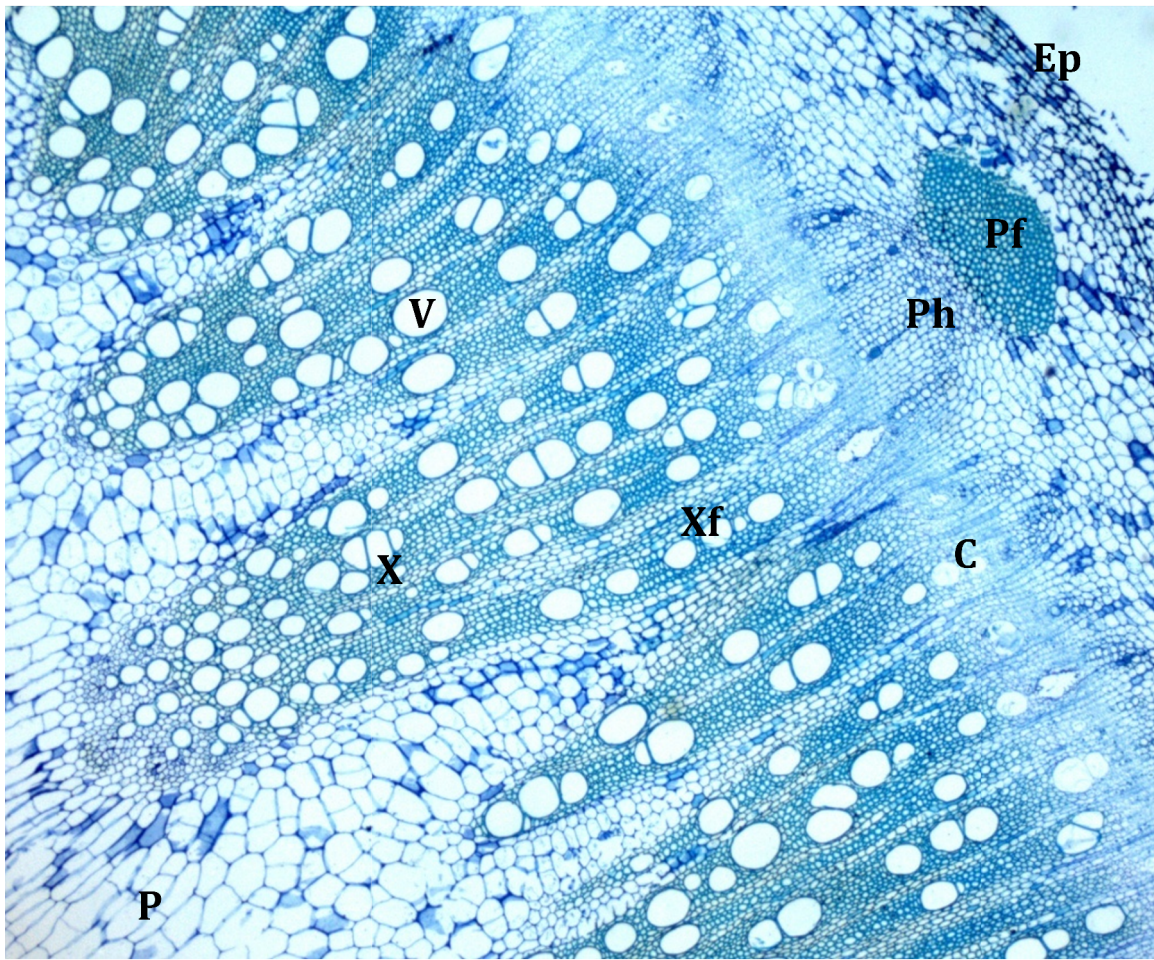


Figure 7a. Paraffin embedded and stained with toluidine blue transverse section of *Helianthus argophyllus* at 8 weeks. Pith (P), xylem (X), xylem fibers (Xf), vessels (V), vascular cambium (C), phloem (Ph), phloem fiber (Pf), and epidermis (Ep).

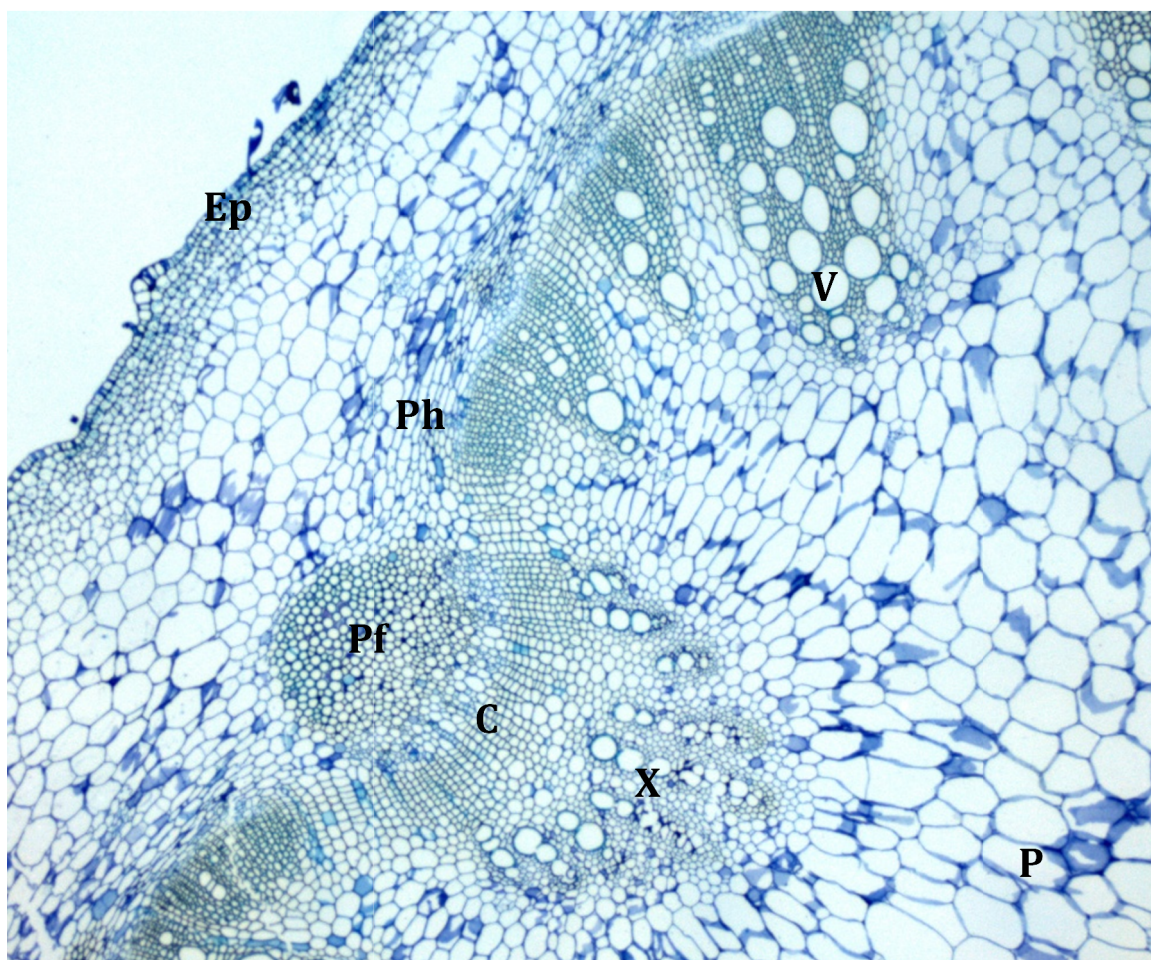


Figure 7b. Paraffin embedded and stained with toluidine blue transverse section of *Helianthus annuus* at 8 weeks. Pith (P), xylem (X), vessels (V), vascular cambium (C), phloem (Ph), phloem fiber (Pf), and epidermis (Ep).

CHAPTER 3: QUANTIFICATION OF DIFFERENTIALLY EXPRESSED GENES IN XYLEM AND PHLOEM TISSUES OF *HELIANTHUS ARGOPHYLLUS*

Introduction

Biofuel production is a rapidly growing area of research as fossil fuels are becoming scarcer and climate change caused by the burning of fossil fuels threatens the environment (Dellomonaco et al. 2010). Currently biofuels are being produced from such varied feedstocks as sawdust, grass cuttings, agricultural waste, and dried manure (Briens et al. 2008). The main feedstocks for biofuel (ethanol) production have been sugars and starch from corn, wheat, barley, and sugarcane. However, the use of corn and other common food commodities has effects on food costs, and biofuel production has already caused concerns in the United States, where food prices have increased with biofuel production (Odling-Smee, 2007). Another issue for production of biofuel using sugar and starch crops is that biofuel production can actually result in increased emissions of carbon dioxide into the environment due to the use of energy-intensive fertilizers and pesticides (Hill et al. 2006). In addition to the detrimental environmental effects of first-generation biofuels, there are economic effects, particularly in poorer countries that struggle to provide enough food for their people (Gomez et al. 2008). While these “first-generation” biofuels appear to be unsustainable because they place stress on the food industry, “second-generation” biofuels are a new direction in which cheap and abundant non-food plant biomass is used in the production of biofuels (Gomez et al. 2008).

The need for a more environmentally and economically efficient source of biofuel is urgent, and the use of plant biomass may be the answer. Ideally, second-generation

biofuels are produced from dedicated biomass crops that yield significant amounts of biomass in a short period of time on with the least amount of land and with little use of fertilizers and pesticides (Gomez et al. 2008). Lignocellulosic biomass is an abundant and underutilized biological resource and is considered a primary source of material for second-generation biofuels production (US DOE, 2006). With excessive carbon dioxide emission being a drawback of first-generation biofuels, second-generation biofuels, including perennial grasses such as switchgrass, as well as woody species, such as poplar and willow, are expected to be carbon neutral or better yet, carbon negative (Gomez et al. 2008). Life cycle assessments (LCA) used to assess the sustainability of biofuels show that second-generation biofuels may offer a lower carbon footprint than their predecessors (Sheehan 2009).

Plant cell walls constrain cell expansion and provide mechanical strength to the plant structure. In general, there are two types of cell walls in plants, primary and secondary, and they are distinguished by their composition as well as their roles in plant development. All plant cells initially synthesize primary cell walls, which help determine the final cell size and shape. In a subset of cells, secondary cell wall formation occurs after growth cessation, and their primary functions are to provide mechanical support and limit loss of critical metabolites, such as water (Dai et al. 2011, Sato et al. 2010). Lignocellulosic biomass generally contains a preponderance of secondary cell walls. One main component differentiating between the two cell wall types is the amount of cellulose content. Specialized cells in the plant, such as xylem and phloem, control and regulate the deposition of primary and secondary cell walls in accordance to the functional demands (Kalluri et al. 2004).

Xylem and phloem, vascular tissues, perform essential roles in the physiology, development, and structure of the plant. With critical roles in the transport of water and nutrients, transport of signaling molecules, and physical support, the vascular tissues form a continuous network for the growth and maturation of plants (Sieburth et al. 2005, Ye et al. 2002). These specific cell types are regulated and originate from the vascular cambium. The specification of distinct cells proliferating from the procambium is highly regulated and results in the desired combination for both physical support and defense against abiotic factors, such as drought, wind, and salinity (Delgado et al. 2010). Xylem is composed of conducting tracheary elements, non-conducting parenchyma cells, and xylem fibers. Phloem is composed of conducting sieve elements connected to companion cells and non-conducting parenchyma cells and fibers (Carlsbecker et al. 2005). With both xylem and phloem, two specific cell types with unique roles in plant development differentiating from the vascular cambium, a complex regulatory pathway is responsible for the mechanism of cell specification and patterning of the vascular tissues for their desired capabilities (Delgado et al. 2010). This specialization of cell types complicates the understanding of gene expression particularly in plant development.

Plant tissue structure derives from a myriad of various cell types, each characterized by unique gene expression patterns that ultimately contribute to the overall growth and development patterns of the organism. The non-uniformity of gene expression between various cell types creates a layer of ambiguity for gene expression profiling that can potentially be deciphered through the application of laser-capture microdissection. Laser-capture microdissection (LCM) combines the traditional microscopy and dissection techniques to produce a cutting-edge technology that allows

for the isolation of specific cell types and in some cases even individual cells (Nelson et al. 2006). Isolating intact RNA from specific cellular bundles from stem cross sections allows for the gene expression profiling of xylem and phloem cells in the study of wood formation in sunflower.

Materials and Methods

Plant Materials

Silverleaf sunflower, *Helianthus argophyllus* (accession Arg1820), and common sunflower, *Helianthus annus* (accession HA412-HO), seeds were acquired from the Department of Energy National Renewal Energy Laboratory (Golden, GO, USA). *H. argophyllus* seeds were pre-germinated by placing them on water-saturated filter paper in an enclosed box overnight (Felitti et al. 1997), after which the top of the seed was nicked with a razor blade. Seeds were then placed back in the box, and germination occurred 2-3 days later. Seedlings were subsequently transplanted to flats containing potting soil and grown in a controlled environment chamber at 25°C/20°C, 16 hour/8hour light/dark cycles for four weeks. Plants were then transferred to pots and moved to the greenhouse where conditions were controlled at 25°C with 16-hour light cycles. *H. annus* seeds were sown directly into soil. After germination, plants in flats were grown as described for *H. argophyllus*. Plants were fertilized twice weekly and watered three times a week. Plants were harvested at 4, 8, and 12 weeks. Observations of phenotype indicated that *H. argophyllus* and *H. annus* growth and development were highly similar through the 4-week time point until woody characteristics began emerging around the 8-week time point.

Anatomical Analysis

Cross-sections (5mm) were taken from the top node, 3rd node, and base node of *H. argophyllus* and *H. annus* at eight weeks. After sectioning, tissues were fixed using protocols modified from Harding et al. (2002). Fixed sections were dehydrated over three days, including overnight incubations at 50% aqueous ethanol, 70% aqueous ethanol, and 100% ethanol. . After the final dehydration series into tert-butyl alcohol samples are embedded in paraffin wax and cross-sections of 15µm are made for analysis. Samples are stained with toluidine blue and observed on a Zeiss Axioskop (Carl Ziess AG, Jena, Germany) at 2.5x magnification.

Laser-capture Microdissection

Internodal stem segments were harvested at four, eight and twelve weeks for laser-capture microdissection (LCM) of specific cell types in developing stem (Kunz et al. 2004). Samples from three plants of each species were collected at each time point. To help maximize the quality of recovered RNA, samples were fixed overnight at 4°C in Farmer's Fixative (3:1 ethanol:glacial acetic acid) (Kerk et al. 2003, Takahashi et al. 2010). After fixation, samples were dehydrated over the course of one day in an ethanol to tert-butyl alcohol series protocol modified from Harding et al (2002). Subsequent paraffin infiltration proceeded overnight (Kerk et al. 2003). Sections (15 µm) of embedded tissue were transferred to a PEM-membrane frame slide (No.11505215, Leica Microsystems, Wetzlar, Germany) with a modified protocol of Abbott et al. (2010). LCM was conducted using a Leica LMD7000 (Leica Microsystems, Wetzlar, Germany) and samples were collected from xylem and phloem tissue in each section. Specific cell type

samples were collected from approximately 20-30 cross sections and triplicates were pooled. Paraffin-embedded sections were stored at 4°C in a desiccated chamber until use and have been shown to have a shelf-life of several weeks under these conditions (Kerk et al. 2003).

RNA Isolation and cDNA Synthesis

The RNeasy Plant Mini Kit (Invitrogen Cat. No. 74904, Life Technologies, Grand Island, NY) was used to purify RNA that was extracted from paraffin-embedded samples using protocols modified from Urieli-Shoval et al. (1992). Briefly, cell samples collected by LCM sequentially underwent two washes in xylene, drying in a vacuum centrifuge, two washes in 100% ethanol, and a final drying in a vacuum centrifuge. Samples were then flash-frozen in liquid nitrogen and processed using the RNeasy Plant Mini Kit per instructions from the manufacturer.

Purified RNA was treated with DNase using the TURBO DNA-free Kit (Ambion Cat. No. AM1907, Life Technologies) and following instructions from the manufacturer to ensure removal of any contaminating genomic DNA (Dyer et al. 2009). RNA quantity and quality were initially assessed by A_{260}/A_{280} analysis using a NanoDrop ND-1000 spectrophotometer (Thermo Scientific, Waltham, MA, USA). Results are shown in Table 2. RNA quality was subsequently assessed in greater detail using an Agilent 2100 BioAnalyzer (Agilent Technologies, Santa Clara, CA, USA), and samples receiving RIN scores >5.0 were deemed suitable for RT-qPCR analyses (Fleige et al. 2006). Three samples (Arg1820-3_Phloem, HA412-2_Phloem, HA412-3_Xylem) were suspected of being of poor quality from NanoDrop analysis, and later was confirmed through

BioAnalyzer data with RIN scores <5.0. These samples were dissected again from fresh cross sections and repeated for RNA extraction.

Prior to cDNA synthesis, equal amounts of RNA from individual plants was pooled by tissue, age and species. cDNA was synthesized using SuperScriptIII First-Strand Synthesis for RT-PCR (Invitrogen Cat. No. 18080-51, Life Technologies) following instructions provided by the manufacturer.

RT-qPCR

Oligonucleotide primers for quantitative PCR (qPCR) analysis of candidate gene expression were designed using Primer3 (<http://primer3.sourceforge.net/>). Real-time qPCR (RT-qPCR) was conducted using reactions based on SYBR Green Super Mix (Bio-Rad, Hercules, CA) and following changes in absorbance using an iQ5 iCycler (Bio-Rad) according to manufacturer protocols (Provenzano et al. 2007, Zhou et al 2006). RT-qPCR measurements were performed triplicate for each sample and primer pair. Threshold cycle (Ct) values were determined by the iCycler software. The expression of several housekeeping genes was assessed in each sample, and one of these (glyceraldehyde 3-phosphate dehydrogenase, GAPDH) was used to normalize expression values for all genes analyzed (Fernandez et al. 2008, Roche et al. 2009).

Results

Phenotype Observations

During the first eight weeks of growth, *H. argophyllus* and *H. annuus* differ little with respect to height, size and general appearance (Figure 8). However, around eight

weeks sylleptic branches begin to elongate and stems become noticeably more rigid on *H. argophyllus* plants.

Anatomical Analysis

Cross-sections of stems from 8-week old *H. argophyllus* and *H. annus* showed a number of different characteristics at the cellular level that varied with a developmental gradient from top to bottom of the plant. Starting with the uppermost (1st) internode, *H. argophyllus* and *H. annus* were nearly identical with respect to xylem and phloem tissue development as well as pith volume (Figure 9a-b.). Already by the 3rd internode, *H. argophyllus* has developed a continuous, secondary xylem and a thicker epidermis than *H. annus* (Figure 10a-b.). These and other differences are even more magnified in the basal internodes of each species. The cellular structure in basal internodes of *H. argophyllus* (Fig. 11a) takes on a morphology resembling that common in stems of woody angiosperms. In comparison to *H. annus*, the *H. argophyllus* xylem fiber cells appear smaller with thicker walls, vessels are smaller in diameter, pith volume is greatly reduced, and the epidermis is thicker and contains more polyphenolic materials. In contrast, in the basal internodes of *H. annus* at this age, xylem just begins to reach the circumferential continuity required to be classified secondary xylem and pith parenchyma cells still occupy a substantial volume of the stem (Fig. 11b).

RT-qPCR

17 candidate genes and 4 housekeeping genes were selected based on microarray data (Chapter 2) and a review of the literature in which differentially expressed genes

between xylem and phloem tissues were identified (Fernandez et al. 2008, Hewezi et al. 2006, Roche et al. 2009). Candidate genes were selected for their roles in cellulose, hemicelluloses, and lignin biosynthesis in other wood species, such as *Populus* and *Pinus*.

ACT1, ACT2, PAL1, and TUB2 were selected as housekeeping genes to demonstrate their relatively consistent level of expression between tissue types (xylem and phloem) and species (Arg1820 and HA412-HO). While ACT2, PAL1, and TUB2 were expressed nearly at identical levels across both species and tissue types, ACT1 varied between the species, but was consistent within a particular species but in different tissue types (Figure 14). Y-axis was increased to 10, from the original 1, to show expression levels still varied, as expected, and not completely identical.

RT-qPCR data shows that a majority of the candidate genes was consistent with that of the microarray. Focusing on the xylem tissue comparison endo-beta glucanase (0.08), TED3 (0.18), cellulose synthase (0.24), zinc finger (0.25), della protein (0.3), and cobra-like protein (0.4) all had lower expression levels in Arg1820 than HA412-HO (Figure 12a). There were however a few genes (glucosyl transferase, xyloglucan, somatic embryogenesis receptor-like kinase, MADS box, endo-beta glucanase) of the original 17 candidate genes that were unable to acquire RT-qPCR data. These genes are shown as zeroes (Figure 12 and 13).

As expected, the RT-qPCR data of a select few genes were not consistent with the microarray analysis. Sepallita 1-like mads box (0.077) and glycosyl transferase (0.14) both showed higher expression levels in Arg1820 xylem tissue than its HA412-HO counterpart. Glycosyl transferase shows greater than 1000 fold change in RT-qPCR data

between Arg1820 xylem and HA412-HO xylem (Figure 12a.) with a much greater expression in Arg1820 xylem.

In addition to differential expression between species, there were also significant expression differences between cell types within a species. One in particular, the cellulose synthase gene shows an approximately 100 fold change difference in RT-qPCR data between Arg1820 xylem tissue and HA412-HO xylem tissue (Figure 12a.) but a much smaller fold change in the comparison between Arg1820 phloem and HA412-HO phloem tissue (Figure 13b.). These expression levels are congruent to those found in microarray analysis of Arg1820 and HA412-HO tissue samples. However, when comparing expression levels of cellulose synthase in specific tissue types within species, HA412-HO xylem and HA412-HO phloem displays a marginal fold change (Figure 13a.) in comparison to 150 fold change viewed in Arg1820 xylem and Arg1820 phloem cells (Figure 13b.). A significant increase in expression levels of cellulose synthase is seen in the Arg1820 phloem, counter intuitive at first glance in regards to CesA's role in wood formation.

Discussion

The common sunflower has so far only been grown as a source of oilseed, which can be used for biodiesel production, but the remainder of the plant has not been considered as a source of lignocellulosic biomass for biofuel production (Beckman et al. 2008). This is primarily due to the low-density, pithy stems that characterize common sunflowers. These low-density stems contain relatively little carbon that could be converted to biofuel. However, in desert environments, there are species of sunflowers

that do produce high-density woody stems, including the silverleaf sunflower (*Helianthus argophyllus*) and the Algodones dune sunflower (*Helianthus niveus*).

Lignin, the dominant defining component of lignocellulosic secondary cell walls, is the key to development of plant vascular systems and the great stature of many woody plants (Rogers et al. 2005). Lignin serves as a thermoplastic matrix that crosslinks components of the secondary cell wall to strengthen the mechanical characteristics of plant stems. In addition to structural support, the hydrophobic properties of lignin, in contrast to the hydrophilic properties of the secondary cell wall polysaccharides (cellulose and hemicellulose), allow it to play a critical role in conducting water throughout the plant (Chabannes et al. 2001). This aspect of lignin has obvious ramifications for the increased production of lignified xylem in the silverleaf and Algodones dune sunflowers since plant water-use efficiency (WUE) affects plant fitness and acts as a natural selector in the desert environments where these plants grow (Donovan et al. 2006). The desert floor habitats occupied by the silverleaf and Algodones dune sunflowers are extreme environments that impose much stronger water stress on these plants than is experienced by the domesticated species, common sunflower. Adaptations to the harsh desert environments have likely driven the increased production of lignocellulosic vascular tissues in the desert species (Gross et al, 2004).

Glycosyl transferase has been identified as one of the genes responsible in the development of secondary cell walls. It plays a role in the biosynthesis of glucuronoxylan, a major component in secondary cell walls in addition to lignin and cellulose (Keppler et al. 2010). Glycosyl transferase is required for the initiation, elongation, and termination of the xylan backbone of glucuronoxylan (Pena et al. 2007).

Glycosyl transferase has been shown in poplar (*Populus*) to be associated with secondary wall synthesis, in *Arabidopsis thaliana* to be required for proper vessel morphological development and cell wall thickness to allow for the appropriate cellulose deposition in cell walls (Persson et al. 2005).

Being that Arg1820 is the woody species, a higher expression of glycosyl transferase is expected in the xylem. IRX14 and IRX14L are two closely related glycosyl transferases that have been identified in *Arabidopsis thaliana* and their role in secondary cell wall development are shown through a genetic knockout experiment resulting in an irregular xylem phenotype and overall reduction in xylose content (Keppler et al. 2010). In a similar genetic knockout experiment, IRX8 and IRX9, two other glycosyl transferases in *Arabidopsis thaliana*, were shown to exhibit a collapsed xylem phenotype in mutations. In addition to the irregular xylem development and abnormal vessel morphology, tests for cellulose content also showed a significant decrease (Pena et al. 2007).

The cellulose synthase (CesA) gene family is known to play a regulatory role in cell wall biosynthesis pathway. A total of six CesA proteins have been shown to be required for cell wall biosynthesis in vascular tissues of *Arabidopsis thaliana*, with three associated with primary cell wall development (CesA1, CesA3, CesA6) and three associated with secondary cell wall development (CesA4, CesA7, CesA8) (Dai et al. 2011, Gardner et al. 2003, Tanaka et al. 2003). In primary cell walls, cellulose is important in maintaining cell shape and cell expansion, which is essential during plant growth. Cellulose also is essential in secondary cell walls, where it has its role in mechanical strength for the plant (Wightman et al. 2010).

Further investigation shows that the CesaA in question is CesaA3, a known cellulose synthase in primary cell wall development (Gardner et al. 2003). Primary and secondary cell wall development by the CesaA gene family is differentially expressed in xylem and phloem cells of woody plants, as both are unique in their roles of development. In situ hybridization of *Populus tremuloides* shows the detection of *PtrCesaA3* in xylem and other tissue types in secondary cell wall development stage, while *PtrCesaA4* expression was isolated in the vascular cambium region (Kalluri et al. 2004). This coincides with the RT-qPCR data observed as *PtrCesaA4* is highly similar to CesaA1 of *Arabidopsis thaliana*, which is involved in primary cell wall synthesis (Arioli et al. 1998).

Conclusion

Laser microscopy dissection has provided the technology to isolate specific cell types in the study of gene expression. With the ability to investigate particular cells, further understanding of wood formation can be gathered, as the plant is composed of various specific cell types that each play their own mechanical roles in plant development. The focus on xylem and phloem gene expression is critical as the two cell types are similar in biochemical content, yet vastly different in composition and in result structurally and mechanically different. Drawing from literature and previous experiments, a short list of candidate genes were able to be investigated in these two tissue types of two greatly different sunflower species. But the short list only leaves more to be investigated and discovered in xylem and phloem development.

References

1. Abbott et al. 2010. Laser microdissection of conifer stem tissues: Isolation and analysis of high quality RNA, terpene synthase enzyme activity and terpenoid metabolites from resin ducts and cambial zone tissue of white spruce (*Picea glauca*). *BMC Plant Biology* 10:106
2. Arioli T, Peng L, Betzner AS, Burn J, Wittke W, Herth W, Camilleri C, Hofte H, Plazinski J, Birch R, Cork A, Glover J, Redmond J, Williamson RE. 1998. Molecular analysis of cellulose biosynthesis in *Arabidopsis*. *Science* 279(5351): 717-720
3. Carlsbecker A, Helariutta Y. 2005. Phloem and xylem specification: pieces of the puzzle emerge. *Current Opinion in Plant Biology* 8(5): 512-517
4. Dai X, You C, Chen G, Li X, Zhang Q, Wu C. 2011. OsBC1L4 encodes a COBRA-like protein that affects cellulose synthesis in rice. *Plant Molecular Biology* 75(4-5): 333-345
5. Delgado A, Lee JY, Demura T. 2010. Regulatory Mechanisms for Specification and Patterning of Plant Vascular Tissues. *Annual Review of Cell and Developmental Biology* 26: 605-637
6. Dyer, JH, Maina, A, Gomez, ID, Cadet, M, Oeljeklaus, S, Schidel, AC. 2009. Cloning, Expression and Purification of an Acetoacetyl CoA Tholase from Sunflower Cotyledon. *International Journal of Biological Sciences*. 5(7): 736-744
7. Felitti, SA, Chan, RL, Gago, G, Valle, EM, Gonzalez, DH. 1997. Expression of sunflower cytochrom c mRNA is tissue-specific and controlled by nitrate and light. *Phsiologia Plantarum*. 99: 342-347

8. Fernandez, P, Doslo GAA. 2008. Validating housekeeping genes for quantitative plant gene expression studies in sunflower senescence process. 54^o Congresso Brasileiro de Genetica.
9. Gardiner JC, Taylor NG, Turner SR. 2003. Control of cellulose synthase complex localization in developing xylem. *Plant Cell* 15(8): 1740-1748
10. Harding SA, Leshkevich J, Chiang VL, Tsai CJ. 2002. Differential Substrate Inhibition Couples Kinetically Distinct 4-Coumarate:Coenzyme A Ligases with Spatially Distinct Metabolic Roles in Quaking Aspen. *Plant Physiology* 128: 428-438
11. Hewezi T, Leger M, El Kayal W, Gentzbittel L. 2006. Transcriptional profiling of sunflower plants growing under low temperatures reveals an extensive down-regulation of gene expression associated with chilling sensitivity. *Journal of Experimental Botany* 57(12): 3109-3122
12. Kalluri UC, Joshi CP. 2004. Differential expression patterns of two cellulose synthase genes are associated with primary and secondary cell wall development in aspen trees. *Planta* 220(1): 47-55
13. Keppler BD, Showalter AM. 2010. IRX14 and IRX14-LIKE, two glycosyl transferases involved in glucuronoxylan biosynthesis and drought tolerance in *Arabidopsis*. *Molecular Plant* 3(5): 834-841
14. Kerk NM, Ceserani T, Tausta SL, Sussex IM, Nelson TM. 2003. Laser capture microdissection of cells from plant tissues. *Plant Physiology* . 132:27 – 35 .

15. Pena MJ, Zhong R, Zhou GK, Richardson EA, O'Neil MA, Darvil AG, York WS, Ye ZH. 2007. Arabidopsis irregular xylem8 and irregular xylem9: implications for the complexity of glucuronoxylan biosynthesis. *Plant Cell* 19(2): 549-563
16. Persson, S, Wei, H, Milne, J, Page, G.P. 2005 Large-scale coexpression analysis reveals novel genes involved in cellulose biosynthesis. *Proc. National Academy of Science USA*. 102: 8633-8638
17. Provenzano M, Mocellin S. 2007. Complementary Techniques: Validation of Gene Expression Data by Quantitative Real Time PCR. *Microarray Technology and Cancer Gene Profiling*.
18. Roche J, Hewezi T, Bouniols A, Gentzbittle L. 2009. Real-time PCR monitoring of signal transduction related genes involved in water stress tolerance mechanism of sunflower. *Plant Physiology and Biochemistry* 47:139-145.
19. Sieburth L, Deyholos MK. 2006. Vascular development: the long and winding road. *Current Opinion in Plant Biology* 9(1): 48-54
20. Takahashi H, Kamakura H, Sato Y, Shiono K, Abiko T, Tsusumi N, Nagamura Y, Nishizawa N, Nakazono M. 2010. A method for obtaining high quality RNA from paraffin sections of plant tissues by laser microdissection. *J Plant Res*. DOI 10.1007/s10265-010-0319-4
21. Tanaka K, Murata K, Yamazaki M, Onosato K, Miyao A, Hirochika H. 2003. Three distinct rice cellulose synthase catalytic subunit genes required for cellulose synthesis in the secondary wall. *Plant Physiology* 133(1): 73-83

22. Urieli-Shoval S, Meek R, Hanson RH, Ferguson M, Gordon D, Benditt EP. 1992. Preservation of RNA for In Situ Hybridization: Carnoy's Versus Formaldehyde Fixation. *The Journal of Histochemistry and Cytochemistry* 40(12): 1879-1885
23. Wightman R, Turner S. 2010. Trafficking of the plant cellulose synthase complex. *Plant Physiology* 153(2): 427-432
24. Ye ZH, Freshour G, Hahn MG, Burk DH, Zhong R. 2002. Vascular development in Arabidopsis. *International Review of Cytology* 220: 225-256
25. Zhou G, Richardson R, Morrison E, Nairn J. 2006. The Poplar glycosyltransferase GT47C is functionally conserved with Arabidopsis Fragile Fiber8. *Plant Cell Physiology* 47:1229-1240



Figure 8. Phenotype comparison of *Helianthus annuus* (HA412-HO, right) and *Helianthus argophyllus* (Arg1820, left). Noticeable difference in the branching of Arg1820 in comparison to HA412-HO which only has foliage growth off its primary stem. Both plants are nearly identical in height with HA412-HO being 1.5m from soil line to top apical bud in 5-gallon pot and Arg1820 1.4m from soil line to top apical bud in 5-gallon pot at the University of Georgia Warnell College of Natural Resources Whitehall Greenhouse.

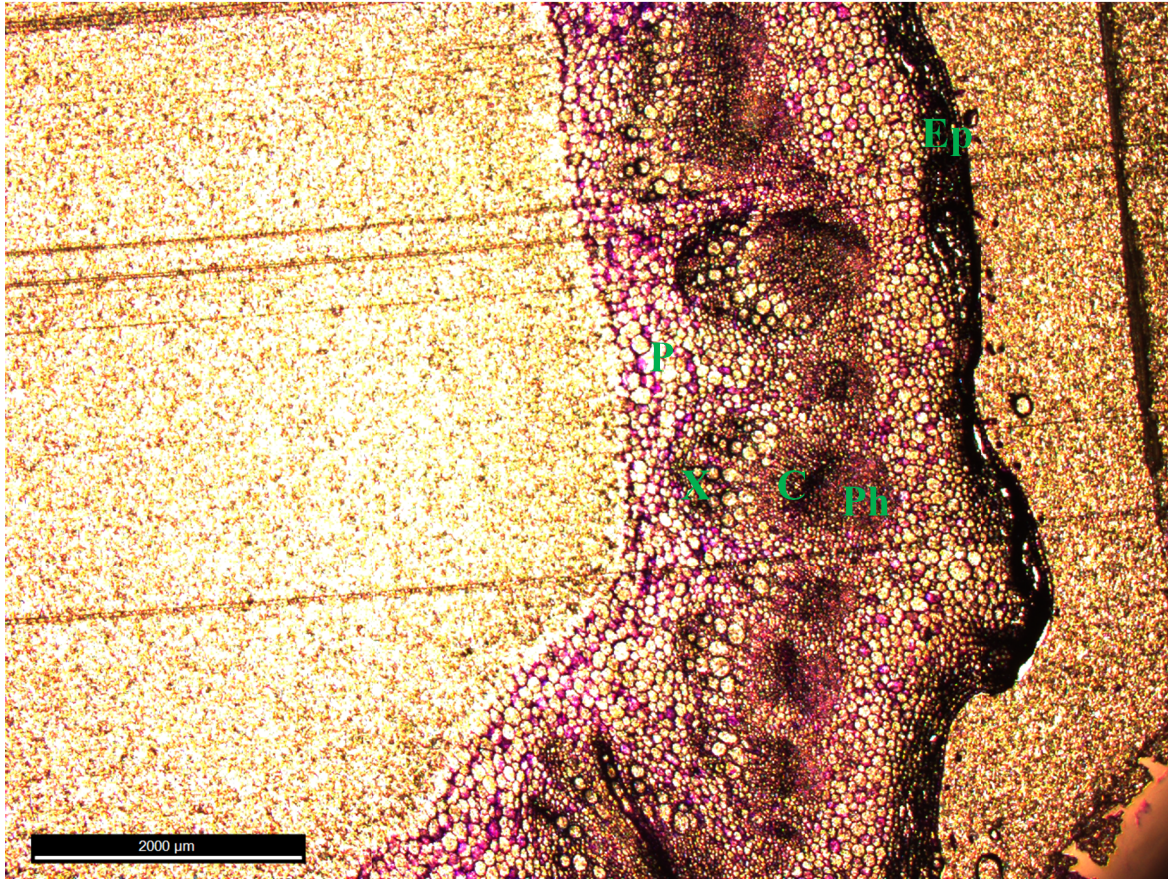


Figure 9a. 15µm cross-section of the 1st node at 8-week of Arg1820 stained with toluidine blue observed at 2.5x magnification. Secondary xylem development has not begun. Xylem fibers are still in the infant stages of development. Nearly identical in structure to HA412-HO. Pith (P), xylem (X), vascular cambium (C), phloem (Ph), and epidermis (Ep). Bulk of pith tissue lost during cross section. Area in the center of the stem section is paraffin embedding medium.

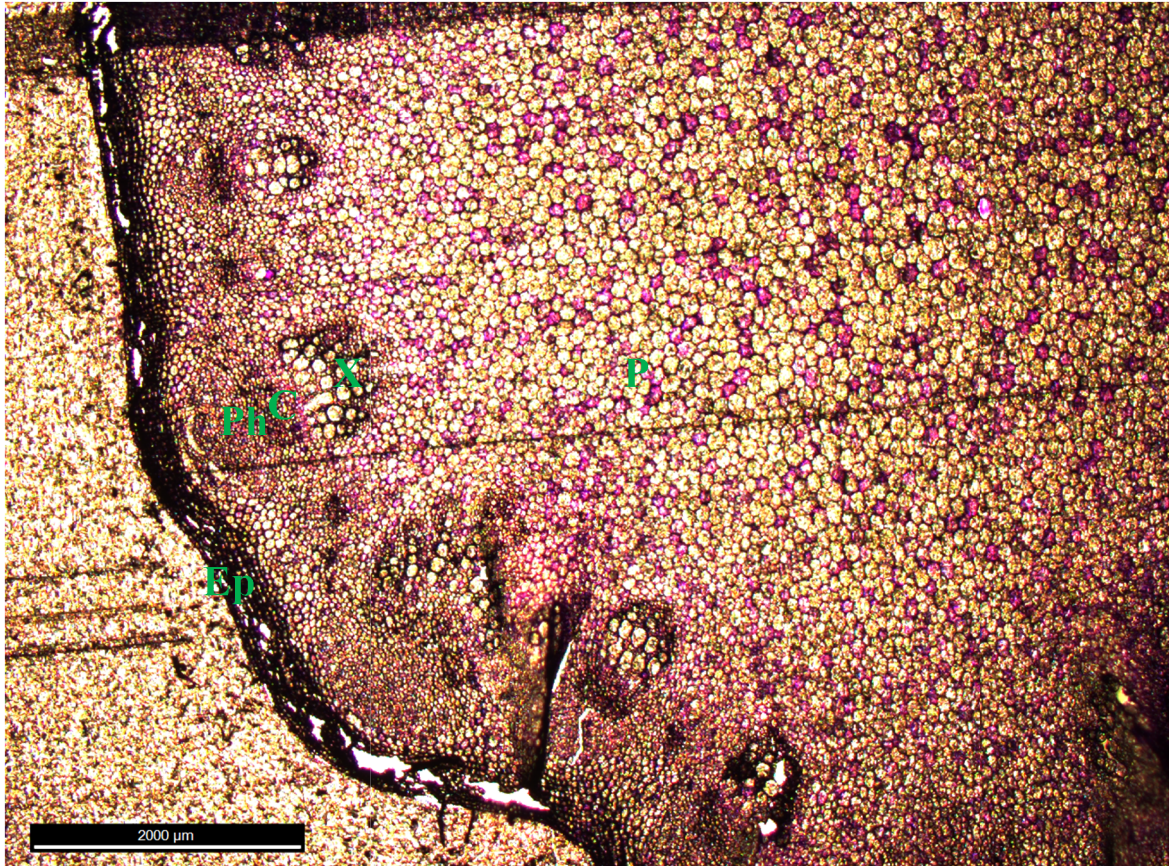


Figure 9b. 15μm cross-section of the 1st node at 8-week of HA412-HO stained with toluidine blue observed at 2.5x magnification. Secondary xylem development has not begun. Xylem fibers are still in the infant stages of development. Nearly identical in structure with Arg1820. Pith (P), xylem (X), vascular cambium (C), phloem (Ph), and epidermis (Ep).

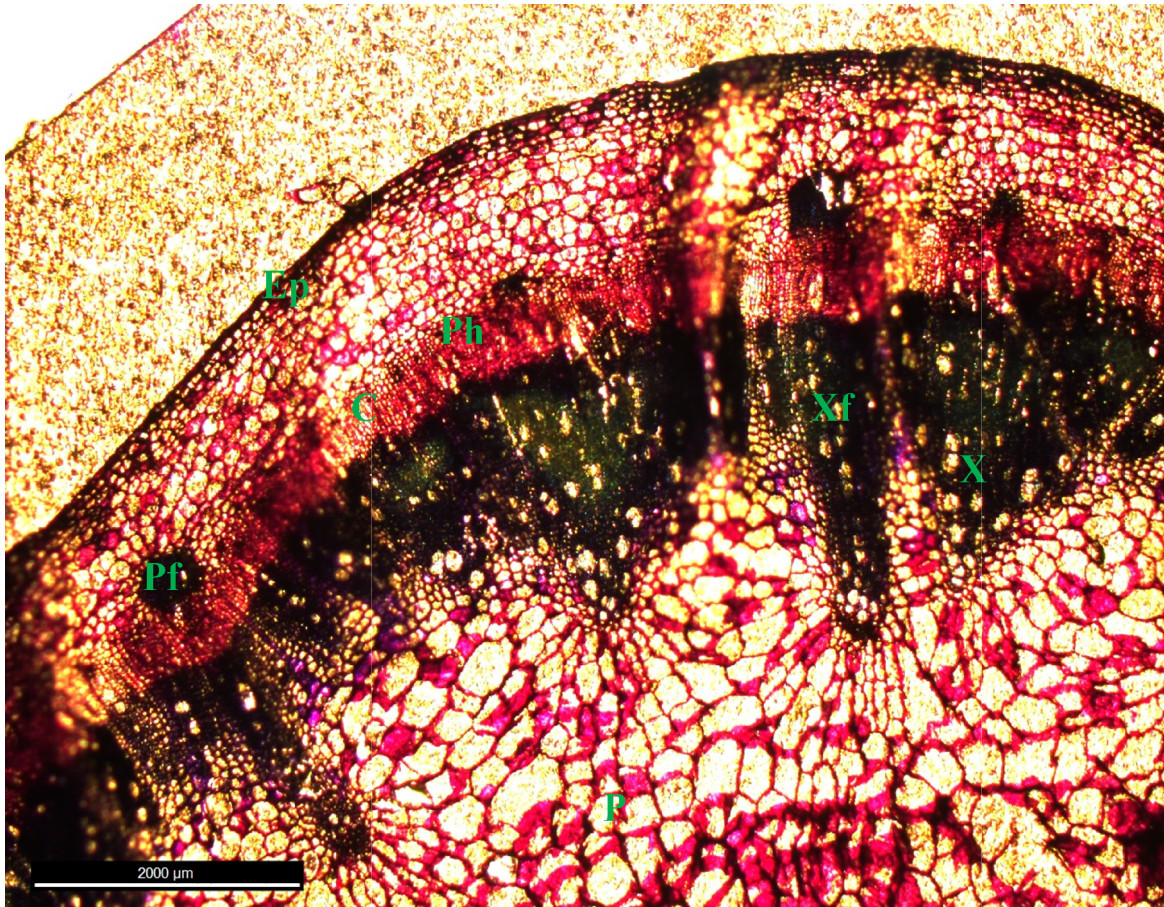


Figure 10a. 15µm cross-section of the 3rd node at 8-week of Arg1820 stained with toluidine blue observed at 2.5x magnification. Secondary xylem has developed and began to form a complete ring around the stem. Pith (P), xylem (X), xylem fibers (Xf) vascular cambium (C), phloem (Ph), phloem fibers (Pf), and epidermis (Ep).

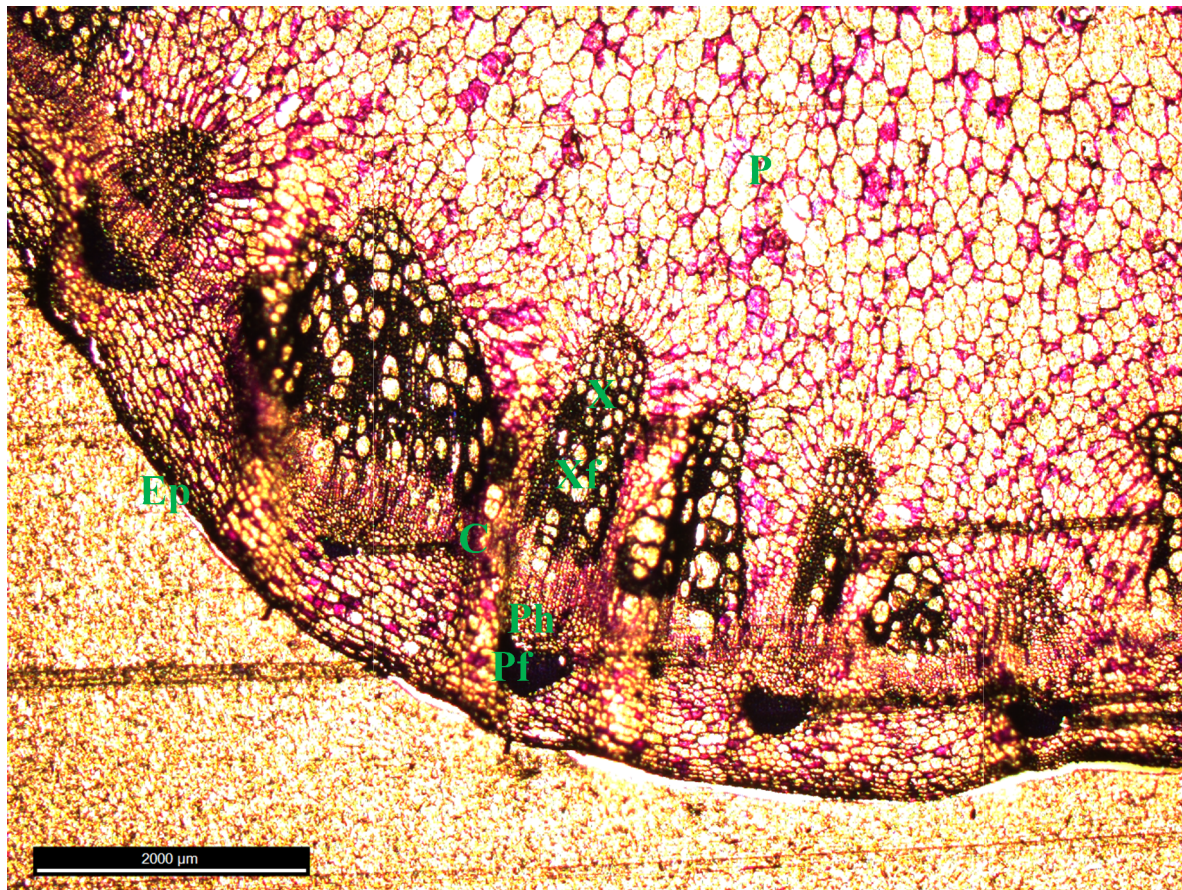


Figure 10b. 15µm cross-section of the 3rd node at 8-week of HA412-HO stained with toluidine blue observed at 2.5x magnification. Secondary xylem has developed at a much slower rate, and a complete ring has not formed as xylem bundles have not connected. Pith (P), xylem (X), xylem fibers (Xf), vascular cambium (C), phloem (Ph), phloem fibers (Pf), and epidermis (Ep).

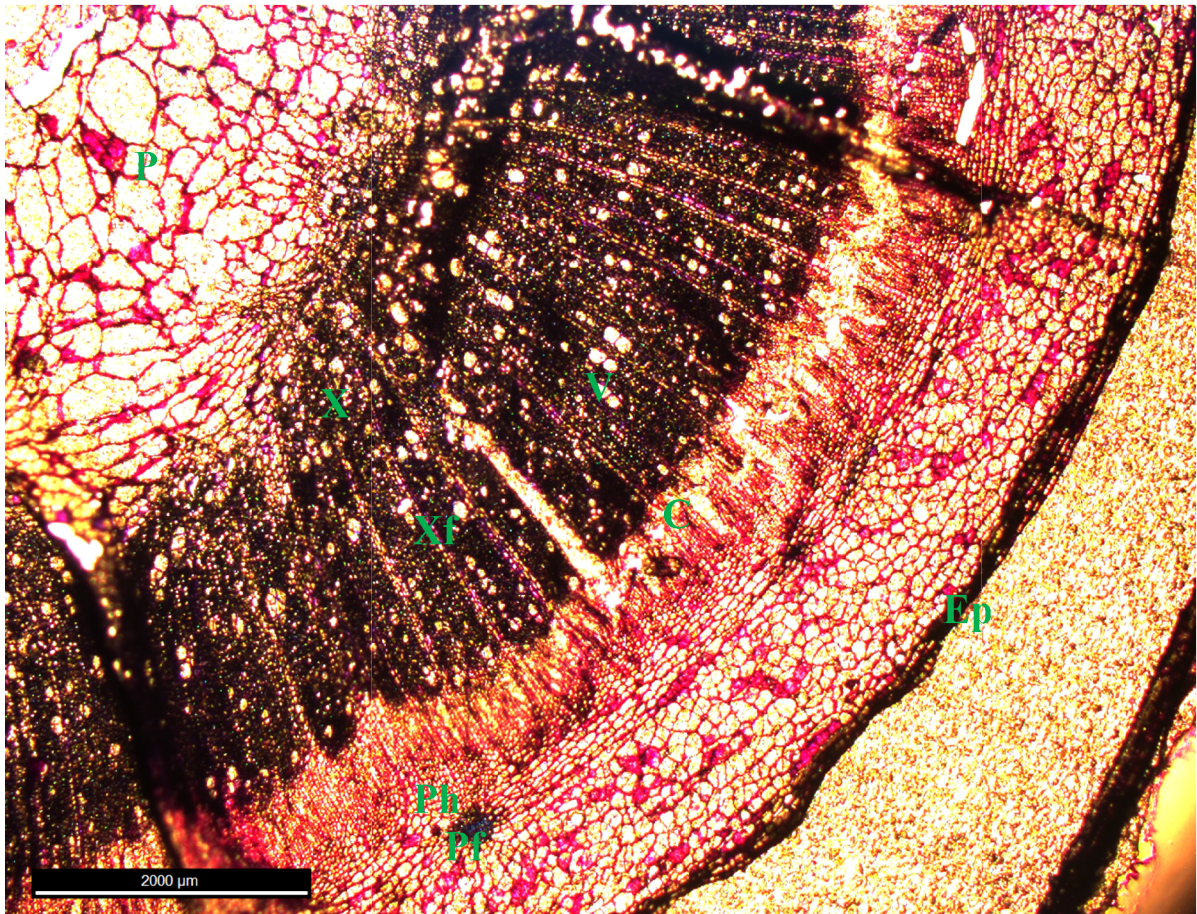


Figure 11a. 15µm cross-section of the base at 8-week of Arg1820 stained with toluidine blue observed at 2.5x magnification. Secondary xylem has completely formed around the circumference of the stem. Vessel elements are small and tightly compact amongst each other in the xylem. Pith (P), xylem (X), xylem fibers (Xf), vessels (V), vascular cambium (C), phloem (Ph), phloem fibers (Pf), and epidermis (Ep).



Figure 11b. 15µm cross-section of the base at 8-week of HA412-HO stained with toluidine blue observed at 2.5x magnification. Secondary xylem has formed however, much of the stems inner regions remains pithy with larger more spread out vessels throughout the xylem. Pith (P), xylem (X), xylem fibers (Xf), vessels (V), vascular cambium (C), phloem (Ph), phloem fibers (Pf), and epidermis (Ep).

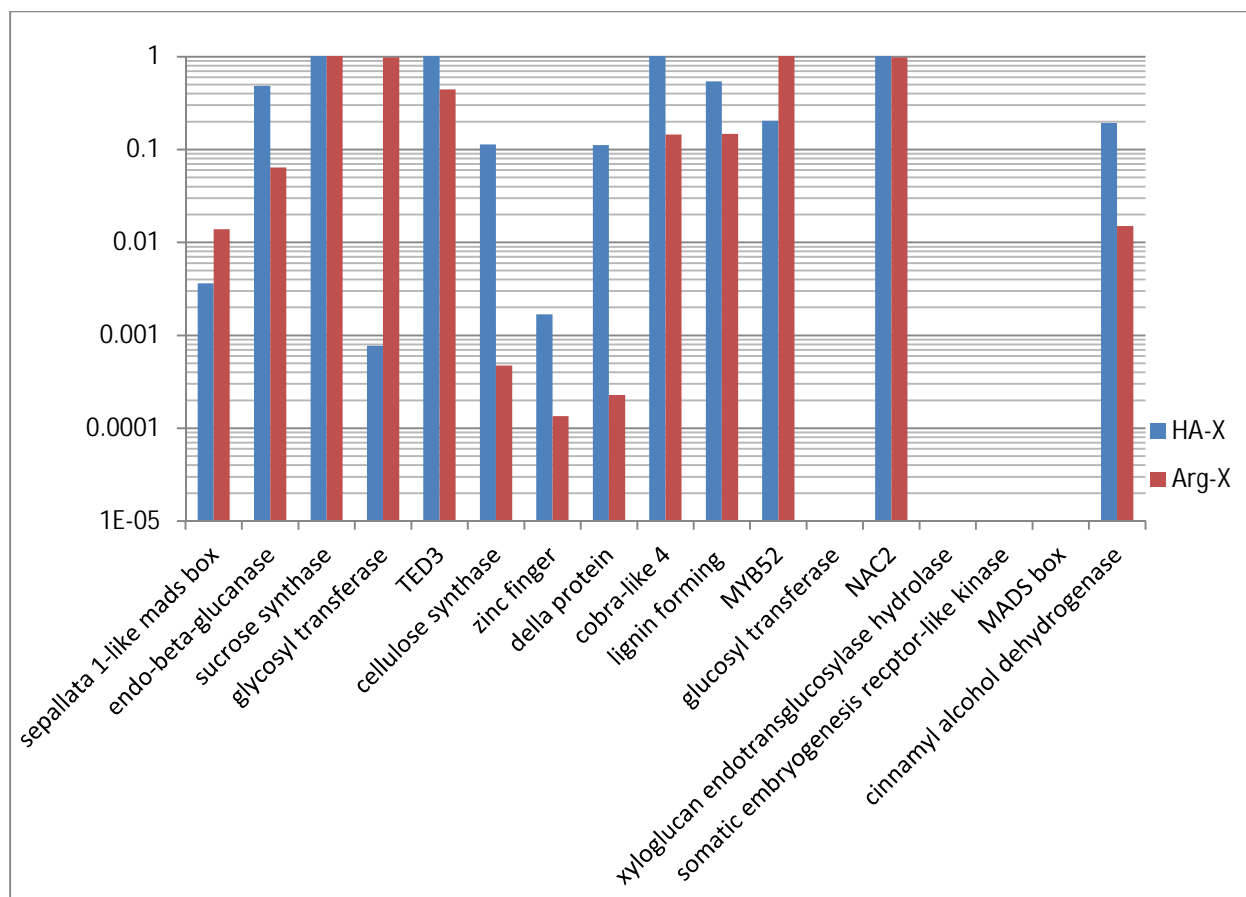


Figure 12a. RT-qPCR expression level comparison between HA412-Xylem and Arg1820-Xylem from the base at the 8-week time point.

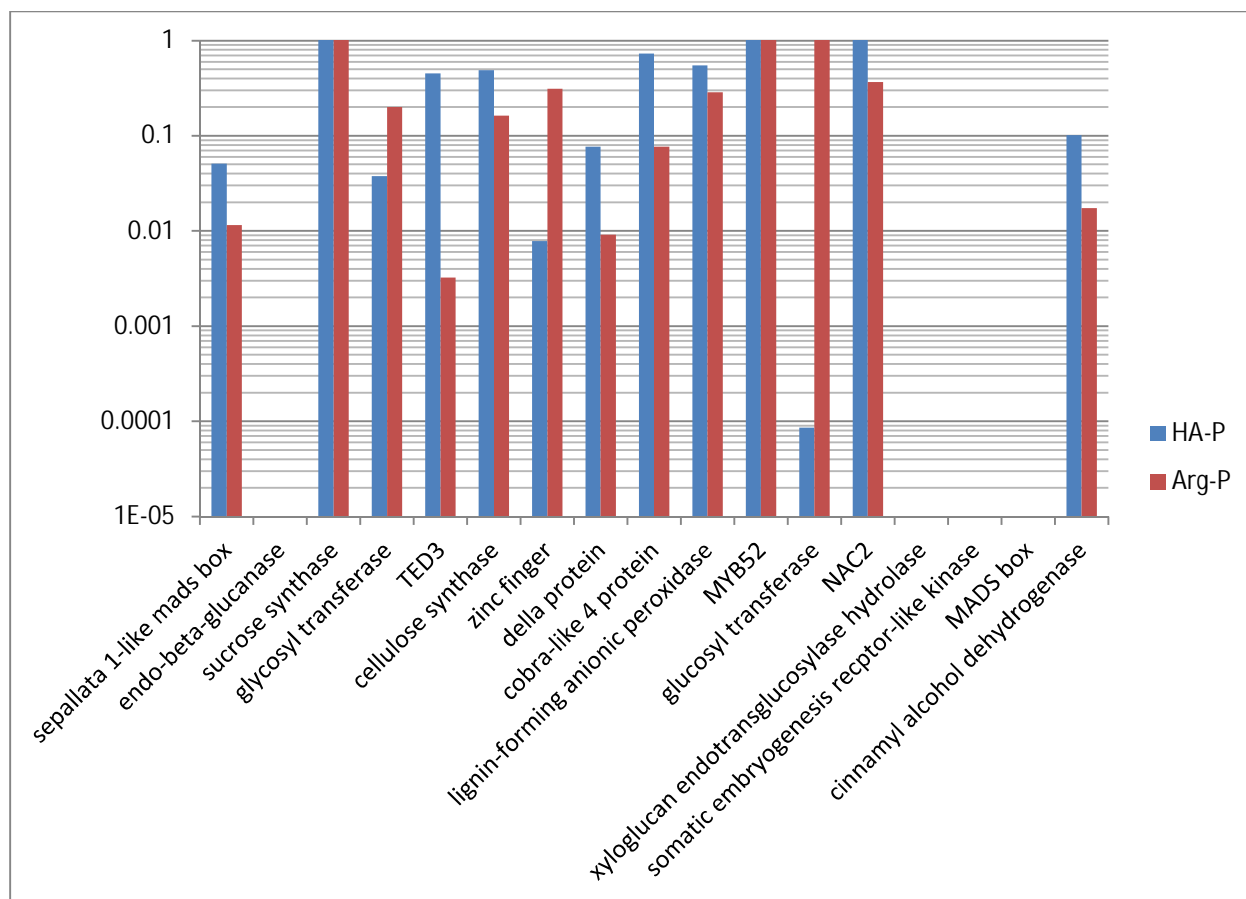


Figure 12b. RT-qPCR expression level comparison between HA412-Phloem and Arg1820-Phloem from the base at the 8-week time point.

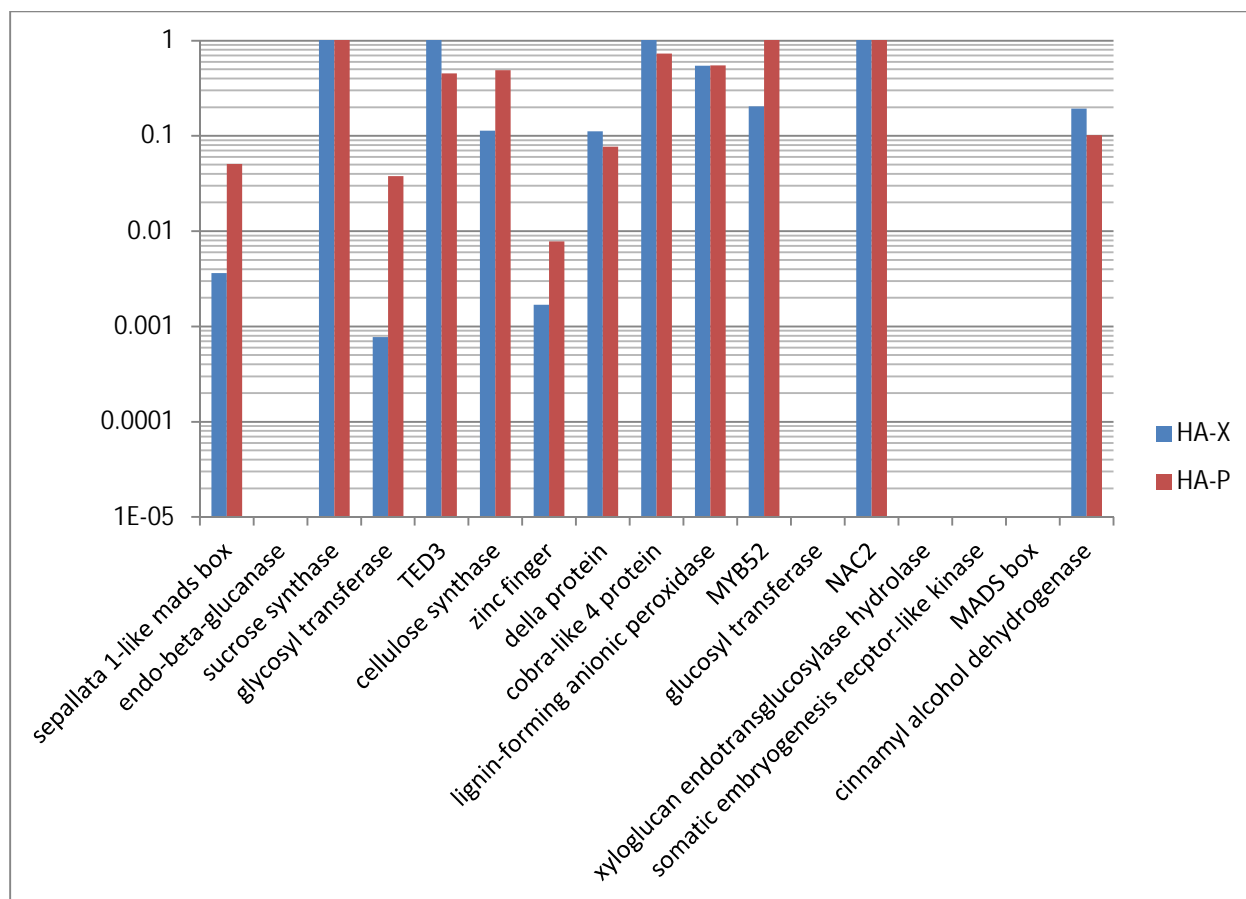


Figure 13a. RT-qPCR expression level comparison between HA412-Xylem and HA412-Phloem from the base at the 8-week time point.

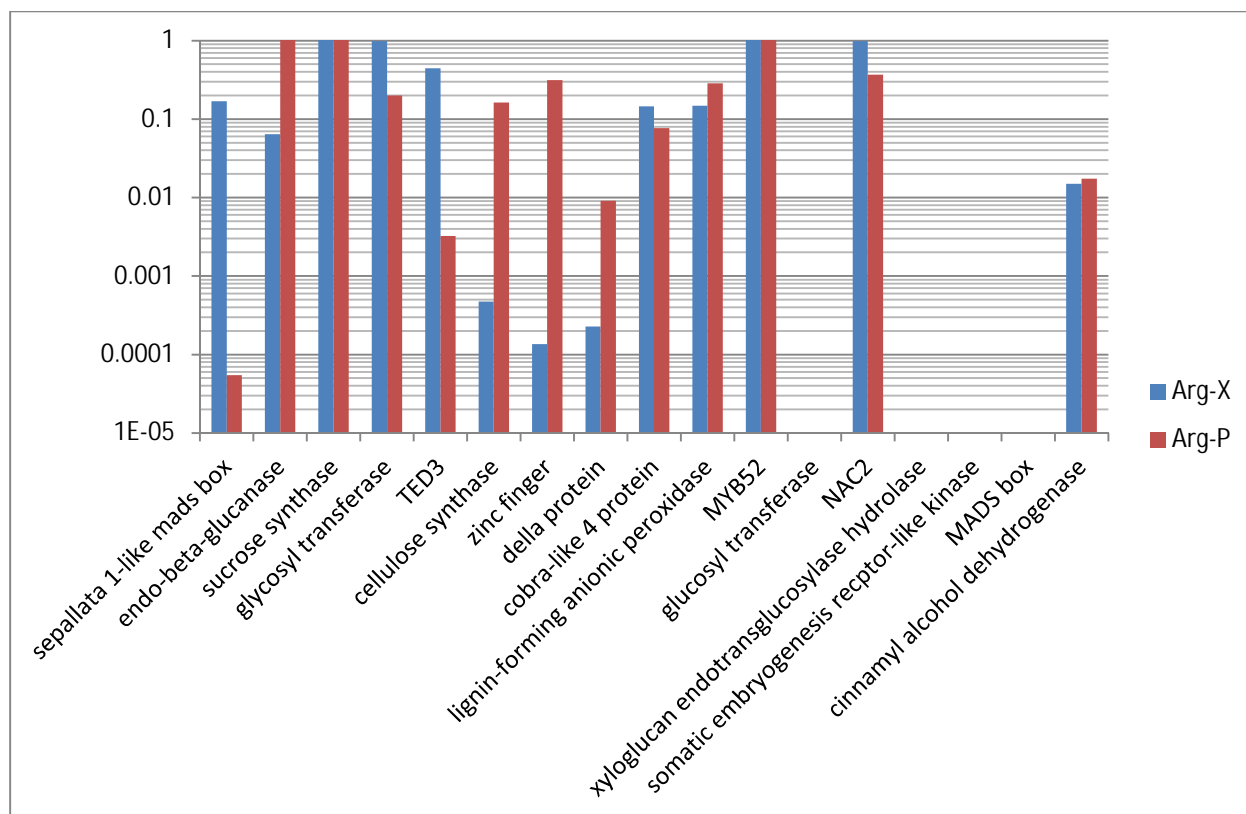


Figure 13b. RT-qPCR expression level comparison between Arg1820-Xylem and Arg1820-Phloem from the base at the 8-week time point.

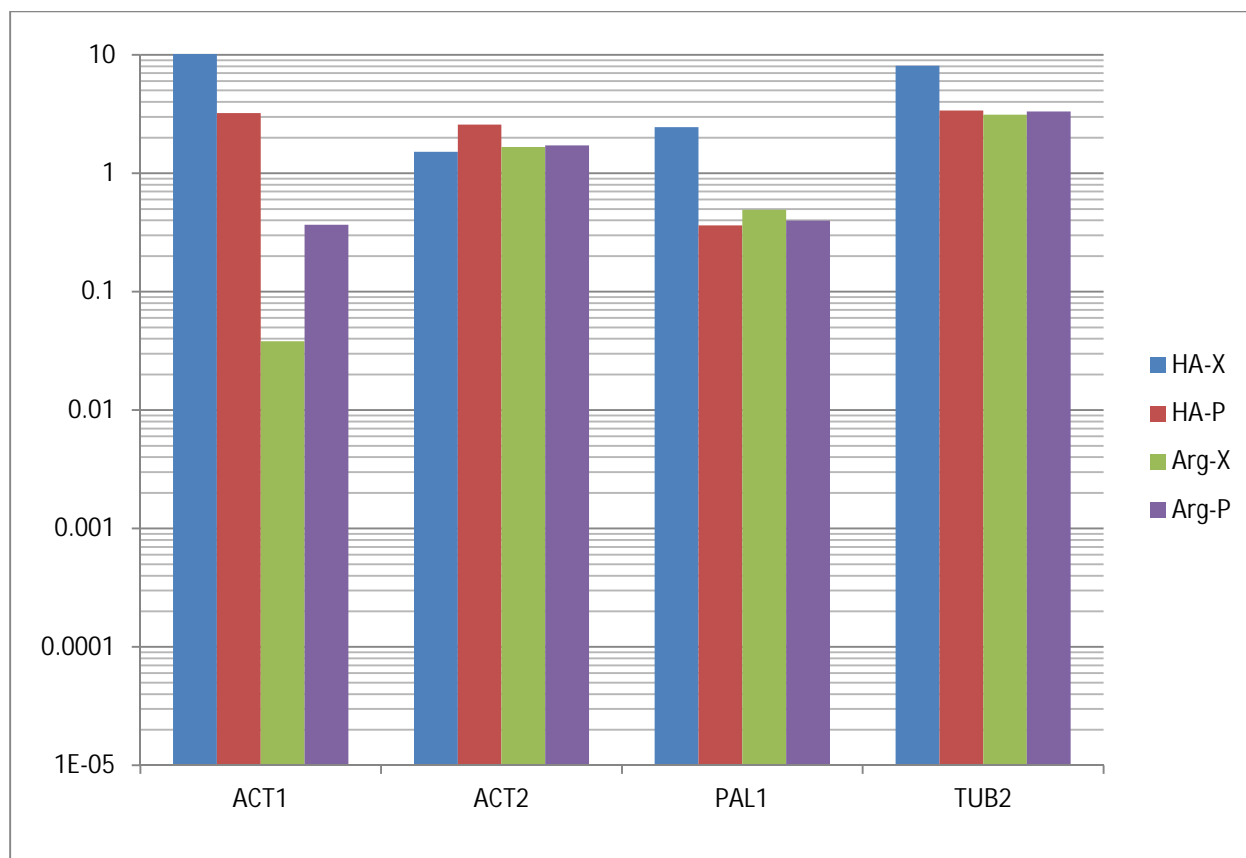


Figure 14. RT-qPCR expression level comparison between all tissue samples (HA412-Xylem, HA412-Phloem, Arg1820-Xylem, Arg1820-Phloem) of four housekeeping genes (ACT1, ACT2, PAL1, TUB2) to show generally consistent expression levels across all species and tissue types.

Sample ID	Concentration (ng/ul)	A260/A280	A260/A230
Arg1820-1_Xylem	170.22	1.87	1.3
Arg1820-1_Phloem	55.05	1.82	1.35
Arg1820-2_Xylem	172.92	1.84	1.37
Arg1820-2_Phloem	91.05	1.81	1.3
Arg1820-3_Xylem	183.15	1.83	1.41
Arg1820-3_Phloem	2.22	1.29	0.36
HA412-1_Xylem	84.61	1.82	1.45
HA412-1_Phloem	54.18	1.81	1.23
HA412-2_Xylem	79.4	1.83	1.44
HA412-2_Phloem	2.65	1.29	0.36
HA412-3_Xylem	1.41	1.40	0.62
HA412-3_Phloem	53.02	1.88	1.37
Arg1820-3_Phloem_R	61.27	1.81	1.39
HA412-2_Phloem_R	69.68	1.82	1.35
HA412-3_Xylem_R	83.46	1.87	1.41

Table 2. NanoDrop data of LCM samples. 3 samples were (Arg1820-3_Phloem, HA412-2_Phloem, and HA412-3_Xylem) were reran and isolated again for RNA. 260/280 ratios range from 1.81-1.88 and 260/230 ratios range from 1.3-1.45. Concentrations range from 53.02ng/ul – 183.15ng/ul. All samples were resuspended in 20ul of DEPC-mpH₂O.

Fold Change	Gene Name
0.077	sepallata 1-like mads box
0.08	endo-beta-glucanase
0.081	sucrose synthase
0.14	glycosyl transferase
0.18	TED3
0.24	cellulose synthase
0.25	zinc finger
0.3	della protein
0.4	cobra-like 4 protein
4.81	lignin-forming anionic peroxidase
5.85	MYB52
6.73	glucosyl transferase
7.32	NAC2
10.07	xyloglucan endotransglucosylase hydrolase
12.38	somatic embryogenesis receptor-like kinase
14.57	MADS box
51.68	cinnamyl alcohol dehydrogenase
housekeeping	ACT1
housekeeping	ACT2
housekeeping	PAL1
housekeeping	TUB2
control	GAPDH

Table 3. 17 genes of interest and their respective fold change values selected from microarray data. GAPDH used to as the control gene normalization of expression levels and 4 housekeeping genes (ACT1, ACT2, PAL1, and TUB2) used to show steady expression levels in like cell types.

CHAPTER 4: SUMMARY OF WORK

Conclusion From Present Work

In the search for a proper source of biomass, *Helianthus argophyllus* has shown the woody qualities needed in the production of second-generation biofuels. In contrast to the common sunflower, *Helianthus annuus*, the development of a complex matrix of secondary cell walls provides *argophyllus* with a physical structure similar to other well studied wood plants such as poplar (*Populus*). Composed of cellulose, hemicelluloses, and lignin, the biological pathway of lignocellulosic biomass, that is the source for energy in biofuel production, has been shown through the gene expression profiling technique microarray and even more specified study of xylem and phloem cells which key regulatory gene families play critical roles in its production.

The cellulose synthase (CesA) family known for its regulatory genes in the production of cellulose for primary and secondary plant cell walls, specifically CesA3 a primary cell wall regulator, is shown to have higher expression rates in HA412-HO in comparison to Arg1820. The COBRA-like protein, specifically COBRA-LIKE4 protein (CBL4), is shown to have a higher expression rate in Arg1820 compared to HA412-HO. In addition to establishing this genes role in wood formation, the limitations of microarray studies was also shown when RT-qPCR analysis displayed revealed exponential folds change difference, albeit in the same direction. The glycosyl transferase gene family, a regulator of the synthesis of xylan, a major component in secondary cell walls was also shown to be highly expressed in Arg1820 in comparison to HA412-HO.

With its critical role in vessel morphology and mechanical structure, expression of glycosyl transferase is congruent with that of the woody species Arg1820.

Further Studies

With the combination of laser microscopy dissection and microarray (LCM) gene expression profiling, the pitfalls of studying plant tissue on a broad scale are exposed in both the dynamic range of microarray data analysis and generality of tissue samples. The plant body is composed of various specific cell types with each having a specific role in the development of the plant. Utilizing the laser microscopy dissection technique has allowed for a closer look at gene expression, but in this study only for a few selected candidate genes. To further expand, xylem and phloem specific tissue samples extracted from LCM can itself be studied in a large scale gene expression profile. Either with the aforementioned microarray technology or to even take a closer look with deep sequencing, such as RNA-seq. Also, additional tissue types, such as the inclusion of pith cells, could provide a better understanding of particular gene expression levels.