University of Alberta

Genetic Analysis of Lignification and Secondary Wall Development in Bast Fibers of Industrial Hemp (*Cannabis sativa*)

by

Susan P. Koziel

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Master of Science
in
Plant Biology

Department of Biological Sciences

Spring 2010
Edmonton, Alberta

Examining Committee

Dr. Michael Deyholos, Biological Sciences

Dr. Warren J. Gallin, Biological Sciences

Dr. Enrico Scarpella, Biological Sciences

Dr. Stephen Strelkov, Faculty of Agricultural, Food and Nutritional Science

Abstract

Industrial hemp (*Cannabis sativa*) is a highly productive crop that is well suited to cultivation in Canada. To better understand the development of bast (phloem) fiber secondary walls and to facilitate reverse genetics screening for improved germplasm, I undertook two sets of microarray experiments. The first compared transcript expression in stem peels at three positions along the length of the stem. The second set of microarray experiments compared transcript expression in adjacent tissue layers along the radial axis of the stem. The transcripts that were enriched in fiber-producing tissues in both studies were consistent with a dynamic program of cell wall deposition. Detailed qRT-PCR analysis of specific lignification genes identified the best targets for reverse genetics. Finally, as a first step towards establishing a virally induced gene silencing (VIGS) system, I identified viruses that produced visual symptoms of infection, although qRT-PCR failed to confirm the infection.

Acknowledgements

I would like to thank The Alberta Research Council for allowing me the time and funding this research, especially Dr. John Vidmar for his overall support of the project; Dr. Jan Slaski, for making the ARC germplasm collection available to me and providing information about the varieties; and Marie Gorda, for lending me some of her excess plants and suggestions on growing healthy hemp plants. I would also like to thank the numerous folks at the University of Alberta who allowed me to become a part-time M.Sc. student; Mary de Pauw and Melissa Roach for their advice and technical support while I was at the university; Anthony Cornish and Troy Locke from MBSU who provided technical assistance with the microarray and qRT-PCR equipment; and Ryan Edgar who spent hours helping me section stems and prepare the microscopy slides.

Table of Contents

Index of Tables

Index of Figures

Table of Abbreviations

Abbreviation	Definition
4CL	4-coumarate CoA Ligase
4CL1	4-coumarate CoA Ligase alignment one
4cl1	4-coumarate CoA Ligase primer set one
4CL2	4-coumarate CoA Ligase alignment two
4cl5	4-coumarate CoA Ligase primer set five
A linkage	Monolignol beta-O-4 linkage
AGO1	Argonaute protein
AGP	Arabinogalactan
AGPs	Arabinogalactan proteins
AMV	Alfalfa mosaic virus
AMV-0196	Alfalfa mosaic virus PV-0196
ARC	Alberta Research Council
ArMV	Arabis mosaic virus
ArMV-0045	Arabis mosaic virus PV-0045
ArMV-0232	Arabis mosaic virus PV-0232
At[followed by a gene name]	Gene in *Arabidosis thaliana*
At[followed by a number]	*Arabidopsis* TAIR accession number
ATKT2P	*Arabidopsis thaliana* potassium transporter 2 protein
ATP	Adenosine triphosphate
AUX/IAA protein	Auxin/Indoleacetic acid induced transcription factor
B linkage	Monolignol beta-5 linkage
BGAL	Beta galactosidase
bgal1	Beta galactosidase primer set one
BLASTn	Basic Local Alignment Search Tool - nucleotide
BMV	Brome mosaic virus
C linkage	Monolignol beta-beta linkage
C3H	p-Coumarate 3-hydroxylase
C4H	Cinnamate 4-hydroxylase
c4h1	Cinnamate 4-hydroxylase primer set one
CAD	Cinnamyl alcohol dehydrogenase
CBL4	Cobra-like 4

Abbreviation	Definition
CCoAOMT	Caffeoyl-CoA O-methyltransferase
CCR	Cinnamoyl-CoA reductase
CCR1	Cinnamoyl-CoA reductase alignment one
ccr1	Cinnamoyl-CoA reductase primer set one
CCR2	Cinnamoyl-CoA reductase alignment two
ccr2	Cinnamoyl-CoA reductase primer set two
cDNA	Complementary DNA
CESA	Cellulose synthase
CesA	Cellulose synthase A
CesA - family	Cellulose synthase gene family
CESA1	Cellulose synthase alignment one
cesa1	Cellulose synthase primer set one
CESA2	Cellulose synthase alignment two
cesa2	Cellulose synthase primer set two
CESA3	Cellulose synthase alignment three
CESA4	Cellulose synthase alignment four
cesa4	Cellulose synthase primer set four
CG	Coniferin-beta-glucodisase
CLUSTAL W	Sequence alignment program
CMV	Cucumber mosaic virus
CMV-0037	Cucumber mosaic virus PV-0037
CMV-0453	Cucumber mosaic virus PV-0453
CND41	41kD Chloroplast nucleoid DNA-binding protein
COMT	Caffeate O-methyltransferase
COMT1	Caffeate O-methyltransferase alignment one
COMT2	Caffeate O-methyltransferase alignment two
comt2	Caffeate O-methyltransferase primer set two
COMT3	Caffeate O-methyltransferase alignment three
COMT4	Caffeate O-methyltransferase alignment four
Cy3	Cyanine 3 fluorescent dye label
Cy5	Cyanine 5 fluorescent dye label
CYP81E8	Cytochrome P450 family protein
D linkage	Monolignol 5-5 linkage
DHN1	Dehydrin 1
DICER	Double-stranded RNA specific endoribonuclease
DMT1	Ferrous ion membrane transport protein
DNA	Deoxyribonucleic acid

Abbreviation	Definition
DOXX	Unknown protein family
DSMZ	German Collection of Microorganisms and Cell Cultures
dsRNA	Double stranded RNA
DUF	Domain of Unknown Function
E linkage	Monolignol 5-O-4 linkage
EIN3-binding F-Box protein	Ethylene insensitive 3 transcription factor binding F-box protein
ELISA	Enzyme-linked immunosorbent assay
ESP	Epidermis from internode containing snap point
EST	Expressed sequence tag
F linkage	Monolignol beta-1 linkage
F1 ATP synthase	ATP synthase, H+ transporting mitochondrial F1 complex
F5H	Ferulate 5-hydroxylase
f5h2	Ferulate 5-hydroxylase primer set two
flagellar L-ring	Flagellar lipopolysaccaride ring
FOM	Figure of merit analysis
G	Guaiacyl monolignol subunit
GID1	Gibberellin insensitive dwarf 1 protein
H	p-Hydroxyphenyl monolignol subunit
HCT	Reversible acyltransferase with p-hydroxycinnamoyl-CoA:D quinatetransferase and p-hydroxycinnamoyl-CoA:shikimate p-hydroxycinnamoyltransferase activity
ICTVdB	The universal database of the international committee on taxonomy of viruses
ITIS	Integrated Taxonomic Information System
JIM14	Labelled antibodies to specific Arabibogalactan proteins
JIM5	Labelled antibodies to specific homogalacturons
JIM7	Labelled antibodies to specific homogalacturons
KMC	K-mean cluster analysis
KNN neighbours	K-nearest neighbour algorithm
KPO4	Potassium phosphate
LIM1	LIM-homeodomain transcription factor
LM2	Labelled antibodies to specific Arabibogalactan proteins
LM7	Labelled antibodies to specific xylogalacturon
LTPs	Lipid transfer proteins
M1G	Coniferyl alcohol
M1H	p-Coumaryl alcohol

Abbreviation	Definition
M1S	Sinapyl alcohol
MIPC	Membrane intrinsic protein C - aquaporin
mRNA	Messenger RNA
N	Reaction did not occur
NaPO4	Sodium phosphate
NCBI	National Center for Biotechnology Information
OMT	o-Methyltransferase
OMT2	o-methyltransferase 2
OSB	Oriented strand board
OSTIP2.1	*Oryza sativa* tonoplast intrinsic protein 2.1 - aquaporin
P8	Cell-wall P8 protein
PAL	Phenylalanine ammonia-lyase
PAR1	Phenylacetaldehyde reductase
PAR-1B	Photoassimilate-responsive protein 1B
PAS	Phloem from internode above snap point
PB1	Octicosapeptide/Phox/Bem1p domain containing protein
PBS	Phloem, including bast fiber, from internode below snap point
PCA	Principle components analysis
PCR	Polymerase chain reaction
PCS3H	p-Coumaroyl shikimate 3'-hydroxylase
PEG	Polyethylene glycol
PEX-S	Analysis comparing phloem, including bast fiber, to the epidermis and xylem at the snap point; aka ESP-PSP-XSP comparison
PO4 Buffer	Phosphate buffer
PopMV	Poplar mosaic virus
PSP	Phloem, including bast fiber, from internode containing snap point
PTGS	post-transcriptional gene silencing
PVX	Potato virus X
qRT-PCR	Quantitative real time PCR
rbcS1	Ribulose-1,5-bisphosphate carboxylase small subunit 1
RCI2A	Rare cold inducible 2A
RISC	RNA induced silencing complex
RNA	Ribonucleic acid
RNAi	RNA interference

Abbreviation	Definition
RSH3	RELA/SPOT homolog 3
RT reaction	Reverse transcriptase reaction
RT-PCR	Reverse transcriptase PCR
S	Syringyl monolignol subunit
S1	Xylem secondary cell wall outer layer
S2	Xylem secondary cell wall the middle layer
S3	Xylem secondary cell wall the inner layer
SAB-P	Analysis comparing the phloem, including bast fiber, above, at, and below the snap point; aka PAS-PSP-PBS comparison
SAM	Significance analysis of microarrays
sulfate ABC transporter	Sulfate ATP-binding cassette transporter - periplasmic sulfate-binding protein
SuSy	Sucrose synthase
TAIR	The *Arabidopsis* Information Resource
tBLASTn	Translated Basic Local Alignment Seach Tool - nucleotide
THC	Tetrahydrocannabinol
TILLinG	Targeted Induced Local Lesions in Genomes analysis
TMV	Tobacco mosaic virus
TRV	Tobacco rattle virus
TUB	Tubulin
UDP	Uracil diphosphate
UGT	Uridine diphosphate glucose coniferyl alcohol glucosyl transferase
UPF0057	Uncharacterized protein family 0057
VIGS	Virally induced gene silencing
VIGS	Virally Induced Gene Silencing
WI12	Wound-induced cell wall protein
XSP	Xylem from internode containing snap point
XTH7	Xyloglucan endotransglucosylase-hydrolase
Y	Reaction occurred

Chapter 1

Introduction and Literature Review

1 Introduction and Literature Review

1.1 History of Hemp

Cannabis sativa L. (hemp) has been cultivated for approximately 10,000 years as a fiber, oil seed, and medicinal plant, although the original purpose of its cultivation is still debated (Sneader 2005), (Russo *et al.* 2008), (Mukherjee *et al.* 2008). The use of hemp for medicine is recorded in Susruta, a 1000BC text from India; and in Pen ts-ao Ching, a Chinese text (Mukherjee *et al.* 2008). Hemp is thought to have originated in China; however, recent molecular analysis of ancient seed from 2500BC suggests two centers of origin for the plant: Northern European/Siberian and Hindustani. (Mukherjee *et al.* 2008). The current Integrated Taxonomic Information System (ITIS) classification follows the Linnaean assignment of hemp to a single species, *Cannabis sativa*; with two subspecies: indica and sativa. According to these systems, the nearest relatives to hemp, and the only other member of the Cannabacea family, are hops, i.e. *Humulus spp.* However, the recent phylogenetic analysis of the chloroplast intergenic spacer and the ribosomal intertranscribed spacer regions by Mukherjee based on archeological data suggests a more complicated evolutionary history. Modern *Cannabis* growing wild in the same locale as the archeological dig that yielded the hemp seed showed a close relationship with *Humulus japonicus* (Japanese hops), due possibly to interspecific hybridization. Crosses between Humulus and *Cannabis* have not been reported in scientific literature but grafting

of *Cannabis* onto hops has been done by horticulturalists. Other isolates of modern hemp that were tested, including a low tetrahydrocannabinol (THC) fiber genotype 'Tochigishiro' and a genotype of a medicinal indica type were more closely related to each other (Figure 1-1).

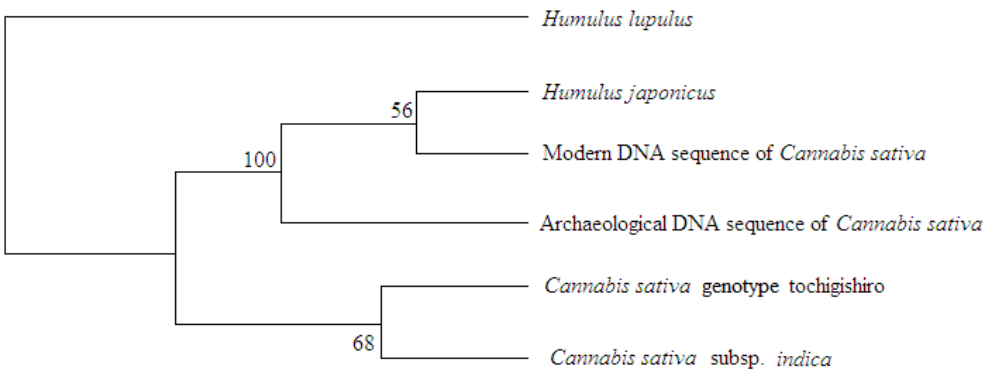


Figure 1-1: Bootstrapped maximum parsimony tree of *Cannabis* and *Humulus* species (Mukherjee *et al.* 2008)

The initial dispersion of hemp into Europe and throughout Asia occurred in pre-historical times. Landraces are domestic populations of plants that have not been purposefully selected or bred. Fiber landraces originated in Europe and East Asia, whereas medicinal and oil landraces originated in South Asia and Central Asia, respectively. Dispersal to the new world was led by the fiber landraces in the 1500s for the manufacture of rope and sail cloth, with the medicinal varieties following 300 years later in the 1800s (Clarke 1999). The initial fiber landraces cultivated in North America escaped and became weeds throughout the US and

Southern Canada (Van Zant *et al.* 1979). Cultivation and breeding of fiber hemp in North America ended in 1923 when *Cannabis* was placed on a restricted substances list, but continued in Europe and China.

Medicinal varieties of hemp have up to 39% w/w THC content (percent dry weight of leaves) (Pijlman *et al.* 2005). Breeding for low THC crops in Europe, which contain less than 0.3% w/w (percent dry weight of leaves) THC (specifically delta-9-tetrahydrocannabinol), and recent changes to regulations have allowed for renewed cultivation of oil and fiber varieties of hemp in Canada. The parentage of the modern European fiber varieties can be grouped into three major regions: North and Central Russia/Ukraine and Finland region (North-Central ecotypes); Mediterranean, Balkan, Turkey, and Caucasus region (Southern ecotypes); and China, Japan, and Korean region (Far-Eastern ecotypes) (de Meijer 1995). The relationships between the European fiber varieites can be seen in Appendix 1 – European Fiber hemp Varieties. Two of the most widely grown Canadian varieties, Anka and Carmen, are of unknown parentage (Slaski 2008).

Although a wide diversity of hemp germplasm is available, it can be difficult to develop stable varieties; so there is still a large amount of cultivation of the

various landraces. This, as well as out-crossing described below, contributes to the relatively high heterozygosity of this crop.

Hemp plants can either be male, female, or hermaphroditic, and varieties can be monecious, dioecious, or are subdioecious. Subdioecious means that most of the population will be monecious and hermaphroditic with a small number of unisexual males or females spontaneously occurring. While most varieties are dioecious; a number of monecious – hermaphroditic, and subdioecious varieties are available, as is one female unisex genotype (Clarke 1999). Sexual determination in hemp is a multigenetic trait with not only two sex chromosomes but autosomes involved (Bosca 1999). The sex of hemp can be influenced by environmental factors such as crowding, temperature, or competition for nutrients or water. In the Anka genotype, a low THC monecious genotype developed for dual use as an oil/fiber crop, there appears to be a stress-dependent component to sex selection. Unstressed plants will not produce male plants, and will continue to grow, producing only immature seed (Dear and Koziel 2006). These types of instabilities in the sex can complicate the ability to selectively breed hemp, and often breeders will look for the trait in both male and female plants before crossing them.

Hemp produces large amounts of wind-borne pollen that, due to its size, can travel long distances, which requires breeding plots to be separated by large distances to prevent unintended pollination. In most dioecious varieties hemp is an obligate outcrosser, with most hermaphroditic plants incapable of self-pollination, and continuous inbreeding resulting in sterile seeds or non-vigorous plants (Clarke 1999). Along with the problems in sex stabilization; which can cause a cross to become dioecious from monecious parents or unisex from dioecious parents potentially leading to loss of a line, fixing desired traits in the crop becomes space and time intensive. Often to produce varieties, clonal propagation is used in selective breeding rather than the more traditional methods of crossing and backcrossing.

In Canada, industrial hemp has been cultivated since 1998 when a ban on its cultivation was lifted. Licenses to cultivate approved hemp genotypes that contain a THC level of 0.3% or less in the leaves and flowering parts can be obtained from Health Canada. Canada produced 19,458 hectares of hemp in 2006; but lack of processing facilities for the fiber caused the cultivated area to drop to 3,259 hectares of hemp in 2008 (Chaudhary 2008). The provinces with the largest areas under cultivation are Saskatchewan and Manitoba followed by Alberta. The Prairie Provinces account for 95% of Canada's total production of hemp.

(Chaudhary 2005, Revised 2008). There are 27 low THC genotypes approved by Health Canada for industrial use (Table 1-1).

Variety	**Exempt from THC Testing ***
Crag	Yes
USO 14	Yes
USO 31	Yes
Anka	Exempt from testing in Ontario
Alyssa	Exempt from testing in Manitoba
Zolotonosha 11	Exempt from testing in Manitoba
C S	No
Carmagnola	No
Carmen	No
Deni	No
ESTA-1	No
Fasamo	No
Fedrina 74	No
Felina 34	No
Ferimon	No
Fibranova	No
Fibriko	No
Fibrimon 24	No
Fibrimon 56	No
Finola	No
Kompolti	No
Kompolti Hibrid TC	No
Kompolti Sargaszaru	No
Lovrin 110	No
UC-RGM	No
Uniko B	No
Zolotonosha 15	No

Table 1-1: List of approved *C. sativa* L. cultivars for 2007 in Canada (Health Canada 2007).
*Exemption is based on past testing of a variety consistently showing a THC content of 0.3% w/w.

There is growing interest in using Canadian-grown hemp fiber in the production of materials such as oriented strand board (OSB), architectural moldings, car parts, and as an alternative to wood pulp for paper. This has motivated researchers to seek a better understanding of the development of bast fibers in hemp to facilitate the breeding of fibers with properties optimized for specific industrial uses. These designer fibers will have altered lignin content, which affects fiber hydrophobicity, length and strength.

There are many uses for hemp so designer fibers need to be tailored to each process. Decreases in lignin content improve digestibility of the fibers in pulp and paper processes, composite production, cellulosic biofuel production and as animal forage. In animal forage lignin hinders the digestive enzymes from breaking down the cell wall, limiting the nutritional value of many plants to ruminants (Li, Weng and Chapple 2008). During pulping of cellulose fibers for paper production, lignin is removed by energy intensive and chemically intensive processes such as bleaching. Thus, decreases to lignin will increase the efficiency and reduce the environmental impact of paper production (Li, Weng and Chapple 2008). Composite production uses either the cellulosic fiber as reinforcements for composites (Rouison, Sain and Couturier 2006) or the lignin as filler or co-reactants in various resins and plastics (Thielemans, 2002). The use of cellulosic fiber; usually separated from lignin before being made into composites, was

observed to increase in tensile strength with decreasing fiber diameter (Rouison, Sain and Couturier 2006), (Thielemans *et al.* 2002). The increase in tensile strength with smaller diameter fibers is due to strength to hold the microfibrils together being lower than the tensile strength of the fiber. So, a larger fiber having more microfibrils is more likely to suffer an intercellular failure leading to an overall decrease in the fiber tensile strength. Thus, biologically decreasing the diameter of hemp fiber for composites use is desirable. Decreasing the lignin content for composites uses is also desirable for the same reasons as for pulping. The ability to manipulate lignin quantity and quality in a plant is important in the chemical reactions where lignin is added to different types of resins and plastics in composites (Thielemans *et al.* 2002). In using hemp for cellulosic biofuel, decreases in lignin content and modification of lignin type leads to an increase in biofuel production (Chen and Dixon 2007). This indicates that lignin modification, reduction, or increase may be advantageous depending on the process.

Fiber length and strength are properties important in composite production where they directly relate to the strength and flexibility of OSB and moldings, in which longer, or stronger fibers are required; however, these characteristics are also important in the harvesting and processing of the fibers, where fiber strength and length can add costs to processing. Hydrophobicity of hemp fibers is important to

the composites industry where the hydrophilic nature of the polysaccharides in fiber interferes with their ability to adhere to the plastic or resin matrix of composite. Lignin is in contrast hydrophobic in nature. The overall hydrophilicity of fibers is currently reduced by chemical treatments before adding the fiber to the composite; eliminating this need would be a benefit (Mutjè *et al.* 2007).

1.2 Bast Fiber Development

Bast fiber is a type of sclerenchyma present in the cortex of stems of many species. Bast fiber cells are found in the phloem of the plant, and at maturity contain primarily hemicellulose and cellulose held together in bundles by molecules including pectin and lignin. Mature bast fibers are harvested from plants such as hemp, flax, kenaf, jute, and ramie to produce textiles, composites, paper, and a variety of other materials. An elementary hemp fiber is one cell, 20mm to over 100 mm in length.

Hemp microfibrils are strands of cellulose that form the structural component of the cell wall of an elementary fiber (Figure 1-2). These cellulose strands, which make up 67-78% of the raw fiber, are embedded in a network of hemicellulose (5.5-16.1%), and pectin (0.9-4.3%) (Pejic *et al.* 2008). Lignin (3.7-8%) is contained to the area between the fiber cells (the middle lamella).

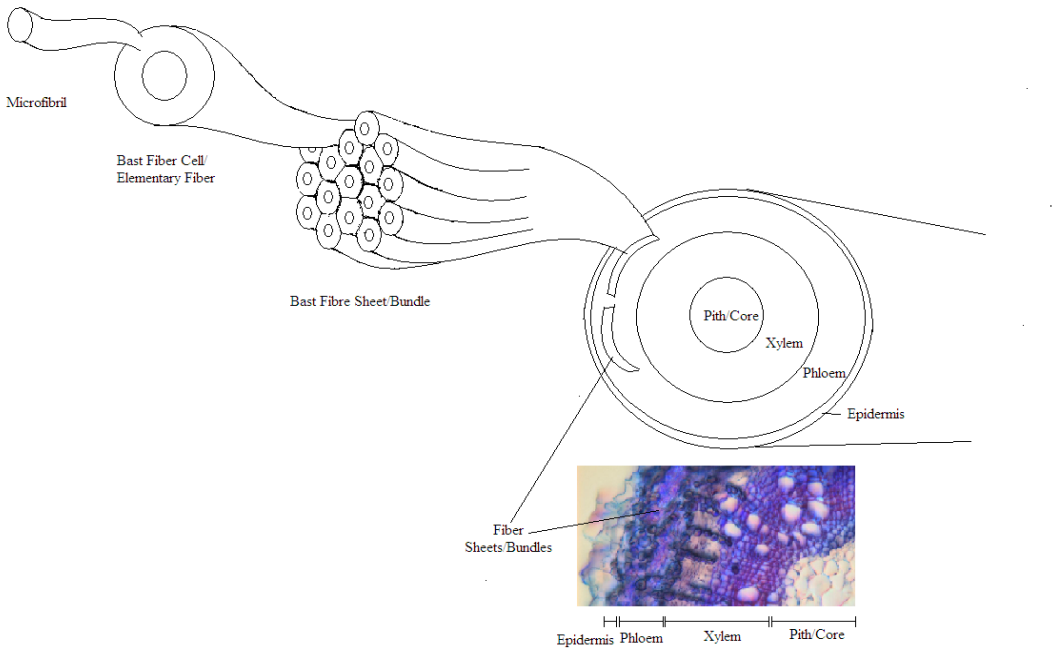


Figure 1-2: Diagram of bast microfibers in a plant stem.
Inset picture: a mature hemp stem cross section stained with safraninin and fast green, at 100X magnification under a fluorescence microscope.

Both primary and secondary fibers are seen in mature stems of hemp. The primary fibers can be over 100mm long, have thicker cell walls and are larger in diameter, 15-25μm, than the secondary fibers (Crônier, Monties and Chabbert 2005). Secondary fibers are shorter, around 2mm long, and thinner than primary fibers (Crônier, Monties and Chabbert 2005). Primary fibers develop near the apex of the plant from the protophloem of the procambium; then later from the secondary phloem of the vascular cambium. They initially elongate, growing intrusively past and into the surrounding cells, which have stopped growing. Once they have

achieved their full length, they start developing thickened cell walls. These walls eventually occlude the cytoplasm causing the death of the mature fibers. This progressive development of primary fibers is evident as a continuum along the stem, with younger fibers occurring in the apex, and older ones nearest the base. The region in which the primary fibers change from intrusive growth to wall thickening is called the snap point. Secondary fibers tend to be much shorter than the primary fibers and are found only in more mature tissues, near the base of the stem.

Chemically, the cell walls of fibers contain 65% sugar, 8% protein, 3% lignin, and 0.2% ferulic acid in the apical region; and 81% sugar, 5% protein, 4% lignin, and 0.3% ferulic acid in the basal region (Crônier, Monties and Chabbert 2005). The sugar content and the ferulic acid content increase as the plant matures; and the protein content decreases. The lignin content remains relatively constant during maturation. Secondary fibers contain less lignin then primary fibers, and do not differentiate in the apical region of the plant. Lignin content in the xylem of hemp is twice as high as in the bast (van den Broeck *et al.* 2008). Glucose is the major sugar in bast fiber cells, at 66% of total sugars in the apical region and 82% in the basal region, followed by galactose, at 9% of total sugar in the apical region and 4% in the basal region (Crônier, Monties and Chabbert 2005). Glucose content increases during maturation to approximately 84% of the total sugars; while the

content of all the other sugars except mannose fall. Mannose levels initially are around 3% in both the apical and basal regions, but increase to 5% in the apical region and 4% in the basal region during maturation (Crônier, Monties and Chabbert 2005).

Labeled antibodies to specific homogalacturons, JIM 7 and JIM5, xylogalacturon, LM7, and arabibogalactan proteins, LM2 and JIM14, were used by Blake et al in 2008 to examine the cell wall components during the development of hemp stems and determine the developmental changes in secondary cell walls. In hemp, homogalacturonan epitopes JIM7 and JIM5 are found in mature primary fibers, whereas LM 7 is observed in the spaces between developing fiber. Arabinogalactan-protein epitope LM2 is specific to developing fiber, and JIM 14 is observed in both developing and mature fiber. Crystalline cellulose, and to a lesser extent mannan epitopes are observed in developing fiber, while xylan is seen in both developing fiber and xylem. Xyloglucan epitopes are not detected in developing fibers. However, xyloglucan is seen in mature primary fibers, along with crystalline cellulose and xylan, which are seen in both primary and secondary fibers. Mannans are detectable in primary fibers and to a much lesser extent secondary fibers (Blake *et al.* 2008).

The biochemical pathways involved in fiber development are common to fibers developing in the hurd (xylem), bast (phloem), and trichome fibers (epidermis and seed). Closely related pathways are also found in trichome resins (epidermis). In hemp, the pathways involved in hurd are the pentose phosphate pathway, aromatic amino acid biosynthesis, lignin biosynthesis, or the phenylpropanoid biosynthetic pathway and the C_1 metabolic pathway (van den Broeck *et al.* 2008). These pathways have also been shown to be active in a fiber-epidermis (de Pauw *et al.* 2007). Of the pathways involved in fiber development, control of lignin biosynthesis is particularly interesting for industrial purposes.

1.3 Lignin Biosynthesis

Lignin covalently cross-links microfibillar cellulose-hemicellulose chains, which make up bast fibers (Chabannes *et al.* 2001) and is often referred to as "natures glue" in engineering (Dunky 2003). It also waterproofs the cell wall, allowing transport of water through the plant, and imparts strength and stiffness to the stem. Lignin is removed from the cellulosic fibers before they are used for pulp and paper or for composite reinforcement, in part because it can cause wear on machinery increasing maintenance and repair costs. The composition of lignin differs between hurd (xylem-based fibers) and bast fibers, which requires a change of the chemistry used in processing. However, changes or increases in lignin affecting fiber hydrophobicity may prove useful to the composites industry

(Felby, Hassingboe and Lund 2002). In the biofuel industry, a decrease in lignin causes an increase in the amount of fuel produced. Lignin imparts strength to the cell walls due to its increased cross-linking, leading to the issues with machinery and forage digestibility explained previously.

The lignin biosynthetic pathway, though well studied, has still not been completely elucidated and some of its steps are still debated. The lignin biosynthetic pathway, also called the phenylpropanoid biosynthetic pathway, leads to the production of monolignols. It converts phenylalanine to the monolignols through a series of steps starting with the deamination of phenylalanine, successive hydroxylations of the aromatic ring, O-methylation of the phenol group, and conversion of the carboxyl side chain to an alcohol (Figure 1-3).

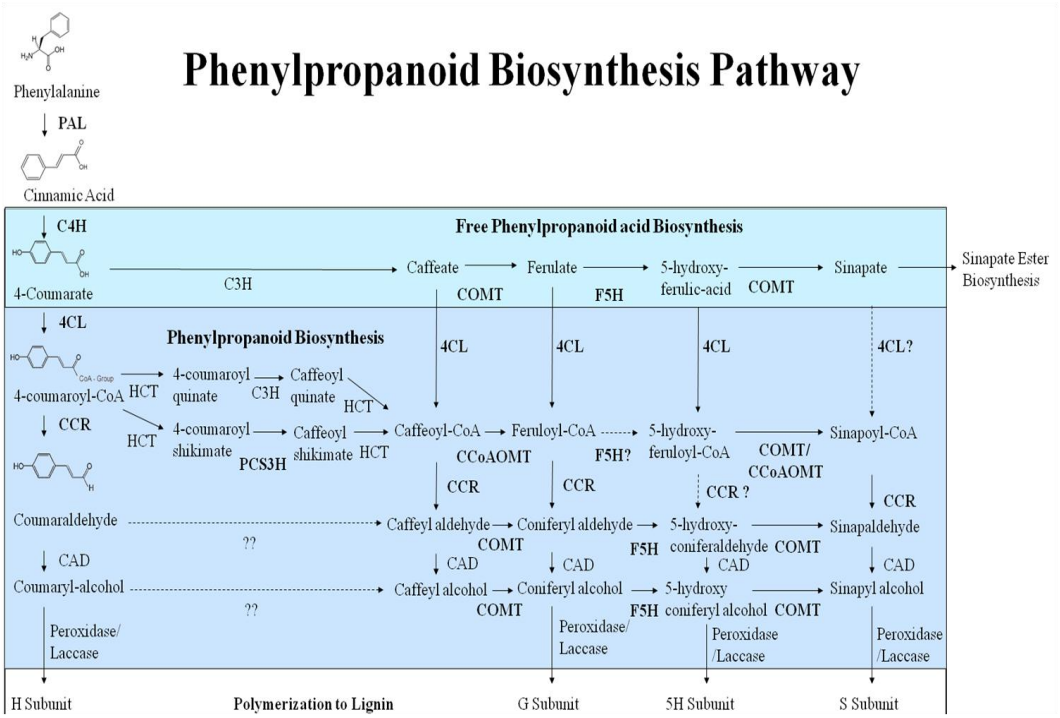


Figure 1-3: Phenylpropanoid and monolignol pathway.
PAL - phenylalanine ammonia-lyase; C4H - cinnamate 4-hydroxylase; C3H - *p*-coumarate 3-hydroxylase; COMT - caffeate *O*-methyltransferase; F5H - ferulate 5-hydroxylase; 4CL - 4-coumarate CoA Ligase; HCT - reversible acyltransferase with *p*-hydroxycinnamoyl-CoA:D quinatetransferase and *p*-hydroxycinnamoyl-CoA:shikimate *p*-hydroxycinnamoyltransferase activity; PCS3H - *p*-coumaroyl shikimate 3'-hydroxylase; CCoAOMT - caffeoyl-CoA O-methyltransferase; CCR - cinnamoyl-CoA reductase; CAD - cinnamyl alcohol dehydrogenase. Question marks indicate unknown genes or genes whose activity has not been confirmed. Enzymes in bold are those later examined in this thesis.

Lignins are a heteropolymer made from a mix of mainly three monolignols: p-coumaryl alcohol (M1H), coniferyl alcohol (M1G), and sinapyl alcohol (M1S); which differ in their degree of methoxylation. The monolignols produce subunits in the polymer called: p-hydroxyphenyl (H), guaiacyl (G), and syringyl (S). The ratios of these subunits differ in different plants, cell types, cell wall layer, and

changes at different maturity stages and in different environmental conditions. Lignification is the polymerization of monolignols in a reaction which adds subunits onto the end of a growing chain. Usually the subunits are linked at the beta position of the monolignol to the 5-O-4 unit of the growing polymer via a radical coupling reaction producing a beta-O-4 (called an A linkage), beta-5 (a B linkage), beta-beta (C), 5-5 (D), 5-O-4(E), or a beta-1 (F). Dimerization of monolignols also occurs producing a variety of dimers linked beta-O-4 (A linkage), beta-5 (B), or beta-beta (C). Beta-O-4 linkages are the most easily cleaved by chemical processes, and provide a base for industrial processes (Boerjan, Ralph and Baucher 2003). The abundance of the different linkages can be determined by the amount of each monomer in the lignin polymers; so the ratios of H, G, and S components in the fiber source are important to industry.

Other monolignols are also present in plants and are seen incorporated into the lignin polymers of transgenic plants that have altered production levels of the three major monolignols. These are derived from alternate monomers or have been through reactions that alter the side chains such as acetylation, oxidation, and incorporation of non-methylated precursors, aldehydes of the precursors, or ferulates. While these alternate lignin monomers are seen in plants that are deficient in monolignol production, they are also seen in low levels in normal plants (Boerjan, Ralph and Baucher 2003).

Transport of synthesized monolignols to the cell wall, where they are incorporated into lignin polymers, is hypothetically accomplished by coniferin-beta-glucosidase (CG) and uridine diphosphate glucose coniferyl alcohol glucosyl transferase (UGT) found in secondary cell walls of the xylem. The polymerization of monolignols occurs via dehydrogenation through a variety of enzymes such as peroxidases, laccases, polyphenol oxidases, and coniferyl alcohol oxidase. The number of peroxidases found in plant tissues has obscured the identification of a specific peroxidase isozyme in the role of monolignol dehydrogenation. Laccases are expressed during the lignification of cells in a number of species. However, transgenic aspen and poplar with down-regulated genes for laccase did not affect lignin content or composition (Boerjan, Ralph and Baucher 2003).

In xylem cell differentiation, the best characterized lignin system, secondary cell walls consist of three layers in the xylem: S1, the outer layer; S2 the middle layer; and S3 the inner layer. Deposition of lignin is preceded by carbohydrate deposition when S1 formation is initiated. Lignification proceeds through the secondary wall once the polysaccharide matrix of the S2 layer is complete, and finally the bulk of lignin is deposited in the S3 layer after hemicellulose and cellulose is deposited. H units are typically deposited first, then G, and finally S units (Boerjan, Ralph and Baucher 2003). Unlike xylem fibers where the bulk of S units are deposited in the S3 layer, the high amounts of crystalline cellulose found

in bast fiber may impair the lignification, leading to a consistent S/G ratio in mature fiber. Hemp fiber like flax bast fiber has lignification restricted to the middle lamella, but is more lignified and has a lower proportion of H units and a higher S/G ratio then flax (Crônier, Monties and Chabbert 2005) (Figure 1-4).

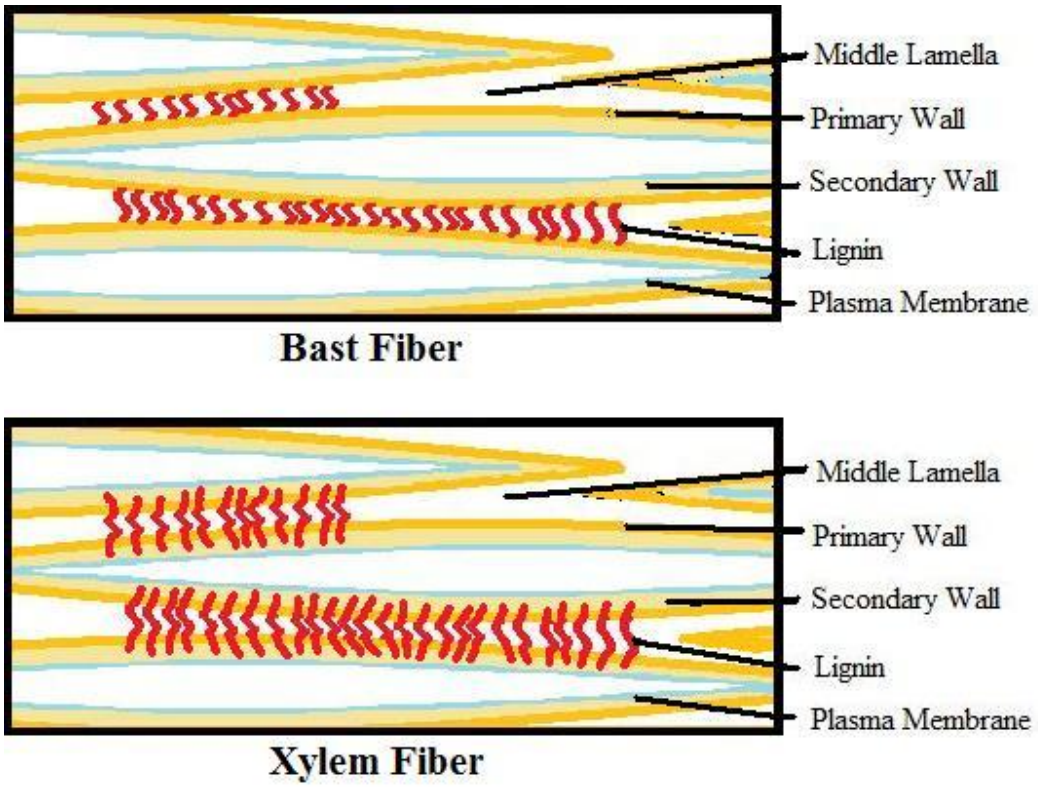


Figure 1-4: Diagram showing the difference in lignin deposition in xylem and phloem fiber

In hemp bast fiber, the S/G ratio was highest in basal samples of stem and did not change significantly as the plant matured in either the apical or basal regions. This indicates that hemp bast contains condensed lignin that is resistant the chemical

degradation used to measure monolignol contents. H units occur in larger proportions in younger plants and in the apical region, but their amounts do not increase as the plant matures which leads to a decrease in their proportion in older plants. The presence of H units is indicative of highly condensed lignin but the decrease of the proportion of H units in later depositions suggests a less condensed polymer exists in later cell walls (Crônier, Monties and Chabbert 2005).

The phenylpropanoid pathway is highly conserved in plants, despite the differences within spatial and temporal deposition of monolignols within plant cell types and cell wall layers. Phenylalanine ammonia lyase (PAL) is part of a multigene family in most plants with individual genes theoretically playing specific metabolic roles. Cinnamate 4-hydroxylase (C4H) in alfalfa has at least two forms. C4H-1 is involved in lignification, but the role of C4H-2 has not been completely determined. 4-coumarate CoA ligase (4CL) is a multigene family with different substrate specificities, and spatial/temporal expression patterns. Caffeate *O*-methyltransferase (COMT) and caffeoyl-CoA *O*-methyltransferase (CCoAOMT) are usually part of a multigene family but aspen only has one known CCoAOMT isoform. Cinnamoyl-CoA reductase (CCR) has two known isoforms in *Arabidopsis* and maize but homology searches in *Arabidopsis* suggest that it may be part of a larger gene family. In other plants, such as poplar, tobacco,

sugarcane, and loblolly pine. CCR only has one known isoform, despite there being three different substrates for the enzyme. Cinnamyl alcohol dehydrogenase (CAD) exists in various isoforms in angiosperms, but only one is recognized in gymnosperms. The presence of *p*-coumarate 3-hydroxylase (C3H), ferulate 5-hydroxylase (F5H), reversible acyltransferase with *p*-hydroxycinnamoyl-CoA:D quinatetransferase and *p*-hydroxycinnamoyl-CoA:shikimate *p*-hydroxycinnamoyltransferase activity (HCT), and *p*-coumaroyl shikimate 3'-hydroxylase (PCS3H) isoforms in *Arabidopsis* have not been determined (Anterola and Lewis 2002).

Transgenic studies in numerous plant species allow insight into the larger phenotypic effects seen when the lignin biosynthesis pathway is altered. COMT, PAL, and CCoAOMT down-regulation improves digestibility in ruminants (Boerjan, Ralph and Baucher 2003), (Guo *et al.* 2001). PAL down-regulation reduces G units and C4H down-regulation reduces S units in transgenic tobacco (Sewalt *et al.* 1997). PAL over-expression increases levels of chlorogenic acid in leaves, and PAL and C4H down-regulation decreases chlorogenic acid (Boerjan, Ralph and Baucher 2003), (Bate *et al.* 1994). Down-regulation of CCoAOMT in alfalfa and poplar decreases the G units, causing an increase in the S/G ratio (Guo, Chen *et al.* 2001), (Meyermans *et al.* 2000). In CCoAOMT down-regulated poplar the lignin was also less condensed and less cross-linked (Zhong *et al.* 2000).

Alternatively, down-regulation of COMT reduces S units and substitutes 5H units into the lignin polymer and causes an overall reduction of lignin in alfalfa, maize, and poplar (Boerjan, Ralph and Baucher 2003), (Jouanin *et al.* 2000). However, 5H is not cleavable in alkaline solutions so COMT down-regulated poplar has lignin which is more difficult to extract from pulp (Boerjan, Ralph and Baucher 2003), (Lapierre *et al.* 1999).

Plants which are down-regulated in 4CL, CCoAOMT, CCR, CAD, or COMT all show a red/brown discolouration in the xylem (Baucher, Chabbert *et al.* 1996), (Higuchi *et al.* 1994). 4CL down-regulation in tobacco, *Arabidopsis*, and aspen reduces lignin content and the amounts of p-coumaric, ferulic, and sinapic acids (Boerjan, Ralph and Baucher 2003), (Hu *et al.* 1999), (Kajita *et al.* 1997), (Lee *et al.* 1997). It reduces S units in tobacco, G units in *Arabidopsis*, and has a negligible effect on the S/G ratio in aspen; but these differences may be due to the presence of numerous isoforms of 4CL present it the plants (Anterola and Lewis 2002), (Lee *et al.* 1997), (Hu *et al.* 1999). Aspen with transgenically down-regulated 4CL contains less lignin, more cellulose, and grows taller (Hu *et al.* 1999). Down-regulation of CCR causes an increase in S/G ratio in tobacco due to a reduction in G units, but the lignin is more condensed in part from the increase in tyramine ferulare units incorporated into the polymer (Boerjan, Ralph and Baucher 2003), (Piquemal *et al.* 1998). CCR *Arabidopsis* mutants have an

expanded S2 wall with cellulose microfibrils that lack cohesion (Boerjan, Ralph and Baucher 2003), (Jones, Ennos and Tunner 2001). The lignin in these plants shows a high amount of cross-linking due to the buildup of hydroxycinnamates. CAD down-regulation is only slightly decreases the lignin content of the plants, despite the fact that it is the last step in the pathway (Baucher, Chabbert *et al.* 1996). This is thought to be due to the incorporation of other phenolics; which compensate for the lack of the major monolignols. Poplar down-regulated for CAD have an altered lignin structure that decreases the amount of chemicals used to remove lignin in wood pulp, but which have growth and development that are not significantly affected (Baucher, Chabbert *et al.* 1996).

The last sets of transgenic plants to be discussed are F5H and C3H mutants. F5H mutants in *Arabidopsis* show no S units (Anterola and Lewis 2002), (Chapple *et al.* 1992). Up-regulation of F5H gives lignin almost entirely composed of S units in *Arabidopsis*, poplar, and tobacco (Boerjan, Ralph and Baucher 2003), (Franke *et al.* 2000). Down-regulated C3H *Arabidopsis* mutants have highly reduced lignin contents and what lignin is there is mostly composed of H units (Boerjan, Ralph and Baucher 2003), (Franke, Hemm *et al.* 2002).

1.4 Objectives

Given the importance of secondary wall development, and specifically lignin biosynthesis to industry, the identification of the relevant genes in *C. sativa* bast fiber is important. In an attempt to identify and characterize a series of target genes for the Alberta Research Council (ARC) biofibers reverse genetics hemp TILLinG platform, I first analyzed two sets of microarray experiments; one comparing the phloem, including the bast fibers, above, below and at the snap point, the other comparing snap point epidermis, phloem (including bast fiber), and xylem, to identify genes up-regulated in bast fiber. Then I employed qRT-PCR to further examine the expression pattern of some of the genes found in the microarray analysis, or genes that were suspected of being active in lignin synthesis through a literature search. I compared the microarray and qRT-PCR results to establish an expression profile for each gene of interest. Alignments of ESTs and phylogenetic trees of various isoforms of each gene were made to expand the information known about the ESTs in question. I also attempted to develop a virally induced gene silencing (VIGS) system for rapidly evaluating the function of genes of interest found in the above study for changes in fiber characteristics.

Chapter 2

Microarray analysis of developing hemp stems

2 Microarray Analysis of Developing Hemp Stems

2.1 Introduction

The development of phloem-derived bast fibers has had relatively little genetic research applied to it compared to that done on the fibers that make up the woody core of a plant. The bast fibers of flax (*Linum usitatissimum* L.) have the most detailed genetic research of all the plants that are used to commercially produce bast fiber (Gorshkova and Morvan 2006), (Roach and Deyholos 2007), (Blake *et al.* 2008). In flax, a tightly cellulose-bound tissue-specific galactan is the major non-cellulosic polysaccharide (Salnikov, Ageeva and Gorshkova 2008), with other pectin galactins, galacturnic acid, arabinogalactan, galactomanan, rhamnogalacturonan, and xyloglucan making up the vast majority of the remaining polysaccharides (Gorshkova and Morvan 2006). Galactose is the main non-cellulosic sugar in hemp during its vegetative stages, at 9% of the total non-cellulosic sugars in immature fibers and 4% in mature fibers below the snap point (Crônier, Monties and Chabbert 2005). Apart from sugars, which make up 65-80% of the total cell wall of hemp, proteins at 5-8% and lignin at 3-4% are the other major components. Ferulic acid, which may indicate protein and polysaccharide linkages, is almost the exclusive cell wall bound phenolic acid present at 0.2-0.6%.

Microarrays have been used to characterize differences in gene expression in stem peelings from three axial segments of the hemp stem (de Pauw *et al.* 2007)

defined by the authors as the top (apical), mid (snap-point), and bottom (basal) regions of the stem. Above the snap point, transcripts for expansin, pectinesterates, and glycosyl hydrolases were relatively abundant. At the snap point and below, stem peels were enriched in genes for secondary wall deposition including transcripts involved in cellulose deposition (e.g. sucrose synthase (SuSy), cellulose synthase A (CesA), cobra-like 4 (CBL4)); and other putative cell-wall processes (arabinogalactan proteins (AGPs), lipid transfer proteins (LTPs), glycosyl transferase), and enzymes in the phenylpropanoid pathway (caffeate O-methyltransferase (COMT), ferulate 5-hydroylase (F5H)). The results were similar to the flax model (Roach and Deyholos 2007), in which stem peels above the snap point were enriched in some transcripts associated with cell elongation, while enrichment of secondary wall related transcripts were most evident at the snap-point and below.

Microarrays have also been used to characterize differences in gene expression in other tissues of hemp, including comparisons of bast fibers and xylem fibers (Reijmers *et al.* 2005). The biochemical pathways involved in fiber development seen when comparing xylem secondary wall formation to bast are the pentose phosphate pathway, aromatic amino acid biosynthesis, lignin biosynthesis, or the phenylpropanoid biosynthetic pathway; and the C_1 metabolic pathway (van den Broeck *et al.* 2008). These same pathways are also seen enriched in bast fiber in

the basal region, when compared to bast from apical regions (de Pauw *et al.* 2007) and make up a generalized set of secondary wall formation pathways.

Since the pathways in both the Reijmers study and the dePauw study are the same, there is a need to look at both studies to answer the question of what the differences of the phloem fibers compared to the xylem fibers at the snap point are but not above or below it. Also the epidermis, which contains trichomes, may be obscuring results for genes involved in bast fiber development. Here we present a microarray examination of more tissue types. We look at gene expression in phloem fiber with specific attention to bast from the node containing the snap-point compared to the nodes above and below it (SAB-P); as well as a comparison between the bast fiber and the other major tissue components present in the stem: the epidermis, and the xylem at the snap point (PEX-S).

2.2 Methods

2.2.1 Tissue Dissection

The following tissues were dissected (Table 2-1) and (Figure 2-1) from six pools of six to ten five-week old plants of *C. sativa* genotype Carmen, where the internode containing the snap point was manually defined for each individual plant (Appendix 2 – Snap Point Determination). First the nodes above and below the internode are cut, and the layer containing the fiber and phloem are stripped from the xylem and core. Then the epidermis is dissected from the phloem. Above the snap point the epidermis and phloem layers are quite soft, and tear easily. Below the snap point the mature fibers are more easily dissected from the epidermis (Figure 2-2). At the snap point the epidermis was always stripped from fiber/phloem starting at the bottom to the top of the internode as it was easier to dissect the layers this way.

Abbreviation	Tissue Description
PAS	Phloem from internode above snap point
ESP	Epidermis from internode containing snap point
PSP	Phloem (including bast fiber) from internode containing snap point
XSP	Xylem from internode containing snap point
PBS	Phloem (including bast fiber) from internode below snap point
PEX-S	Experiment comparing snap point internode tissues: phloem, epidermis, and xylem
SAB-P	Experiment comparing phloem above, below and at the snap point.

Table 2-1: Table of dissected tissue type and experiment abbreviations

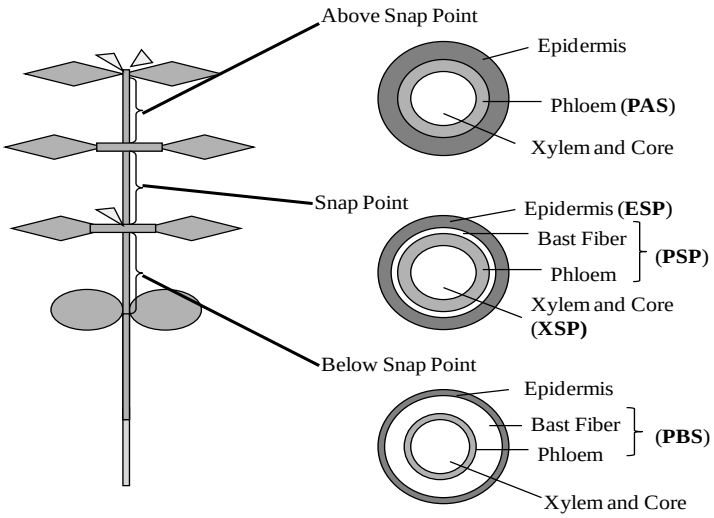


Figure 2-1: Diagram of dissections: Black and white schematic:
Above the snap point only phloem and elongating fibers were dissected (PAS). At the snap point epidermis (ESP), the phloem layer, including bast fiber (PSP) and the xylem (XSP) layers were dissected. Below the snap point phloem layer, including bast fiber was collected (PBS).

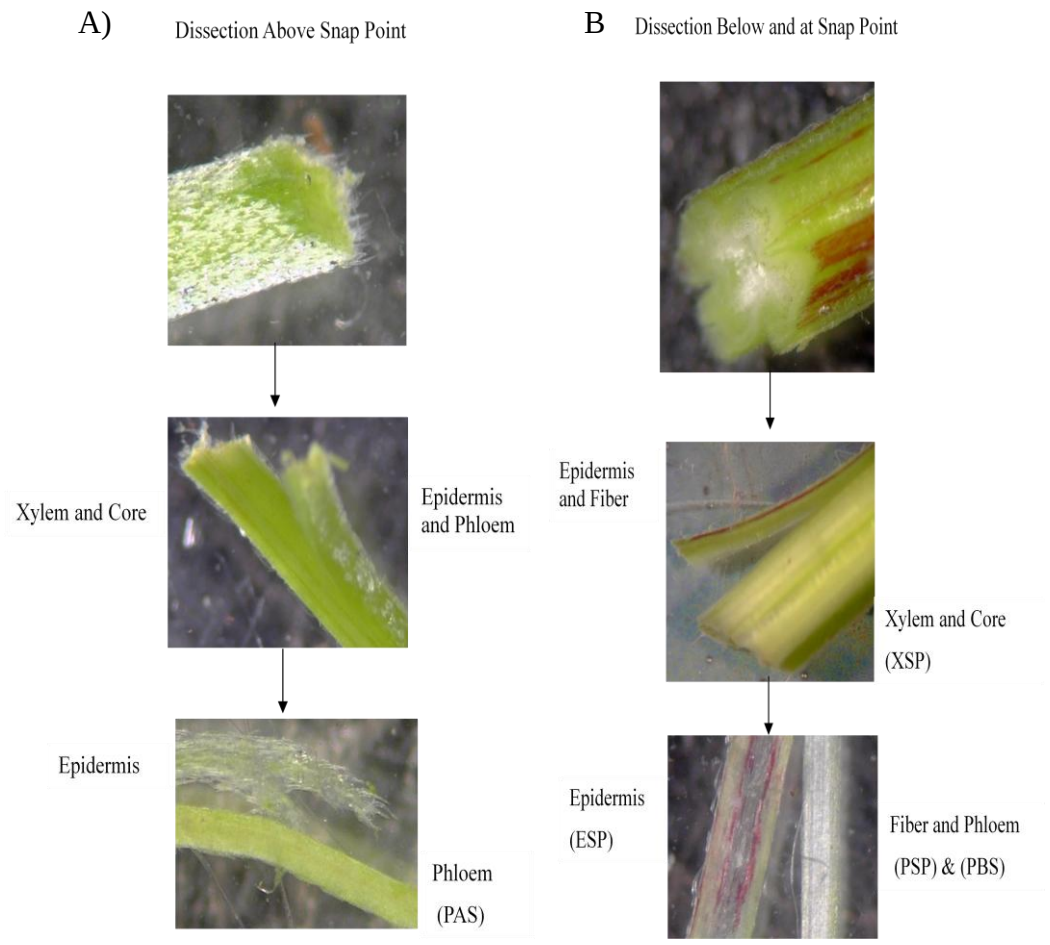


Figure 2-2: 2X dissection scope light microscopy pictures of dissection. A: above the snap point. B: below the snap point. Not shown: Pictures of dissection at the snap point. The internode containing the snap point at its top resembles the above snap point picture, and at its bottom resembles the below snap point picture.

2.2.2 RNA Extraction

Total RNA was extracted from the various tissues via the Qiagen RNeasy Mini Plant Kit protocol (QIGEN Group 2006) modified by adding 50uL 20% high molecular weight polyethylene glycol (PEG), and 10uL Beta mercaptoethanol per 1 mL of RLC buffer. Multiple aliquots of each tissue were processed through separate Qiashedder columns, and then combined via processing them through a single Qiaquick extraction column. All the samples were treated with DNase using the TURBO DNA-free Kit from Ambion.

2.2.3 Microarray Processing

Custom microarray slides based on a cDNA library of stem peels of 4 week-old *Cannabis sativa* L. genotype Carmen were used (de Pauw *et al.* 2007). For hybridization, MMuLV reverse transcriptase (Fermentas) was used to synthesize cDNA from 500ng of total RNA from three biological replicates of three tissue types for the first experiment (PEX-S): snap point internode epidermis (ESP), phloem, containing bast fiber (PSP), and xylem (XSP). The same was done with 500ng of total RNA two biological replicates of three tissue types for the second experiment (SAB-P): internode phloem above snap point (PAS), and phloem containing snap point (PSP), and below snap point (PBS). The cDNA was primed with poly-A-capture oligomers from the Oyster Dye microarray labeling kit allowing each to be tagged with 3' end labeled fluorescent primer (Genisphere, Hatfield, PA, USA).

The hybridizations were set up for each pair of possible hybridizations for each three comparisons (Table 2-2).

Hybridization	cDNA labeled withCy3 Dye	cDNA labeled with Cy5 Dye
1	PSP	ESP
2	XSP	ESP
3	ESP	PSP
4	XSP	PSP
5	ESP	XSP
6	PSP	XSP
7	PSP	PAS
8	PBS	PAS
9	PAS	PSP
10	PBS	PSP
11	PAS	PBS
12	PSP	PBS

Table 2-2: Table of microarray hybridizations.

Arrays were hybridized at 65°C overnight using a 24X50mm cover slip in a humidified centrifuge tube. Cy3 and Cy5 hybridizations were performed for 5hours at 65°C and scanned using the Array Worx (Applied Precision, Issaquah, WA, USA). Two sets of the PAS-PSP-PBS arrays were processed (12 hybridizations) and three sets of the ESP-PSP-XSP arrays were processed (18 hybridizations). Spot intensities were determined using Spotfinder v3.11 and normalized through MIDAS v2.19 via the Lowess algorithm.

2.2.4 Microarray Statistics

MEV v4.2 was used to determine significant results using SAM (Significance analysis of microarrays) in a multiclass test with a false gene discovery median rate of 2.02% or 80.16 for ESP-PSP-XSP and 1.39% or 80.66 for PAS-PSP-PBS. The K-means clustering support for the samples (three clusters – each hybridization and dye flip set) was used, at 500 iterations with a threshold co-occurrence of 80%, to be sure all the samples grouped as expected. The single slide which had a processing error did not group with any of the clusters in the K-means clustering support was eliminated from the data set. The significant genes found via SAM analysis were checked with a figure of merit to determine the number of clusters to use for K-means clustering of the genes. K-means clustering was used to group the genes into 12 groups so the expression profiles of the significant genes could be determined. Principal components analysis, with a KNN neighbours input of 10 was completed to determine how much overlap was seen in each sample set.

2.3 Results

Dissection of the phloem from the bast fibers was not possible using a light microscope so the bast fiber and the other phloem cells were considered as one unit. Two sets of microarray comparisons were completed on mRNA prepared from the same samples, the first compared phloem including bast fiber to the

epidermis and xylem at the snap point (PEX-S), and the second compared the phloem (including bast fiber) above, at, and below the snap point (SAB-P).

2.3.1 Gene Expression in Snap Point Phloem (Including Bast Fiber) Compared to Epidermis and Xylem (PEX-S)

After an initial SAM analysis a K-means clustering support analysis on the samples showed one of the hybridizations as an outlier. The removed slide had very little cy5 reaction, due to a processing mistake, so the data from that slide was removed from further analysis and the SAM results were reprocessed without the outlier. From a fiber-enriched cDNA library microarray of 7053 independent clones, 2790 showed a significant increased expression level in snap point epidermis, bast fiber/phloem, or xylem/core tissues compared to the other two tissues based on a multiclass SAM analysis. SAM and clustering analysis are discussed further below. Principal components analysis (PCA) of the genes shows that the fiber/phloem and the xylem/core components at the snap point are more closely related to each other than to the epidermis (Figure 2-3).

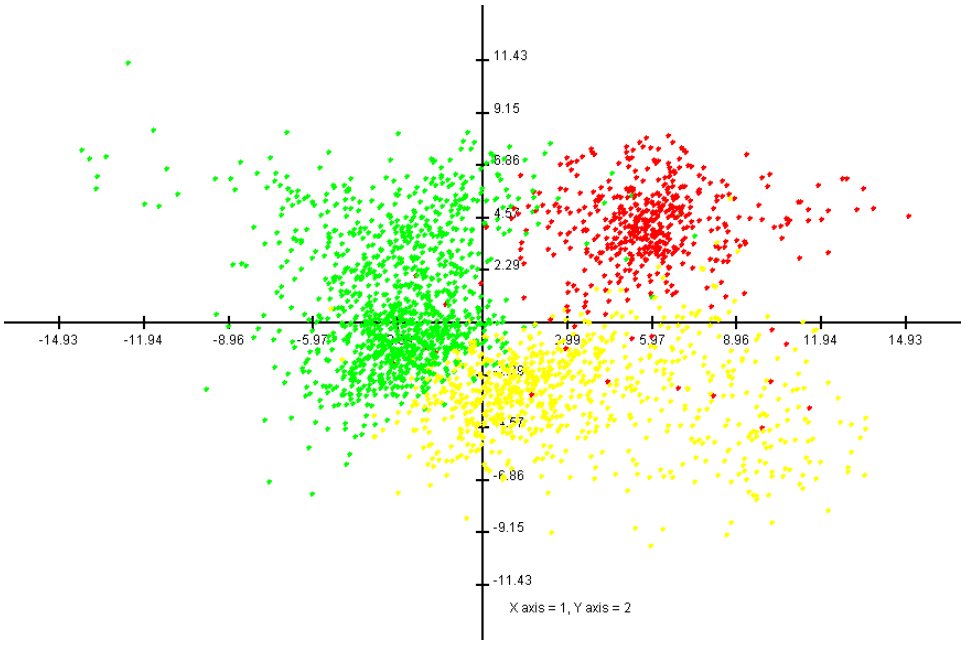


Figure 2-3: PCA analysis of the SAM significant genes in the phloem. Epidermis (ESP) – yellow, Phloem and bast fiber (PSP) – green, and Xylem (XSP) – red.

Figure of merit (FOM) analysis, a statistical method used to determine the best number of clusters to use as parameter for clustering analysis, for the significant genes determined an optimal number of clusters to use in the K-mean cluster (KMC) analysis of the genes. The FOM suggested that the optimal number of clusters was approximately nine. Nine clusters were initially created, but that only resolved one cluster with increased expression in epidermis. In an attempt to distinguish the genes increased in epidermis but also increased in phloem (over xylem) more clusters were created. The optimal number that would distinguish the two possible expression patterns in the epidermis (epidermis greater than

phloem and greater than xylem; or epidermis greater than xylem and greater than phloem) was determined to be 12.

For the twelve clusters from the KMC analysis performed on the significant genes found via SAM analysis, the log ratio expression values for all the replicates slides were averaged and plotted to give expression graphs for each cluster; which were then arranged in three groups: those most highly expressed in snap point epidermis, phloem, or xylem (Figure 2-4). There are 13 possible combinations for the clusters (Table 2-3), one of which would not show up as statistically significant (ESP=PSP=XSP). In this analysis there were no statistically significant genes that fell into the combinations ESP=XSP>PSP, ESP=PSP>XSP, or ESP>XSP>PSP. This could be due to the library used to make the microarray slides was made from phloem and epidermal tissue, and there are not enough xylem genes to distinguish a statistical different between the epidermis and xylem. There were genes in the epidermis that showed increased expression compared to the Phloem (ESP>PSP>XSP and ESP>XSP=PSP). The twelve clusters that were determined using KMC analysis showed 2 clusters with increased expression in epidermis, 5 with increased expression in phloem, and 5 with increased expression in xylem. The extra clusters for each tissue are ones that show less extreme differences in the analysis. For example in phloem PSP>XSP=ESP is expected; but PSP>XSP>=ESP and PSP>XSP=ESP are seen.

Cluster combinations	Cluster seen in KMC analysis
ESP=XSP>PSP (same as XSP=ESP>PSP)	No
ESP=PSP>XSP (same as PSP=ESP>XSP)	No
PSP=XSP>ESP (same as XSP=PSP>ESP)	Yes
ESP>XSP>PSP	No
ESP>PSP>XSP	Yes
PSP>XSP>ESP	Yes
PSP>ESP>XSP	Yes
XSP>ESP>PSP	Yes
XSP>PSP>ESP	Yes
ESP>XSP=PSP (same as ESP>PSP=XSP)	Yes
PSP>XSP=ESP (same as PSP>ESP=XSP)	Yes
XSP>ESP=PSP (same as XSP>PSP=ESP)	Yes
XSP=ESP=PSP	Not significant

Table 2-3: KMC Cluster combinations for PEX-S experiment.
Epidermis (ESP), Phloem and bast fiber (PSP), and Xylem (XSP) at the snap point.

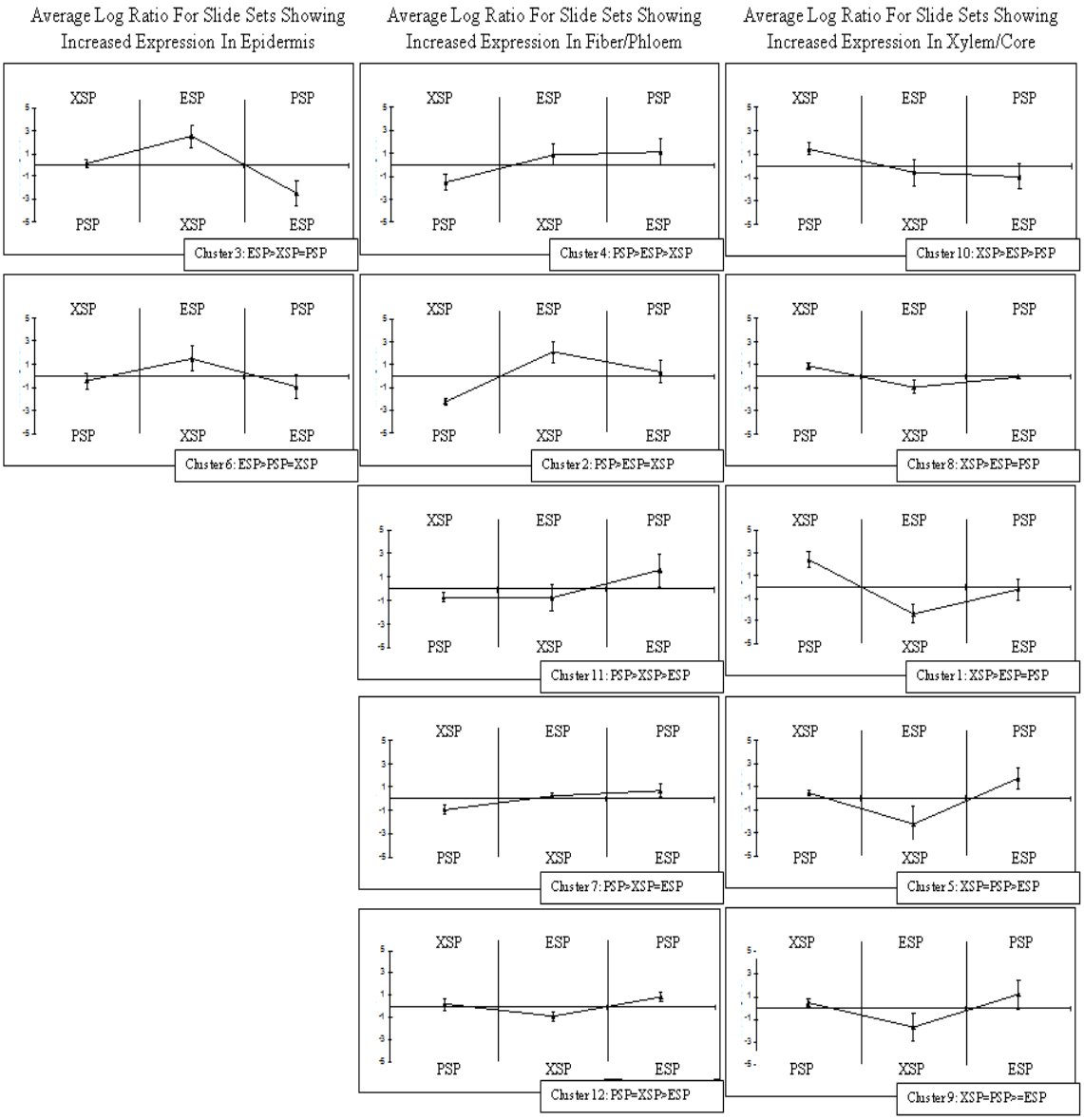


Figure 2-4: PEX-S experiment graphs of cluster analysis showing expression pattern and average log intensity ratios for each cluster.
XSP: snap point xylem, ESP: snap point epidermis, PSP: snap point phloem including bast fiber.

The microarray slides were made by spotting 7053 cDNA ESTs in duplicate (a total of 14113 spots). The multiclass SAM analysis was performed to determine the statistically significant spots. SAM analysis carries out permutations of gene specific t-tests and computes the strength of the gene expression compared to a response variable. It does not assume an equal variance or an independence of any of the genes. A total of 3186 spots (of which 813 corresponded to genes that had functions that could be inferred from annotation) were assigned to clusters (Table 2-4). 10927 spots did not show significant differential expression and fell below a delta of 0.48 in multiclass tests according to SAM analysis. The difference of 396 spots between the total number of genes in the clustering and the total number of significant genes in the SAM analysis is due to replicate spots on the microarray placed in different clusters during the analysis.

Cluster Definition	Number of Annotated Spots	Number of Un-known Spots	Number of Un-sequenced Spots	Total Number of Spots	Final Class
Cluster 1: XSP>ESP>=PSP	38	5	42	85	XSP
Cluster 2: PSP>=ESP>XSP	64	6	115	185	PSP
Cluster 3: ESP>XSP>=PSP	33	10	110	153	ESP
Cluster 4: PSP>ESP>XSP	45	5	100	150	PSP
Cluster 5: XSP>PSP>ESP	39	14	80	133	XSP
Cluster 6: ESP>PSP>=XSP	124	22	536	682	ESP
Cluster 7: PSP>ESP>=XSP	78	14	194	286	PSP
Cluster 8: XSP>ESP>=PSP	91	8	206	305	XSP
Cluster 9: XSP>PSP>=ESP	92	12	184	288	XSP
Cluster 10: XSP>=ESP>PSP	58	9	201	268	XSP
Cluster 11: PSP>XSP>=ESP	55	8	167	230	PSP
Cluster 12: PSP=XSP>ESP	96	16	309	421	PSP
Totals	813	129	2244	3186	

Table 2-4: Numbers of spots in each cluster and its class designation.

In a multiclass SAM analysis the MEV v4.2 program, the option to choose genes that show a specific fold change does not exist. So, for the significant spots determined by SAM analysis, the log2 intensities for the duplicate spots were averaged for each tissue and standard error was calculated. The tissue with the highest log2 intensity was used to assign each set of duplicate spots into one of three groups (ESP - 178, PSP - 169, and XSP - 367 annotated genes) (Table 2-5). The tissue classes were then divided into functional classes based on the those used in de Pauw (2007) and a graph was made of the percent of the total number of annotated genes (for this experiment – 714 genes) at each functional class for each tissue (Figure 2-5). Then the values for each tissue were subtracted in a pair wise manner and those with a 1.5 fold change or higher in one of the two comparisons was used to create the log 2 intensity graphs for each functional class.

Functional Class	ESP	PSP	XSP	Totals
Photosynthesis - chloroplast	7	5	27	39
Glycolysis & respiration	4	2	13	19
Amino acid & protein metabolism	11	21	22	54
Secondary metabolism	9	7	26	42
Oxidoreductases	20	13	19	52
Regulation	16	13	48	77
Transport	7	8	52	67
Glycosyltransferases	5	4	9	18
Glycosylhydrolases	12	24	11	47
Other cell wall enzymes	2	1	7	10
Other cell wall proteins	23	21	18	62
Miscellaneous	55	41	79	175
Unknown function	7	4	25	36
Phenylpropanoid biosynthesis	0	5	11	16
Totals	178	169	367	714
Total Including non-annotated genes	826	1152	1014	2992

Table 2-5: Numbers of genes in each functional class, based on a pair-wise comparison of the average intensity for significant genes found by SAM analysis

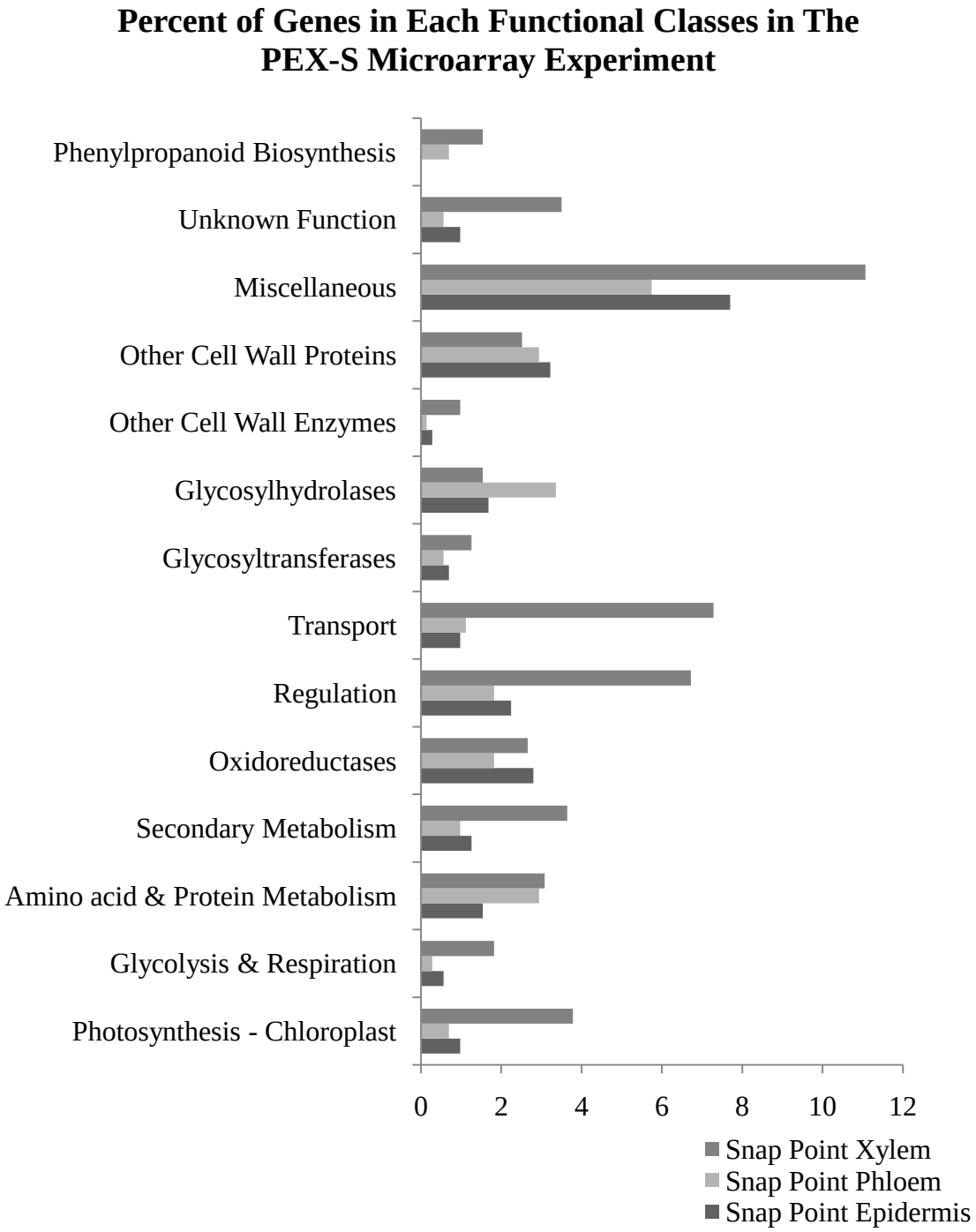


Figure 2-5: Percentage of total number of annotated genes for each functional class in the PEX-S microarray experiment.

Of the 714 annotated genes, 36 were listed as having no known function. Duplicate listings of contigs were deleted, but multiple cluster definitions were retained. The ESP cluster has 178 annotated genes, PSP 169, and XSP 367. See Appendix 3 for Gene Listings.

Genes that did not fit into any other functional class were placed in the miscellaneous class. Of the 175 miscellaneous ESTs, 41 were up regulated in phloem (Figure 2-6). Of these, the ESTs with differences of more than 1.5 log2 units in fiber are: Argonaute (AGO1) protein, ATP Sulfurylase, dehydrin 1, DUF250 - P0028E10.8 protein, two DUF642 - hypothetical proteins, nucleoid DNA-binding - CND41-like protein, stress-related - rubber elongation factor - protein, sulfate adenylyltransferase, and a TUB family protein.

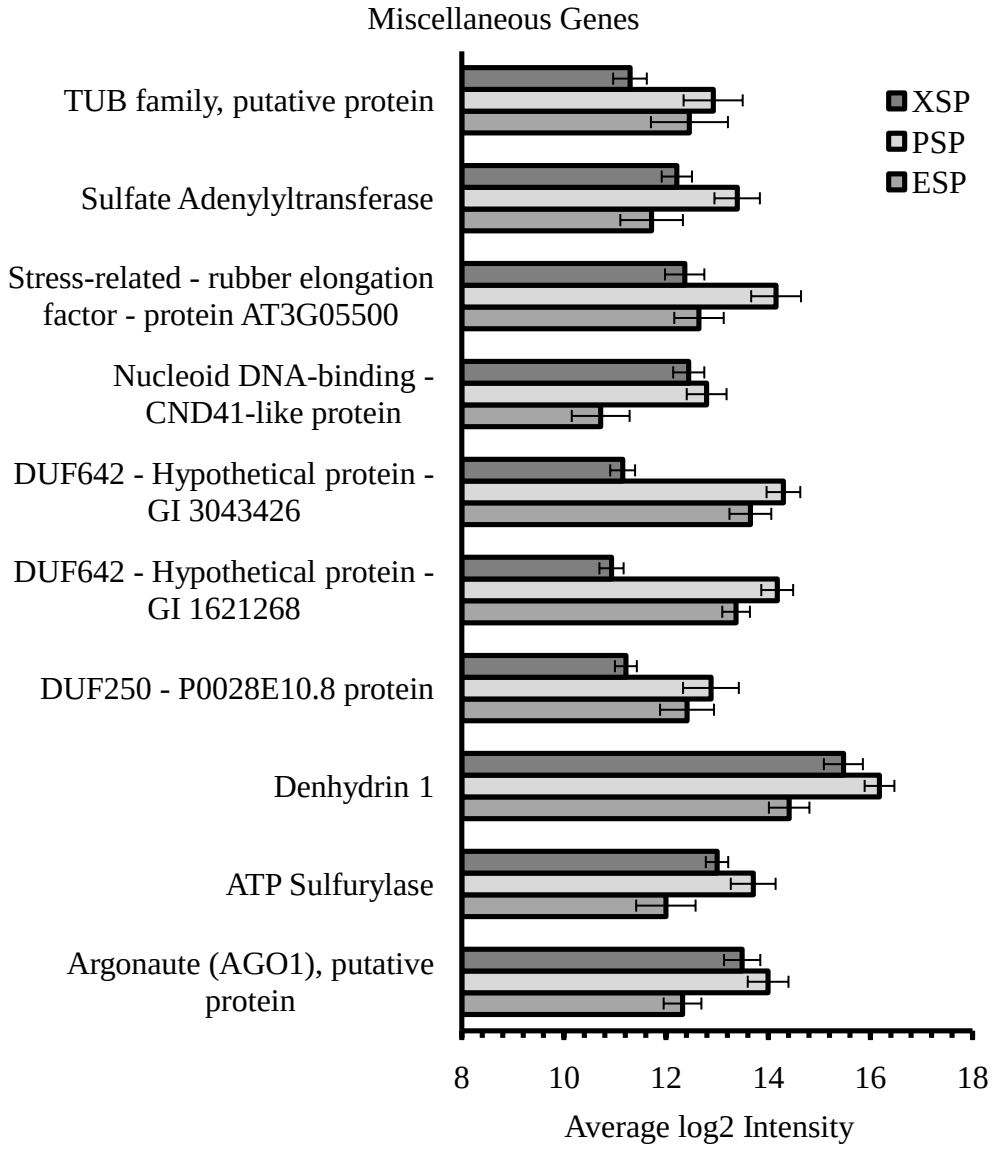


Figure 2-6: Average log2 intensity of selected ESTs of miscellaneous genes showing increased fiber expression in the PEX-S experiment.
Error bars show standard error with an n=28

Twenty-one of the 54 ESTs assigned a functional class of amino acid and protein biosynthesis show up-regulated expression levels in the snap point phloem. Polyubiquitin and a ubiquitin gene showed the differences of more than 1.5 log2 intensity units (Figure 2-7).

Seven of 42 and 5 of 16 ESTs with a functional class of secondary metabolism and phenylpropanoid pathway genes were up regulated in fiber, respectively. Glutathione transferase, S-adenosylmethionine synthase, and S-adenosylmethionine decarboxylase showed the differences of greater than 1.5 log2 intensity units (Figure 2-8).

Thirteen of 52 Oxidoreductases, 13 of 77 regulatory ESTs, and 8 of 67 ESTs involved in transport showed increased expression levels in fiber. The following oxidoreductase and transport ESTs showed differences greater than 1.5 log2 intensities: CYP81E8 - cytochrome P450, cytosolic 6-phosphogluconate dehydrogenase, monodehydroascorbate reductase - seedling isozyme, uricase, MIPC - aquaporin, and a mitochondrial F1 ATP synthase beta subunit (Figure 2-9). None of the regulatory ESTs had a difference in log2 intensity over 1.5.

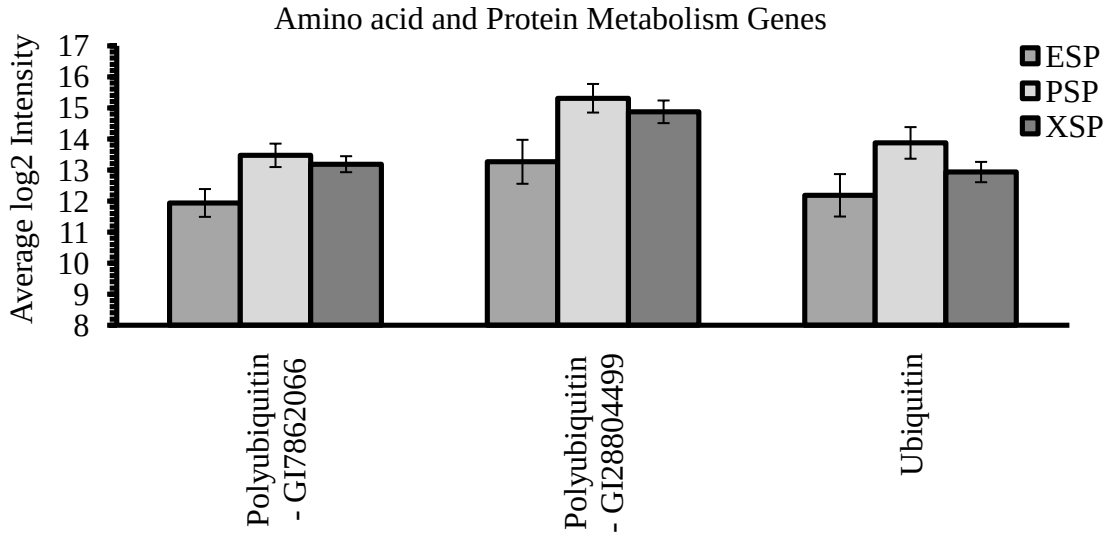


Figure 2-7: Average log2 intensity of selected ESTs of amino acid and protein metabolism genes showing increased fiber expression in the PEX-S experiment. Error bars show standard error with an n=28.

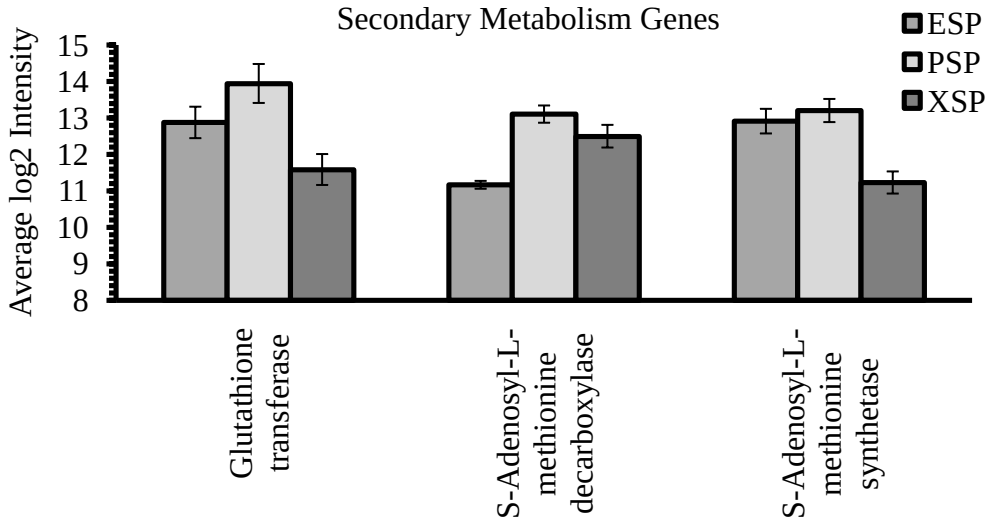


Figure 2-8: Average log2 intensity of selected ESTs of secondary metabolism genes showing increased fiber expression in the PEX-S experiment.
Error bars show standard error with an n=28.

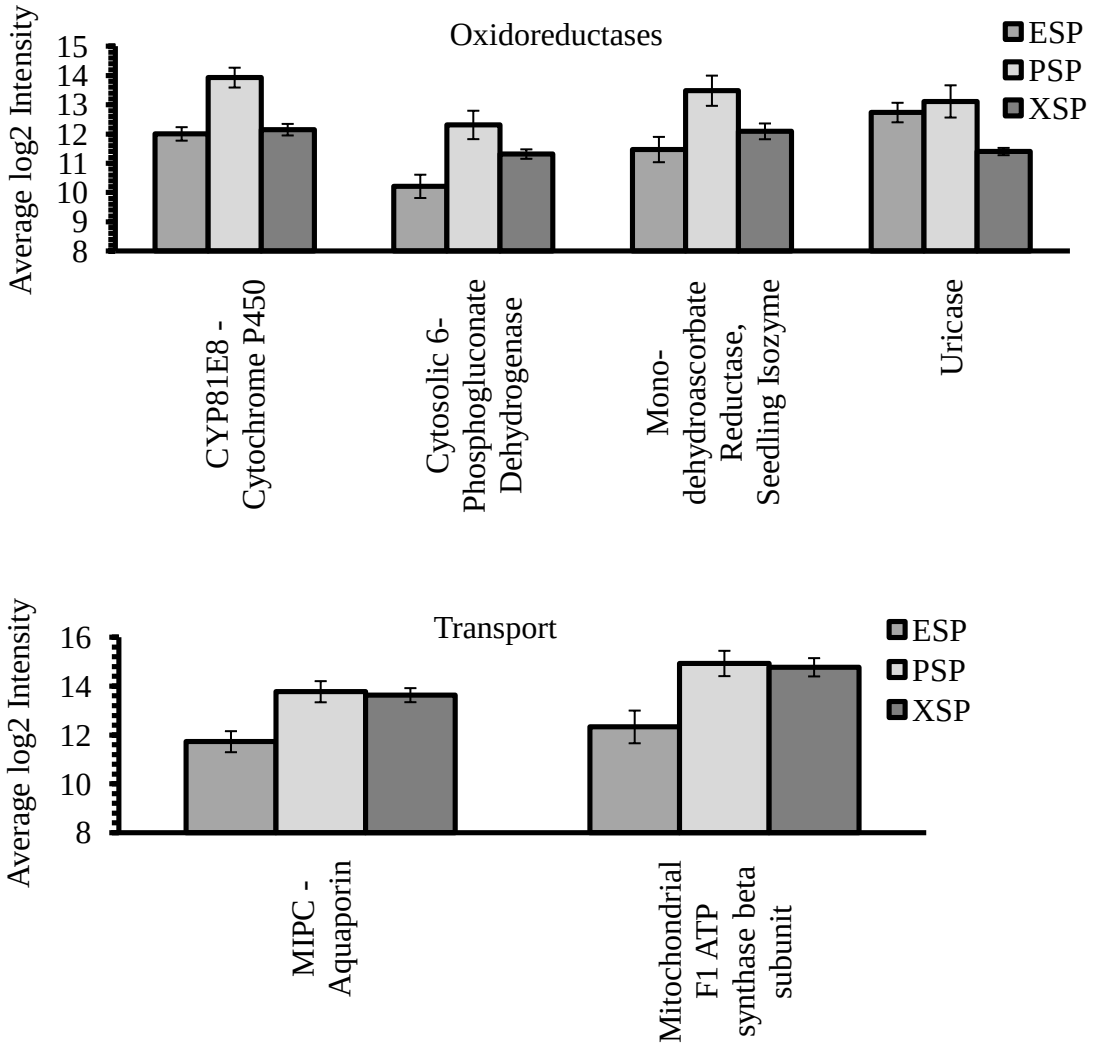


Figure 2-9: Average log2 intensity of selected ESTs of oxidoreductases (above), and transport genes (below) showing increased fiber expression in the PEX-S experiment.
Error bars show standard error with an n=28.

4 of 18 glycosyltransferases, 24 of 47 gycosylhydrolases, and 22 of 72 other cell wall enzymes and proteins assigned functional classes were increased in fiber. The arabinogalactan proteins show some of the most extreme differences of all the proteins examined in these microarrays (Figure 2-10). A specific listing of the ESTs that show 1.5 log2 intensity difference or greater in these groups are: xyloglucan endotransglycosylase, endo-xyloglucan transferase, xyloglucan endotransglycosylase precursor, xyloglucan endotransglucosylase, arabinogalactan protein, hypothetical protein AT5G64430; T12B11.2, and fasciclin-like AGP 15, 11, 6, 5, and 2.

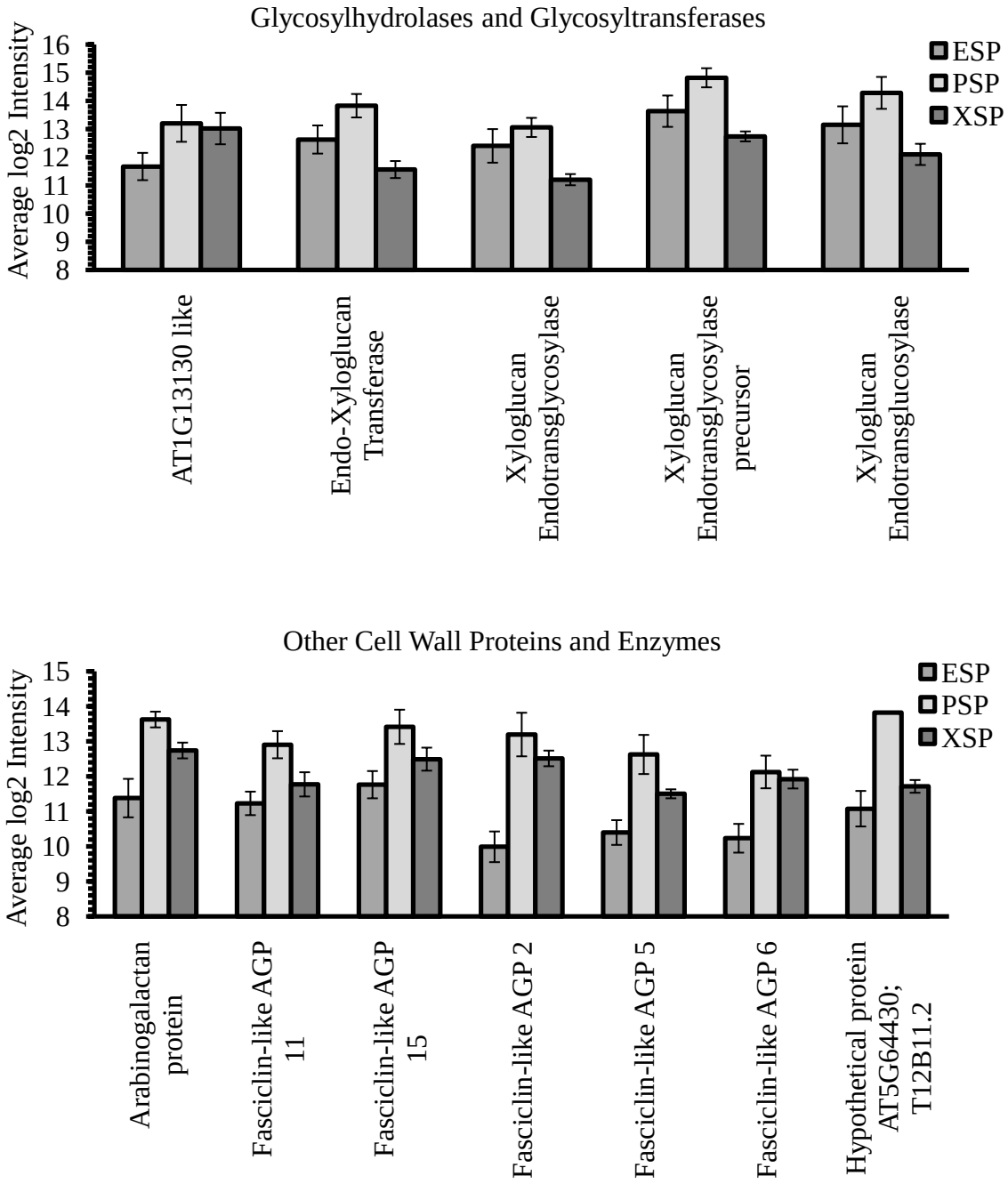


Figure 2-10: Average log2 intensity of selected ESTs of glycosyltransferases and glycosylhydrolases (above), and other cell wall proteins and genes (below) showing increased fiber expression in the PEX-S experiment.
Error bars show standard error with an n=28.

2.3.2 Gene Expression in Snap Point Phloem (Including Bast Fiber) Compared to Phloem From Above and Below the Snap Point (SAB-P)

After an initial SAM analysis a K-means clustering support analysis was performed on the sample, which confirmed that the samples belonged in three groups (each hybridization set and its corresponding dye flip). Of the 7053 independent clones in the fiber-enriched cDNA library microarrays, 3816 showed a significant increase in one of the three sections of phloem above, below, or at snap point from SAM analysis. A further discussion of SAM and clustering analysis is discussed below. Principle component analysis of the significant genes shows three overlapping components, all closely related but distinct (Figure 2-11).

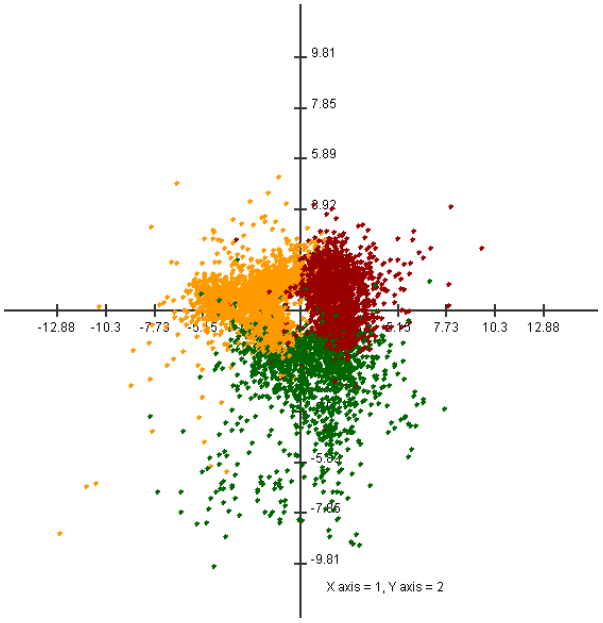


Figure 2-11: PCA analysis of the SAM significant genes at the snap point. Above the snap point (PAS) – yellow, phloem (including bast fiber) at the snap point (PSP) – green, and phloem (including bast fiber) below the snap point(PBS) – red.

Figure of merit analysis determined that 10-12 clusters were optimal for KMC analysis. For each of the twelve clusters of the KMC analysis, the log ratio expression values for all the replicates slides were averaged and plotted to give expression graphs for each cluster. These were organized into three groups: those that showed increased expression at the snap point, below the snap point or above the snap point (Figure 2-12). There are 13 possible combinations for the clusters (Table 2-6), one of which would not show up as statistically significant (PAS=PSP=PBS). All of the combinations were seen in the KMC analysis.

Cluster combinations	Cluster seen in KMC analysis
PAS=PBS>PSP (same as PBS=PAS>PSP)	Yes
PAS=PSP>PBS (same as PSP=PAS>PBS)	Yes
PSP=PBS>PAS (same as PBS=PSP>PAS)	Yes
PAS>PBS>PSP	Yes
PAS>PSP>PBS	Yes
PSP>PBS>PAS	Yes
PSP>PAS>PBS	Yes
PBS>PAS>PSP	Yes
PBS>PSP>PAS	Yes
PAS>PBS=PSP (same as PAS>PSP=PBS)	Yes
PSP>PBS=PAS (same as PSP>PAS=PBS)	Yes
PBS>PAS=PSP (same as PBS>PSP=PAS)	Yes
PBS=PAS=PSP	Not significant

Table 2-6: KMC Cluster combinations for the SAB-P experiment.
Phloem containing bast fiber: above snap point (PAS), at snap point (PSP), and below snap point (PBS).

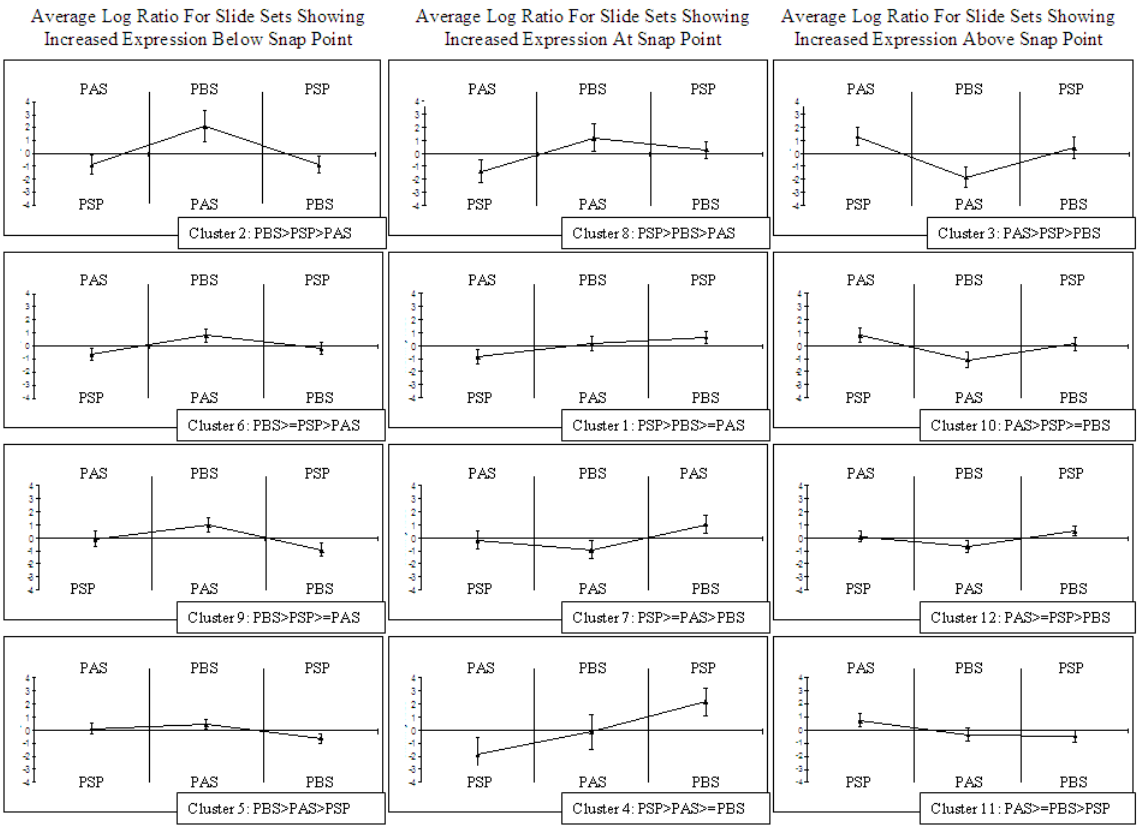


Figure 2-12: SAB-P experiment graphs of cluster analysis showing expression pattern and average log intensity ratios for each cluster.
PAS: phloem layer above snap point, PBS: phloem layer below snap point, PAS: phloem layer at the snap point.

4456 spots, 1087 which corresponded to genes which had function inferred from annotation, were assigned to clusters (Table 2-7). The difference between the total number of spots in the clustering and the total number of significant spots in the SAM analysis is due to replicates that were placed in different clusters during the analysis.

Cluster	Number of Annotated Spots	Number of Un-known Spots	Number of Un-sequenced Spots	Total Number of Spots	Final Class
Cluster 1: PSP>PBS>=PAS	69	12	269	377	PSP
Cluster 2: PBS>PSP>PAS	43	14	95	152	PBS
Cluster 3: PAS>PSP>PBS	67	4	216	287	PAS
Cluster 4: PSP>PAS>PBS	44	7	59	110	PSP
Cluster 5: PBS>PAS>=PSP	151	21	486	658	PBS
Cluster 6: PBS>=PSP>PAS	150	36	394	580	PBS
Cluster 7: PSP>=PAS>PBS	82	8	157	247	PSP
Cluster 8: PSP>PBS>PAS	62	12	154	228	PSP
Cluster 9: PBS>PSP>=PAS	93	16	326	435	PBS
Cluster 10: PAS>PSP>=PBS	71	21	276	368	PAS
Cluster 11: PAS>=PBS>PSP	74	18	286	378	PAS
Cluster 12:PAS>=PSP>PBS	181	37	418	636	PAS
Totals	1087	206	3163	4456	

Table 2-7: Numbers of spots in each cluster and its class designation for the SAB-P experimental set.

The same method used to calculate the fold difference and determine the tissue which had the highest intensity compared to the other two tissues for the PEX-S experiment was used for the SAB-P experiment (Table 2-8). The number of annotated sets of duplicate spots for each tissue is PAS - 247, PSP - 447, and PBS – 259, for a total of 953. The annotated tissue sets were then divided into functional classes based on the classes determined in de Pauw 2007 microarray experiments and a graph of the percentage of each functional class for each tissue type was created (Figure 2-13).

Of the 953 annotated genes 41 have no known function. Duplicate listings of contigs were deleted. PAS showed 170 genes, PSP 127, and PBS 122. See Appendix 4 for Gene Listings.

Functional Class	PAS	PSP	PBS	Totals
Photosynthesis - chloroplast	45	66	5	116
Glycolysis & respiration	5	4	8	17
Amino acid & protein metabolism	10	39	31	80
Secondary metabolism	16	17	14	47
Oxidoreductases	17	45	16	78
Regulation	20	35	51	106
Transport	14	45	34	93
Glycosyltransferases	7	8	2	17
Glycosylhydrolases	12	40	6	58
Other cell wall enzymes	0	3	2	5
Other cell wall proteins	17	30	22	69
Miscellaneous	50	97	59	206
Unknown function	23	12	6	41
Phenylpropanoid biosynthesis	11	6	3	20
Totals	247	447	259	953
Total Including non-annotated genes	1556	898	1663	4117

Table 2-8: Number of genes from the SAB-P experiment in each functional class, based on a pair-wise comparison of the average intensity for significant genes found by SAM analysis.

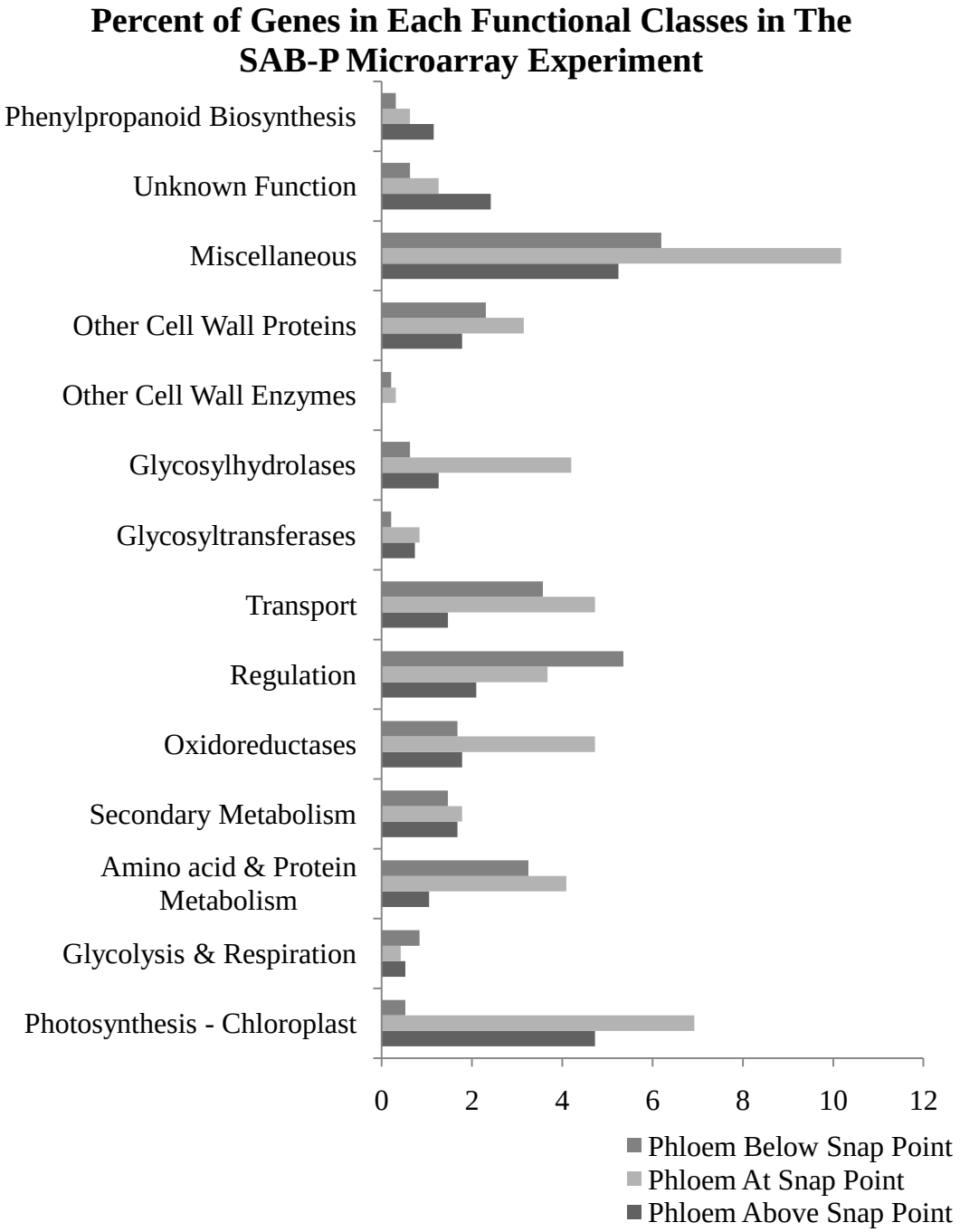


Figure 2-13: Percentage of total number of annotated genes for each functional class in the SAB-P microarray experiment.

Genes which did not fall into any of the other functional classes were placed in a miscellaneous class. Of the 206 genes in the miscellaneous functional class determined to be significant via SAM analysis, 97 were up regulated in fiber. These are: Aluminum-induced protein, two ATP sulfurylases, AUX/IAA protein, dehydrin 1, desiccation-related protein, DHN1 protein, DUF231 hypothetical protein, DUF250 - P0028e10.8 protein, three DUF642 - hypothetical proteins, DUF760 - hypothetical protein, flagellar L-ring protein precursor - lipoprotein, hydrolase - alpha/beta fold family protein, hydrophobic protein RCI2A - UPF0057, hypothetical protein - DOXX, photoassimilate-responsive protein PAR-1B-like protein, RSH3, stress-related - rubber elongation factor - protein, wound-induced WI12. All these genes show increases of 1.5 log2 intensity units or more in the Snap Point Fiber (Figure 2-14).

The miscellaneous genes that are seen in SAB-P expression but not in PEX-S expression are DUF231, DUF760, DUF642, hydrophobic protein RCI2A, AUX/IAA protein, photoassimilate-responsive protein PAR-1B-like protein, armadillo-like helical protein, DOXX, DHN1 protein, dehydrin 1, desiccation-related protein, wound-induced WI12, aluminum-induced protein, RSH3, alpha/beta fold family hydrolase, flagellar l-ring protein precursor - lipoprotein. Those seen in PEX-S expression and not SAB-P expression are sulfate adenylyltransferase, nucleoid DNA-binding - CND41-like protein, argonaute (AGO1) protein, and a TUB family protein.

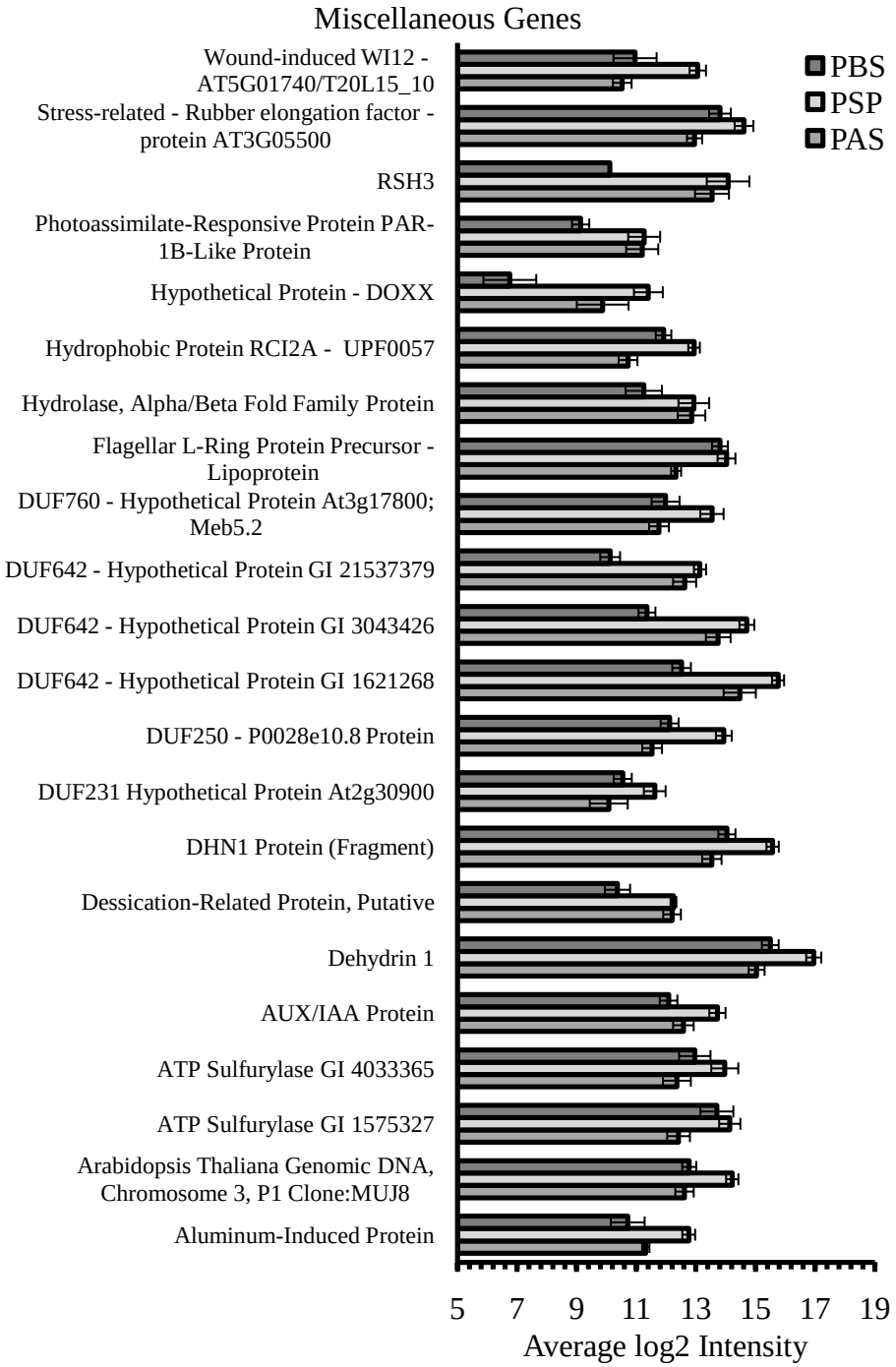


Figure 2-14: Average log2 intensity of selected ESTs of miscellaneous genes showing increased fiber expression in the SAB-P experiment.
Error bars show standard error with an n=16.

Of the total 116 photosynthesis genes and 17 glycolysis and respiration genes that were assigned a functional class, 66 photosynthesis and 4 glycolysis and respiration genes showed increased expression at the snap point. The genes that showed a greater than 1.5 log2 intensity unit (Figure 2-15, Top) increase in expression at the snap point are the hypothetical protein AT2G47860, ribulose 1,5-bisphosphate carboxylase small subunit precursor, ribulose-1,5-bisphosphate carboxylase small subunit rbcS1 , alcohol dehydrogenase, and glyceraldehyde-3-phosphate dehydrogenase A subunit.

39 of a total 80 amino acid metabolism genes were increased in the fiber at the snap point compared to the fiber/phloem above or below it. Only four of them show a difference of 1.5 log2 expression level or more (Figure 2-15, Bottom): mitochondrial processing peptidase alpha subunit, polyubiquitin, ubiquitin homolog, and vacuolar processing enzyme precursor.

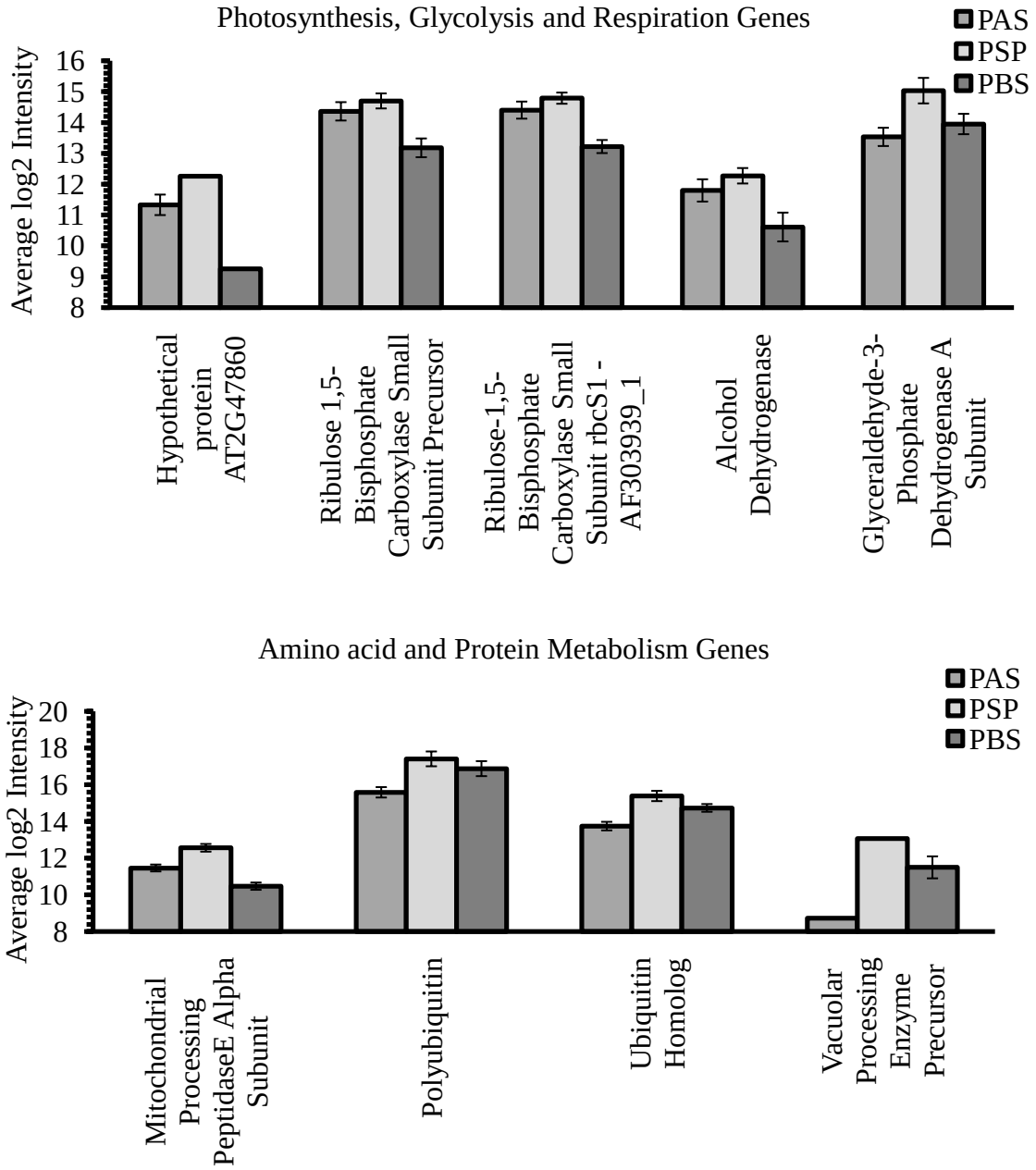


Figure 2-15: Average log2 intensity of selected ESTs of photosynthesis, glycolysis, and respiration genes (Top); and amino acid and protein metabolism genes (Bottom) showing increased fiber expression in the SAB-P experiment. Error bars show standard error with an n=16.

A total 67 secondary metabolism genes and 78 oxidoreductases assigned functional classes were determined by SAM analysis to be significant, of these 23 secondary metabolism genes, and 45 oxidoreductases are increased in snap point phloem. The genes involved in secondary metabolism that have a expression change greater than 1.5 log2 units (Figure 2-16, Top) are glutathione transferase, S-adenosylmethionine decarboxylase, cinnamate-4-hydroxylase, and cinnamoyl-CoA reductase. The oxidoreductases that show expression differences greater than 1.5 log2 intensity units (Figure 2-16, Bottom) between the snap point and other tissues are basic blue copper protein, catalase, cytochrome P450, peroxidase, peroxidase precursor, aldehyde dehydrogenase, and quinone oxidoreductase.

For regulatory and transport genes assigned a functional class, 106 and 93 genes showing significant levels of expression respectively. 35 regulatory and 45 transport genes are increased in phloem, but only 4 regulatory genes (GID1-like protein 1, leucine rich repeat Protein Precursor, transcription factor LIM1, and EIN3-binding F-Box protein) and 5 transport genes (carrot root specific aquaporin, potassium transporter ATKT2P,ferrous ion membrane transport protein DMT1, OSTIP2.1 aquaporin, and sulfate ABC transporter - periplasmic sulfate-binding protein) have an increase of greater than 1.5 log2 units (Figure 2-17).

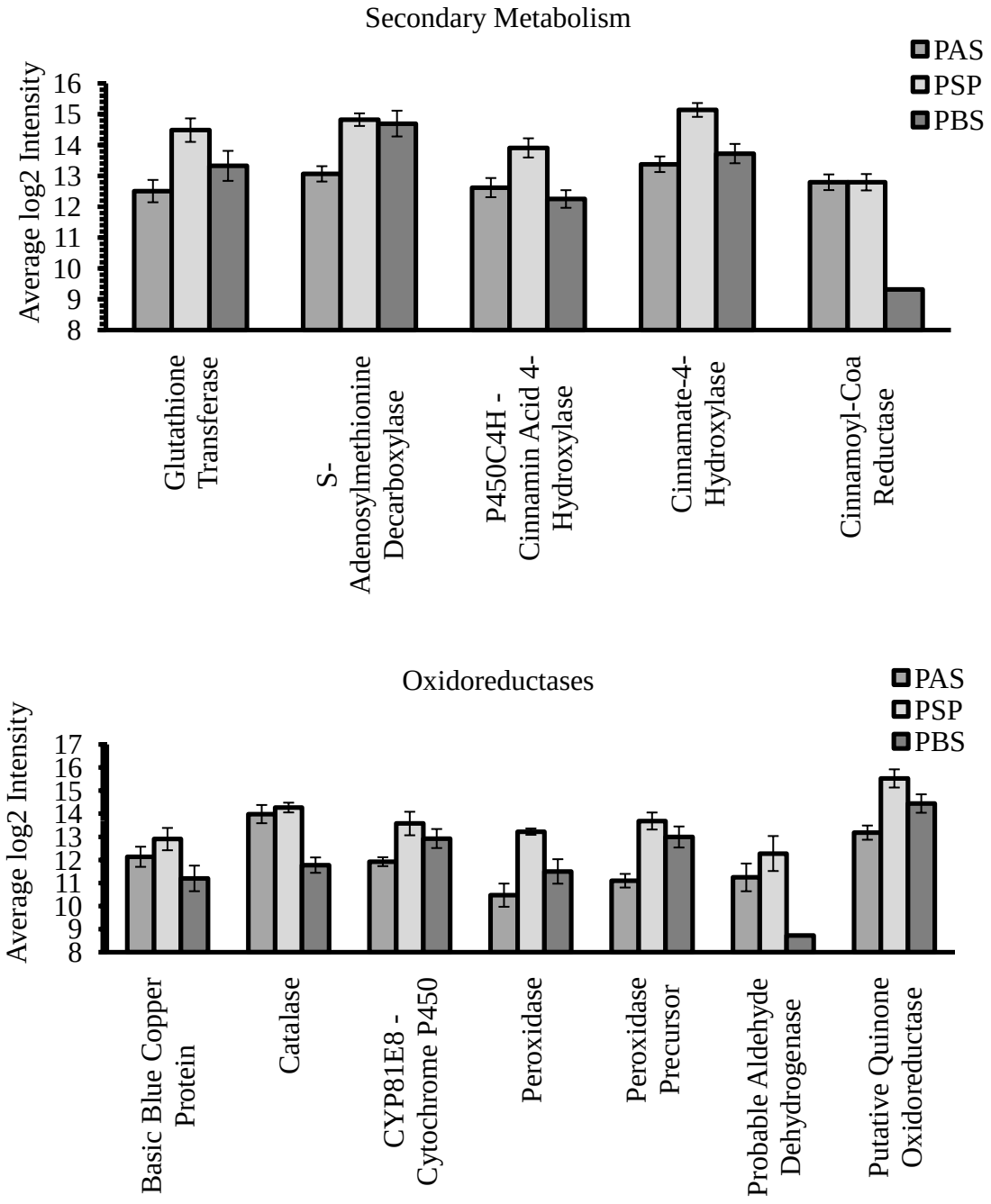


Figure 2-16: Average log2 intensity of selected ESTs of secondary metabolism (Top); and oxidoreductases (Bottom) showing increased fiber expression in the SAB-P experiment.
Error bars show standard error with an n=16.

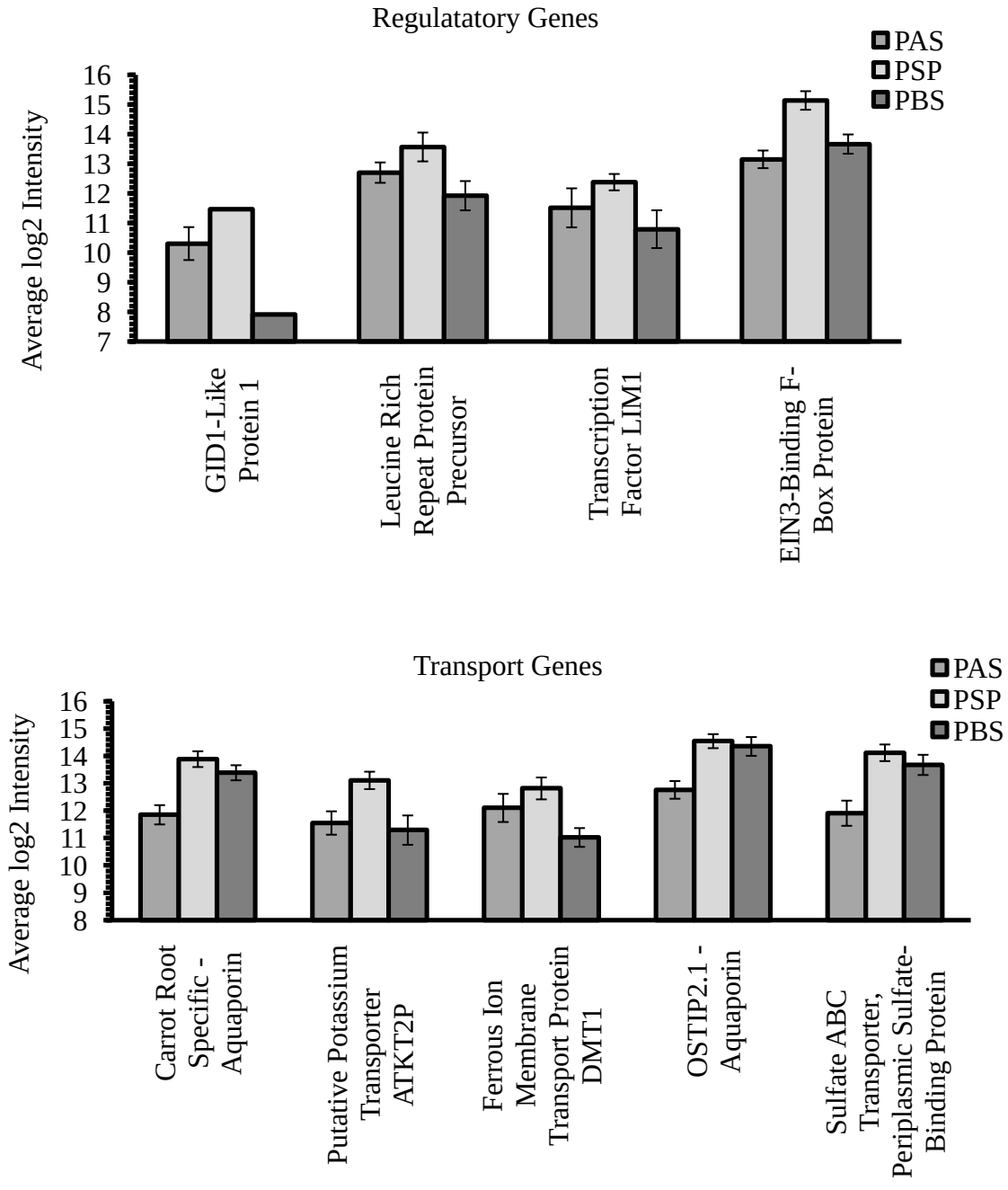


Figure 2-17: Average log2 intensity of selected ESTs of regulatory (Top) and transport genes (Bottom) showing increased fiber expression in the SAB-P experiment.
Error bars show standard error with an n=16.

75 glycosyltransferases and gylcosylhydrolases were found significant in the phloem above, below, and at the snap point using SAM analysis; of these 48 genes showed increased expression at the snap point. Only (1-4)-beta-mannan endohydrolase-like protein, chitinase-like protein, endo-xyloglucan transferase, pectinacetylesterase precursor, pectinesterase-3 precursor, xyloglucan endotransglucosylase-hydrolase XTH7, xyloglucan endotransglycosylase, xyloglucan endotransglycosylase precursor, UDP-glycosyltransferase 89B2, and xyloglucan endotransglucosylase gave a expression level with a difference of above 1.5 log2 expression units (Figure 2-18, Top).

For the remaining cell wall proteins and enzymes 33 of the total 74 genes assigned to the functional class showed increased expression levels in the phloem of the snap point. Those that show the largest differences in expression level are: Phytochelatin Synthetase-Like Protein, Cell-Wall P8 Protein, Pectinesterase, At4g14500, Lipid Transfer Protein Isoform 4, Arabinogalactan Protein, Pectin Acetylesterase, Fasciclin-Like AGP 5 and AGP 15 (Figure 2-18, Bottom).

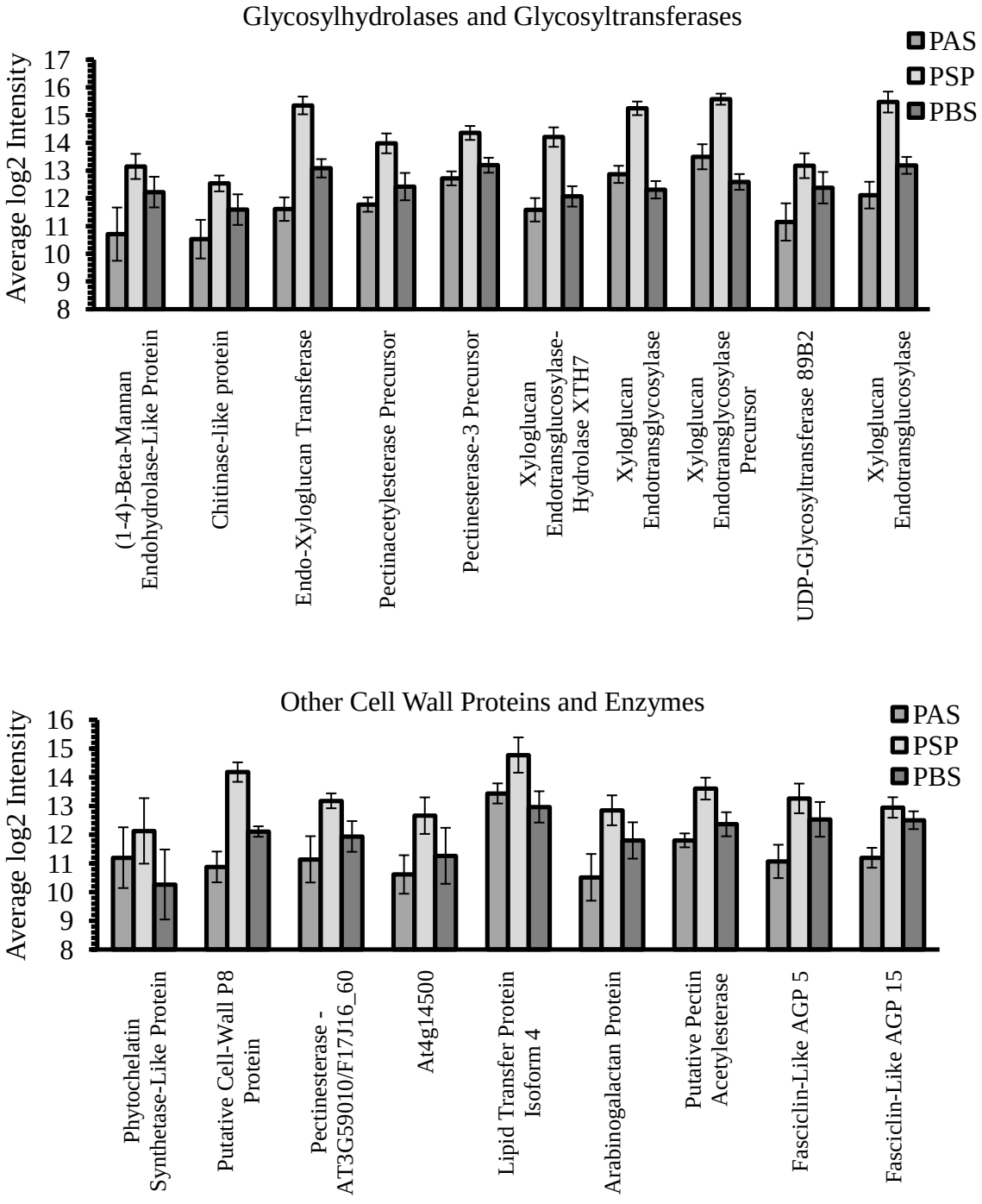


Figure 2-18: Average log2 intensity of selected ESTs of glycosylhydrolases and glycosyltransferases (Top), and other cell wall proteins and enzymes (Bottom) showing increased fiber expression in the SAB-P experiment.
Error bars show standard error with an n=16.

The genes which have a expression level of greater than 1.5 log2 units in both the PEX-S and SAB-P experiments are polyubiquitin, a ubiquitin homolog, xyloglucan endotransglycosylase precursor, xyloglucan endotransglycosylase, endo-xyloglucan transferase, xyloglucan endotransglucosylase, a stress-related - rubber elongation factor, dehydrin 1, arabinogalactan protein, fasciclin-like AGP 15, CYP81e8 - cytochrome p450, glutathione transferase, DUF250, and two accessions of DUF642.

2.4 Discussion

Entrez Gene suggested further functional roles for each of the genes found in both microarray studies (NCBI 2009) (Appendix 5 – NCBI Conserved Domain Database and Entrez Gene References for Function of Genes Found in Microarray Analysis. The genes found to be up-regulated in the SAB-P expression study were involved in: energy production (mitochondrial F1 ATP synthase beta subunit), reorganization and creation of cell walls (xyloglucan endotransglycosylase, xyloglucan endotransglucosylase, and cytosolic 6-phosphogluconate dehydrogenase); the creation, cleavage and shuffling of proteins (S-adenosylmethionine decarboxylase, cytochrome P450, glutathione transferase, ATP Sulfurylase, polyubiquitin, and ubiquitin), cell differentiation (arabinogalactan protein) and polymerization (TUB family proteins) (NCBI 2009). There are also a number of genes implicated in gene silencing, senescence and cell death (CND41-like protein, and Argonaute (AGO1) protein) (NCBI

2009), which could be from the maturing fiber near the lower portion of the snap point. The octicosapeptide/Phox/Bem1p (PB1) domain-containing protein (hypothetical protein AT5G64430) has been implicated in protein-protein interactions in a cell polarity manner (NCBI 2009); which suggests a cell changing at its ends, such as what may occur in fiber growth. The fasciclins (fasciclin-like AGP 15, 11, 6, 5, and 2) are involved in cell to cell adhesion (NCBI 2009) and could be involved in the fiber cells sticking together as they encounter each other during growth. There are also a number of genes that are involved in desiccation (NCBI 2009) which could be due to the stress of the dissection process or due to water being moved out of maturing fiber as the cytoplasm is occluded (DUF250, dehydrin 1, aquaporin, and stress-related - rubber elongation factor protein). Monodehydroascorbate reductase and uricase protects against reactive oxygen species damage as do peroxidases, which have been implicated in lignin degradation (NCBI 2009). DUF 642 plays an unknown function; and decarboxylase could play many different possible roles dependent on the specific type.

The comparison of the expression in snap point node fiber to the fibers or phloem in the nodes above and below (the SAB-P study) gave more genes then in the PEX-S study; but many of those seen in the PEX-S study are also increased in the SAB-P study (ATP Sulfurylase, xyloglucan endotransglycosylase, xyloglucan endotransglucosylase, S-adenosylmethionine decarboxylase, cytochrome P450,

glutathione transferase, polyubiquitin, ubiquitin, arabinogalactan protein, fasciclin-like AGP 15, 11, 6, 5, and 2, DUF 250, dehydrin 1, aquaporin, stress-related - rubber elongation factor protein, and DUF 642) . Like the PEX-S study many of the new genes are consistent with the restructuring of cell walls and potential changes in cell walls during intrusive growth. The genes which have similar types of functions to those seen in the PEX-S study are energy production (quinone oxidoreductase, and rbcS1), reorganization and creation of cell walls (cell-wall P8 protein, phytochelatin synthetase-like protein, and UDP-glycosyltransferase), the cleavage and shuffling of proteins (leucine rich repeat Protein, hypothetical protein AT2G47860, and mitochondrial processing peptidase alpha subunit, glyceraldehyde-3-phosphate dehydrogenase A subunit), cell differentiation (transcription factor LIM1), and senescence (EIN3-binding F-Box protein) (NCBI 2009). Genes which show functions in the SAB-P study which are not seen in the PEX-S study are cell wall breakdown (pectin acetylesterase, (1-4)-beta-mannan endohydrolase-like protein, chitinase-like protein, aluminum-induced protein, and pectinesterase), oxidative waste protection (peroxidase, and catalase) all of which may be involved in lignin degradation (NCBI 2009). Lignin degradation in cells where an increase in lignin is expected may either be due to the stress of the dissection or part of the function that allows the fibers to grow and pass each other. Phenylpropanoid genes (alcohol dehydrogenase, cinnamate-4-hydroxylase, cinnamoyl-CoA reductase, and aldehyde dehydrogenase), and wound-induced WI12 protein which has been

reported to act in cell wall reinforcement (NCBI 2009) are directly responsible for the growth of the fiber cells. There is also a number of lipid, ion, water, and waste transport genes expressed (lipid transfer protein isoform 4, At4g14500, DMT1, potassium transporter ATKT2P, hydrophobic protein RCI2A, periplasmic sulfate-binding protein, vacuolar processing enzyme, PAR-1B-like protein, desiccation-related protein, and TIP2.1 aquaporin) (NCBI 2009) which may be due to the stress of the dissection process, or occlusion of the cytoplasm in the thickening cell walls of the bast fibers. There are two inducible genes which do not seem related to fiber differences (AUX/IAA protein and GID1-like protein 1). AUX/IAA protein is a rapidly induced auxin responsive gene; and GID1-like protein 1 is thought to be a gibberellin receptor (NCBI 2009). Genes with unknown functions (hydrolase - alpha/beta fold family protein, DUF231, DUF760, DOXX, PAR1, and basic blue copper protein) are also seen.

qRT-PCR of the phenylpropanoid genes and the various transcription factors will help further define their activity in the snap point fiber, as would the examination of the various lignin degrading genes. Wound-induced WI12 protein and its role in cell wall reinforcement may have some commercial applications as well – perhaps giving stronger or thicker fibers.

Chapter 3

Tissue-Specific Transcript Abundance of Selected Cell Wall Genes

3 Tissue-Specific Transcript Abundance of Selected Cell Wall Genes

3.1 Introduction

Lignification of various tissues in hemp stems is of potential interest in applied research, because the quality and quantity of lignin affects such properties as fiber strength, hydrophobicity, ease of forage digestion, ease of pulping, and biofuel production. Biochemical and histochemical analyses provide convincing evidence of lignin and its phenylpropanoid subunits within hemp stems ((Crônier, Monties and Chabbert 2005), (Blake *et al.* 2008), (de Pauw *et al.* 2007), (van den Broeck *et al.* 2008)). Yet, previous microarray studies (de Pauw *et al.* 2007), and northern blots (van den Broeck *et al.* 2008) of gene expression in hemp fibers failed to show enrichment of several genes from the phenylpropanoid pathway. This may be because the microarrays used were not comprehensive. Additionally, these microarrays only identified genes that differed in abundance between tissues, and may not have been able to distinguish between closely related genes from different tissues. Therefore, the more precise technique of quantitative real-time PCR (qRT-PCR) was used to analyze tissue-specific transcript abundance of phenylpropanoid pathway and cell wall related genes in hemp stems. My objective was to identify genes in these pathways that were highly enriched in only one type of fiber (i.e. either xylem or phloem). The identification of these tissue-specific genes will allow for possible modification of lignin content and cell wall properties through Alberta Research Council's hemp reverse genetics platform.

Identifying genes in hemp that may be targeted for mutation to modify its lignin content and cell wall structure is currently hampered by the relatively small numbers of ESTs and other sequences published for *Cannabis* (3961nucleotide sequences in the NCBI database). The majority of the 241 sequences not categorized as either ESTs or proteins are markers and microsatellite data for distinguishing drug varieties from fiber varieties. Of the 3746 ESTs, 1565 of them are from a trichome study done by the University of Minnesota, with the remaining EST sequences from two groups researching fiber: 606 from the Deyholos labs library, and 1575 from the van den Broeck study.

When analyzing related genes, qRT-PCR is a more accurate tool than cDNA microarrays and better suited to identifying expression of specific genes in large families because it uses smaller, more specific areas of the genes; so may be able to answer if the fiber and xylem genes have different expression levels not seen in the microarray. We can also examine the changes that occur in phenylpropanoid pathway activity in fiber as the stem ages using qRT-PCR. This will give us a set of candidate genes in hemp to examine the modification of the quantity and quality of hemp fiber.

The microarray work done in the earlier chapter of this thesis did not show large changes in genetic expression of the phenylpropanoid pathway genes in hemp,

with the exception of C4H and CCR in the PEX-S experiment. However, this may be due to the cross hybridization with members of gene families that are very similar, causing false negatives in the microarray. Most of the phenylpropanoid genes are members of gene families, for example 4CL has 5 known members in *Arabidopsis* which are active in different tissues (Costa *et al.* 2005), and COMT has one gene and 13 COMT-like genes in Arabidosis (Raes *et al.* 2003), but a quick examination of the NCBI database gives 6 OMT genes in Hops. C4H has one isomer and CCR has two isomers and five CCR-like isomers in *Arabidopsis* (Raes *et al.* 2003) which may account for an increase in their expression in the microarray experiment. With this in mind a set of quantitative real-time PCR experiments were designed to look at each EST of interest in the microarray library using the same tissues as in the prior microarray work.

The hemp ESTs in the phenylpropanoid pathway (**Error! Reference source not ound.**) found in the microarray library and in the NCBI gene bank is p-coumaroyl shikimate 3'-hydroxylase (PCS3H), phenylalanine ammonia-lyase (PAL), cinnamate-4-hydroxylase (C4H), cinnamoyl CoA reductase (CCR), ferulate 5-hydroxylase (F5H), caffeic acid O-methyltransferase (COMT), caffeoyl-CoA O-methyltransferase (CCoAOMT), 4-coumarate-CoA ligase (4CL) (Figure 1-3).

The genes for cellulose synthase (CESA), and beta galactosidase (BGAL) involved in the synthesis of cell walls have been identified as having increased

expression in both hemp and flax in other microarray experiments ((de Pauw *et al.* 2007), (Roach and Deyholos 2007)) but, like the phenylpropanoid pathway genes, did not show significant changes in their expression patterns in the previous chapter's experiments.

3.2 Methods

3.2.1 Tissue Dissection

C. sativa genotype Carmen tissues were dissected (Table 2-1: Table of dissected tissue type and experiment abbreviations (Table 2-1) and (Figure 2-1) from three pools of five to six five-week old plants in the same manner as was done for the microarray experiment (2.2.1 Tissue Dissection).

3.2.2 Primer Design

All hemp accessions for the genes of interest from Genbank and Deyholos library were compared using ClustalW and 30 primer sets were made based on those alignments (see Appendix 6). After testing the suitability of the primer sets for qRT-PCR the ESTs were aligned to complete gene accessions from Genbank via Sequencher^{TM} version 4.9(Gene Codes Corporation, Inc. 1991-2009) to determine possible duplications within the ESTs.

3.2.3 RNA Extraction

RNA was extracted from the various tissues using Qiagen's RNeasy Mini Plant Kit modified by adding 50 µL of 20% high molecular weight PEG, and 10 µL Beta mercaptoethanol per 1 mL of RLC buffer. All the samples were then treated with DNase using the TURBO DNA-free Kit from Ambion.

3.2.4 Quantitative Real-Time PCR (qRT-PCR) and Analysis

0.1 µg of total RNA was amplified using Invitrogen's SuperScript™ III Platinum® SYBR® Green Two-Step qRT-PCR Kit with ROX and the samples were analyzed via qRT-PCR in two different trial sets matching those done in the microarray study (PEX-S, and SAB-P). The 96 well plates for each primer set and trial set were aliquoted using an Eppendorf EP motion robot. All genes were cycled on an Eppendorf Realplex using the following program (50°C, 2min; 95°C, 2min; (95°C, 15sec; 60°C, 30sec) X40, with a fluorescence reading taken each cycle. Then a melting point analysis was done using the following parameters 95°C, 15sec; 60°C, 15sec; ramped up to 95°C over 20 minutes with continuous fluorescence readings. Tubulin was run as the reference gene for every plate. All plates were analyzed using the relative standard curve method outlined in the Applied Biosystems real-time PCR systems chemistry guide (Applied Biosystems 2004), and a standard error was calculated with an n= 9.

3.2.5 Alignments and Trees

To determine which ESTs aligned, all the hemp EST accessions for a particular isozyme were aligned in Sequencher™ version 4.9(Gene Codes Corporation, Inc. 1991-2009) with a minimum match of 90%. Then, to find discontiguous sequences, the closest BLASTn hit for each EST was imported as guides and all the entries were re-aligned with a with a minimum match of 60%. Once a set of contigs for each isozyme was determined, it was manually examined and any widely variant EST alignments were removed and checked if they aligned better elsewhere; if they did not they were placed on their own. Contigs sequences were created for each alignment that had contiguous alignments by removing the guide sequence. These contigs were used to create a series of trees for each gene of interest.

For each tree, a recent paper describing the gene family in *Arabidopsis* or other model plants was found and the Genbank accessions used were imported into MEGA (Tamura *et al.* 2007), along with the accessions for the highest NCBI BLASTn hits , and the contigs determined in Sequencher™. Then all sequences for a MEGA alignment were manually examined as protein sequences and cropped to the length of the smallest contig, these were then aligned in MEGA by CLUSTAL W, and a tree was made through MEGA using the maximum parsimony (MP) method (Eck and Dayhoff 1966). A bootstrap consensus tree was inferred from 500 replicates with the percentage of replicate trees in which the

taxa clustered together for a branch shown next to the branch (Felsenstein 1985). The initial MP trees were obtained with the random addition of sequences (10 replicates) and the final tree determined from them using the Close-Neighbor-Interchange algorithm (Nei and Kumar 2000) with search level 3.

3.3 Results

To provide a further examination of the expression pattern of secondary cell wall development and lignin biosynthesis in *C. sativa* genotype Carmen, two sets of qRT-PCR analysis were run to examine ten genes of interest (PCS3H, PAL, C4H, CCR, F5H, COMT, CCoAOMT, 4CL, CESA, and BGAL). Currently the genome of *Cannabis* is not sequenced and there is very little information about the number of proteins in each family for the genes being examined, so an alignment of the available *Cannabis* ESTs and a phylogenetic tree to identify the isozymes were produced to aid in interpreting the qRT-PCR analysis results.

Bast fibers begin to undergo secondary wall thickening in a region of the stem known as the snap-point. The snap-point therefore served as a focus for dissections in preparation for analysis of tissue-specific expression patterns. Stems were dissected radially, separating the internode that contained the snap-point into epidermis (ESP), phloem and bast fiber bearing (PSP), and xylem bearing (BSP) tissues. Phloem, including bast fiber bearing tissues, from above

the snap-point (PAS) and from below the snap point (PBS) was also included in these dissections (Table 2-1, Figure 2-1, and Figure 2-2).

To identify transcripts of genes required for lignin deposition and cell wall development specifically in either xylem or phloem fibers of developing hemp plants, I extracted sequences from EST databases for selected phenylpropanoid pathway genes, as well as CesA family members and beta-galactosidase from Genbank (Table 3-1).

Multiple alignments for the ESTs corresponding to each gene were produced, in some cases using as a guide the most similar full-length cDNA of other species that I obtained from NCBI (i.e. highest BLASTn hit). For some enzymes, the corresponding ESTs aligned in distinct clusters, indicating that the enzyme was encoded by more than one gene in hemp. I will refer to these distinct clusters as isoforms. At least 1 primer pair was synthesized for each isoform of each enzyme (Table 3-1).

In cases where there were gaps in alignments, primers from two discontigous ESTs were mixed and the resultant PCR product from *Cannabis sativa* genomic DNA was sequenced to confirm that the ESTs were sections of the same gene (Coffey 2008). The primer sets were also used in rolling circle amplification to determine the full length sequence for the mRNA of the gene (Voicu 2009).

Gene	Gene Name	Accession Number	qRT-PCR Primer Set
C4H	Cinnamate-4-hydroxylase	gi|110190212|gb|EC855348.1|EC855348 EST00086	c4h1
CCR	Cinnamoyl CoA reductase	gi|110190257|gb|EC855393.1|EC855393 EST00131	ccr1
CCR	Cinnamoyl CoA reductase	gi|110190225|gb|EC855361.1|EC855361 EST00099	ccr2
CCR	Cinnamoyl CoA reductase	lcl|MD|02_HEMPSE_RP_011_A12_30MAR2006_096	ccr1
F5H	Ferulate 5-hydroxylase	lcl|MD|HEMPSEQ_RP_006_B08_16DEC2005_062	f5h2
COMT	Caffeic acid O-methyltransferase	lcl|MD|02_HEMPSE_RP_016_G07_30MAR2006_051	comt2
4CL	4-coumarate-CoA ligase	lcl|MD|02_HEMPSE_RP_019_G05_01APR2006_035	4cl1
4CL	4-coumarate-CoA ligase	gi|110190230|gb|EC855366.1|EC855366 EST00104	4cl5
4CL	4-coumarate-CoA ligase	gi|110190231|gb|EC855367.1|EC855367 EST00105	4cl1
CESA	Cellulose synthase	gi|110190255|gb|EC855391.1|EC855391 EST00129	cesa1
CESA	Cellulose synthase	gi|110190253|gb|EC855389.1|EC855389 EST00127	cesa2
CESA	Cellulose synthase	gi|110190245|gb|EC855381.1|EC855381 EST00119	cesa4
CESA	Cellulose synthase	02_HEMPSE_RP_011_H03_30MAR2006_017	cesa1
BGAL	beta-D-galactosidase	HEMPSEQ_RP_007_F02_06FEB2006_006	bgal1

Table 3-1: Accessions of genes examined in qRT-PCR studies PEX-S and SAB-P. A more complete list of all accessions and primer sets is in Appendix 7 – Primer Sets and *C. sativa* EST Accessions Examined for qRT-PCR

.

Primers were screened for increased activity in fiber via qRT-PCR, and those showing an increase in snap point fiber were used for further testing. Eleven primer pairs (4cl1, 4cl5, c4h1, ccr1, ccr2, cesa1, cesa2, cesa4, comt2, f5h2, and bgal1) were used to measure transcript abundance in each of three biologically independent pools of tissue for two sets of experiments similar to the microarray study: phloem including bast fiber from above, below and at the snap point (SAB-P) and epidermis, xylem, and phloem including bast fiber from the snap point (PEX-S).

qRT-PCR was attempted on PCS3H, PAL, and CCoAOMT. Initial screens of PCS3H and CCoAOMT primer sets gave poor qRT-PCR transcription levels (data not shown), and the EST sequences (one for each gene) were not long or complex enough to design more primer sets. PAL, which has seven EST accessions, gave low qRT-PCR transcription levels during initial screening in the snap-point phloem compared to the other candidate genes, and was not examined further (data not shown).

3.3.1 Cinnamate-4-hydroxylase (C4H)

Five EST sequences were found for Cinnamate-4-hydroxylase (C4H); which gave one contiguous sequence when aligned with *Malus domestica* mRNA (apple) (Figure 3-1).

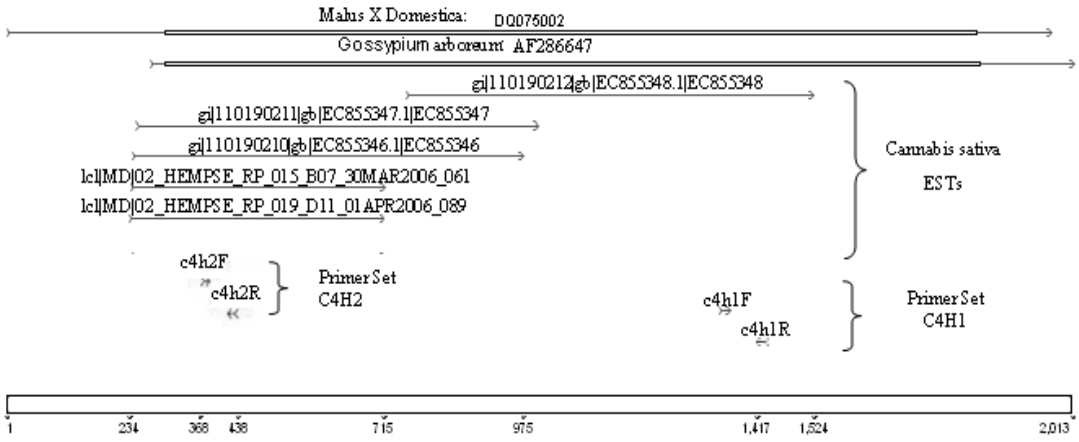


Figure 3-1: Alignment of *Cannabis sativa* ESTs for C4H, and the qRT-PCR primer sets designed for it, against the full length mRNA from nearest BLAST hits to the ESTs.

Initial qRT-PCR screening confirmed a single C4H gene which showed significantly increased expression levels in snap point fiber, and both primer sets showed similar expression patterns. Further qRT-PCR showed an increase in expression in snap point fiber (Figure 3-2), when compared to fiber above and below the snap point; which is consistent with the microarray results. Comparing the epidermis, fiber, and xylem from the snap point yielded no significant increase in fiber expression, and hinted at an increased expression in the epidermis instead. However, the microarray data, although it also does not show significant

differences in expression either, shows a slight increase in fiber expression. It is likely that this is due to the C4H gene being active in xylem – producing xylem fibers; fiber – producing bast fibers, and epidermis – producing trichome resins (Nagel *et al.* 2008) at this stage of development.

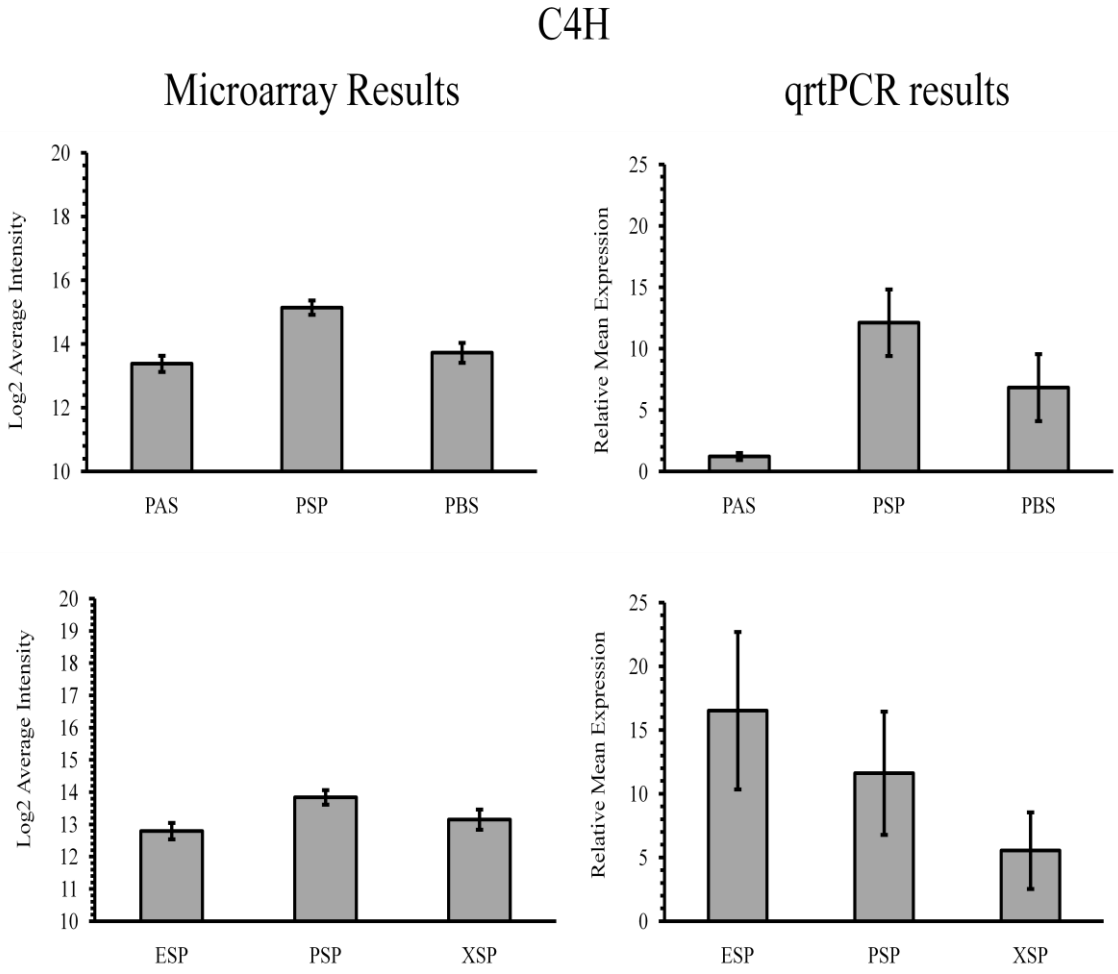


Figure 3-2: C4H normalized to tubulin qRT-PCR expression pattern from primer set c4h1 compared to the microarray results for the same ESTs.
Graphs showing expression above, below, and at the snap point (Top graphs), and showing expression in epidermis, fiber, and xylem (Bottom graphs). qRT-PCR samples were tested at 1/50 dilutions of the cDNA from 5 week old *Cannabis sativa* genotype Carmen. PAS – phloem above snap point, PSP – fiber at the snap point, PBS – fiber below the snap point, ESP – epidermis at the snap point, XSP – xylem at the snap point.

3.3.2 Cinnamoyl CoA Reductase (CCR)

Four ESTs for cinnamoyl CoA reductase (CCR) were found in NCBI Genbank and the Deyholos lab database. Each of these aligns to two separate accessions, suggesting there are at least two separate CCR genes in hemp (Figure 3-3). gi|110190257 most closely aligns to a *Linum album* (flax) CCR mRNA; while gi|110190225 aligns most closely with and *Arabidopsis thaliana* mRNA.

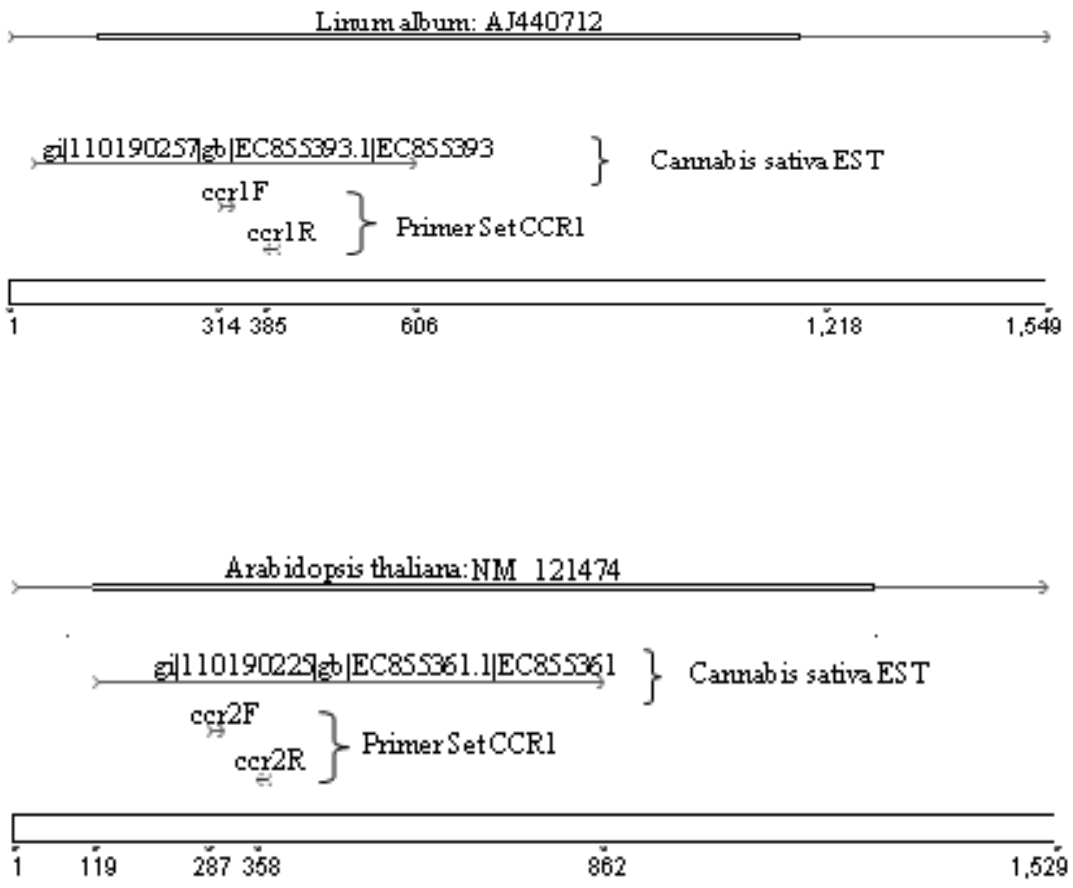


Figure 3-3: Alignments of hemp CCR ESTs,
and the qRT-PCR primer sets designed for them, against the full length mRNA from nearest BLAST hit to the ESTs.

A maximum parsimony tree (Figure 3-4) using the two CCR accessions found in TAIR and the nearest NCBI blast hits places the *Cannabis* CCR1 in the same branch as both *Arabidopsis* CCRs, but shows it as being very loosely related. The other BLAST hits which showed significant similarity for CCR1 are also not very closely related. *Cannabis* CCR2 is most closely related to a CCR-related gene from *Arabidopsis* (At5g14700).

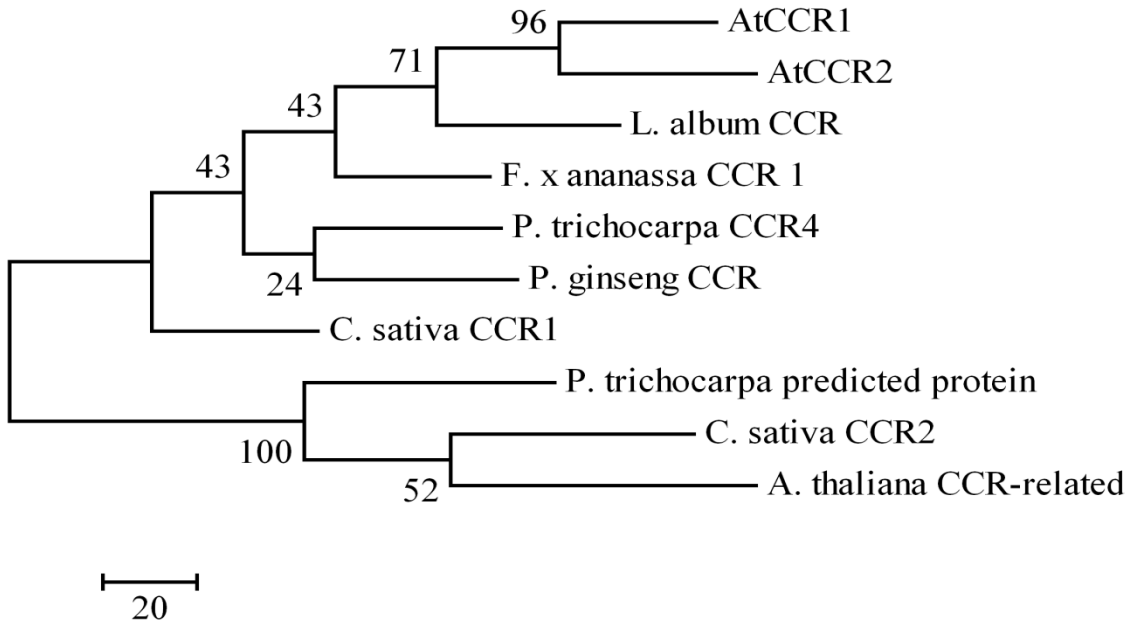


Figure 3-4: Maximum parsimony tree for CCR.
An alignment was made using CLUSTAL W, and a tree was made using the maximum parsimony method (Eck and Dayhoff 1966)of all the *Arabidopsis* CCR accessions from TAIR, the two putative *Cannabis* CCR ESTs, and the two highest NCBI BLAST hits. 211 of 351 total nucleotide positions were parsimony informative. Phylogenetic analyses were conducted in MEGA4 (Tamura *et al.* 2007).

The initial qRT-PCR primer screening showed both CCR ESTs had similar expression patterns both showing increased expression in snap point fiber. Further expression analysis indicates increased expression in fiber below the snap point and in the xylem of the snap point for CCR1 (Figure 3-5). CCR2 shows relatively little change in expression in fiber above below or at the snap point, but expression levels in xylem and epidermis at the snap point are significantly increased (Figure 3-6).

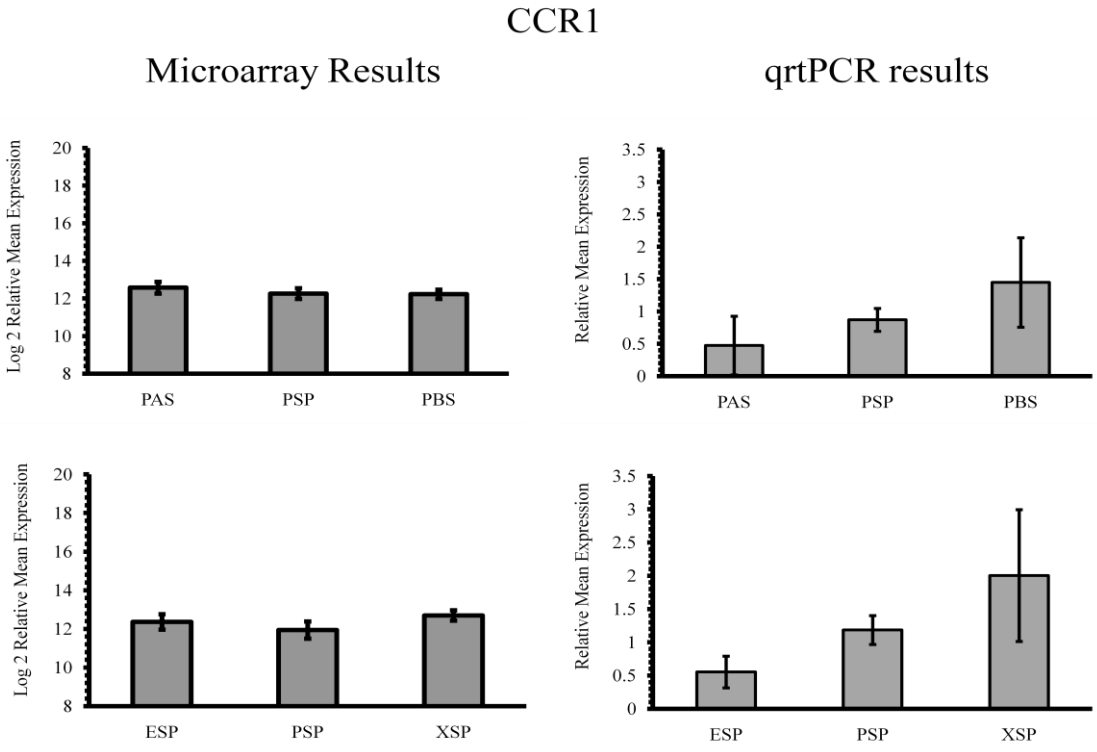


Figure 3-5: CCR normalized to tubulin qRT-PCR expression pattern from primer set ccr1 compared to the microarray results for the same ESTs.
Graphs showing expression above, below, and at the snap point (Top graphs), and showing expression in epidermis, fiber, and xylem (Bottom graphs). qRT-PCR samples were tested at 1/50 dilutions of the cDNA from 5 week old *Cannabis sativa* genotype Carmen.

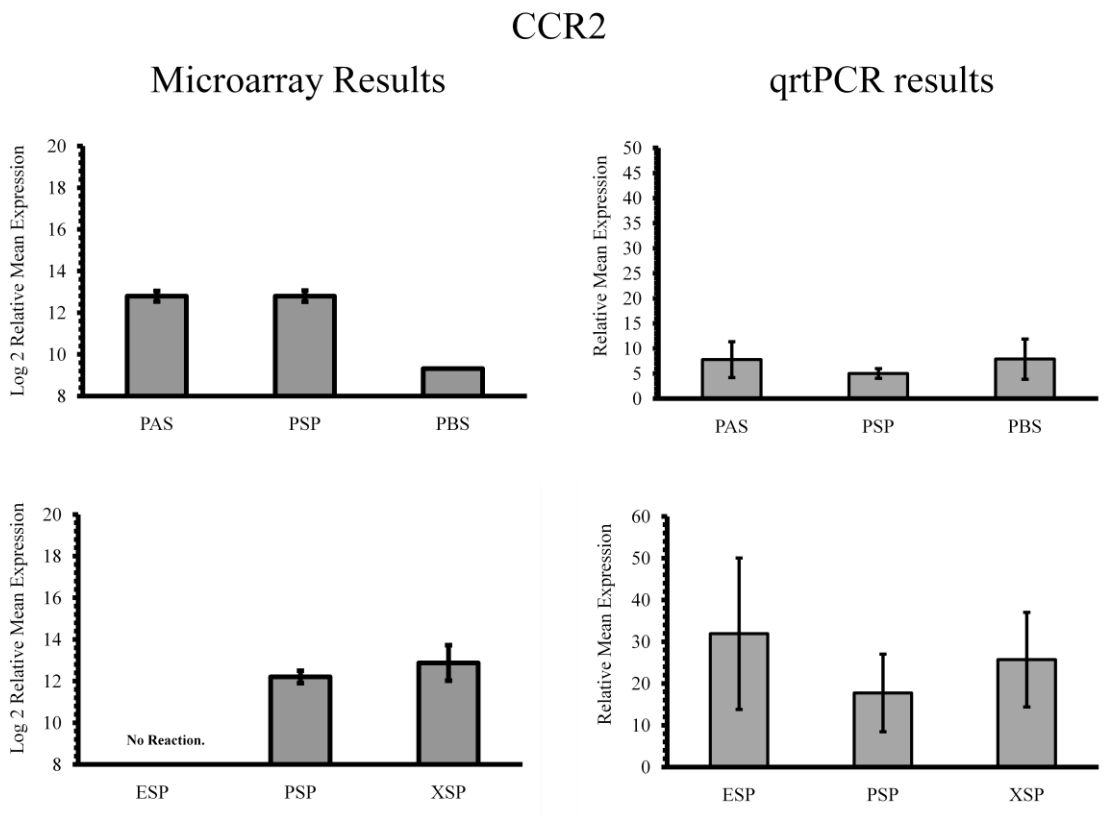


Figure 3-6: CCR normalized to tubulin qRT-PCR expression pattern from primer set ccr2 compared to the microarray results for the same ESTs.
Graphs showing expression above, below, and at the snap point (Top graphs), and showing expression in epidermis, fiber, and xylem (Bottom graphs). qRT-PCR samples were tested at 1/50 dilutions of the cDNA from 5 week old *Cannabis sativa* genotype Carmen.

3.3.3 Ferulate 5-hydroxylase (F5H)

Two F5H ESTs were found both from the Deyholos lab library. These were found to produce a discontiguous alignment when aligned against *Populous tirchocarpa* (Black Cottonwood) (Figure 3-7).

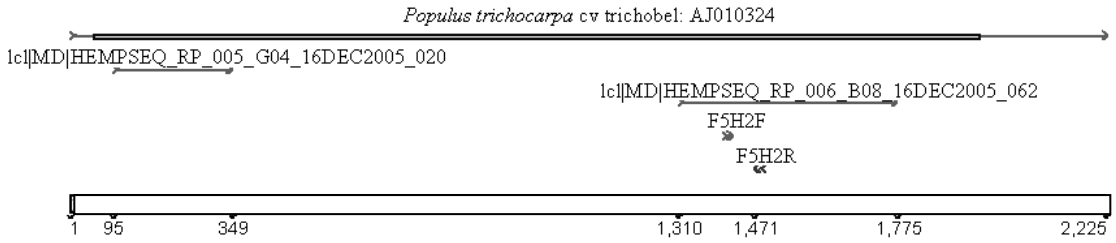


Figure 3-7: Alignment of one *Cannabis sativa* EST for F5H, and the qRT-PCR primer designed for it, against the full length mRNA from nearest BLAST hits to the EST. No primer was designed for the second EST in this instance.

In the initial primer screen the primer set showed an increased in expression in the snap point fiber compared to the stem above or below the snap point. Further studies on F5H showed a subtle increase in snap point fiber compared to the xylem and epidermis (Figure 3-8), though it is not statistically significant. However, a significant increase in fiber from below and at the snap point node compared to the phloem above the snap point. The results from the micro array show a non-significant increase in the fiber compared to either the xylem or epidermis; and the same significant increase in snap point and below snap point

fiber compared to phloem above snap point was seen in the microarray and the qRT-PCR.

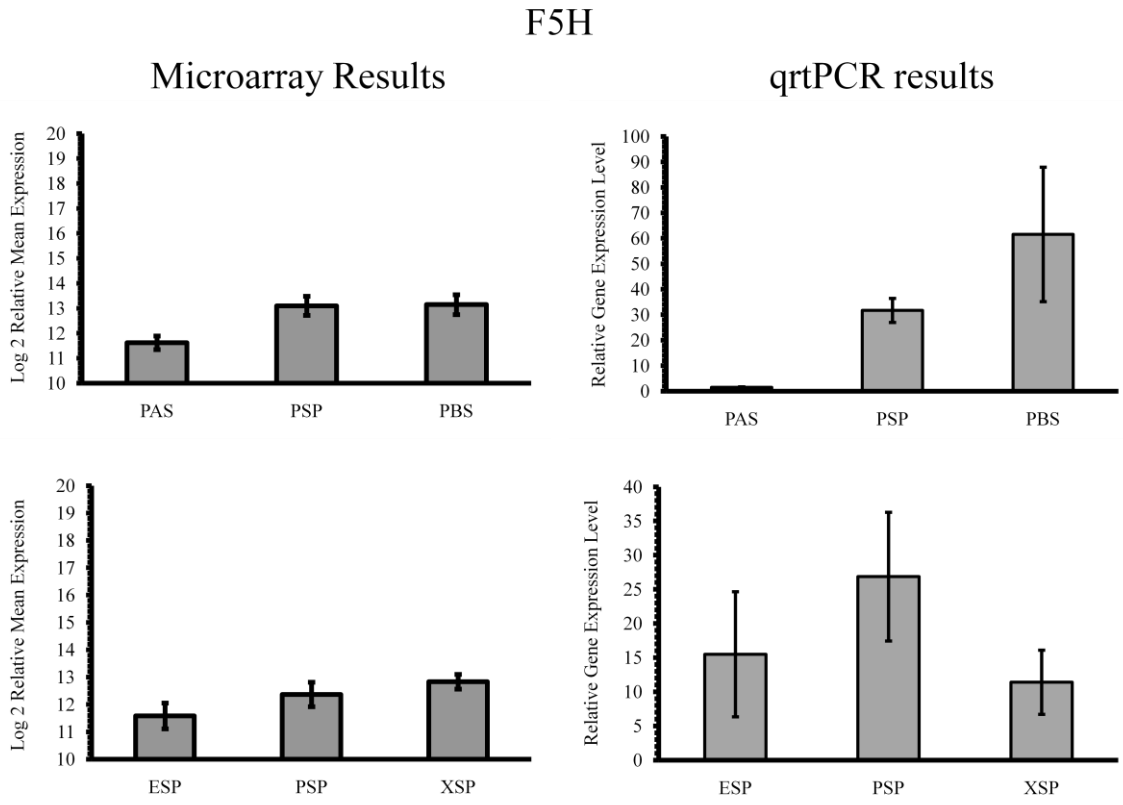


Figure 3-8: F5H normalized to tubulin qRT-PCR expression pattern from primer set f5h2 compared to the microarray results for the same ESTs.
Graphs showing expression above, below, and at the snap point (Top graphs), and showing expression in epidermis, fiber, and xylem (Bottom graphs). qRT-PCR samples were tested at 1/50 dilutions of the cDNA from 5 week old *Cannabis sativa* genotype Carmen.

3.3.4 Caffeic Acid O-methyltransferase (COMT)

Caffeic acid O-methyltransferase COMT was apparently represented by at least 18 ESTs, 3 from Genebank and 15 from the Deyholos Lab. Four different alignments were determined for these ESTs (Figure 3-9); however one of these alignments (alignment 2) may be the 3' end for one of two of the other alignments (alignment 1 or 3). This suggests that there are either two or three COMT isomers: alignment 1, 2, and 3 being one COMT; and alignment 4 being a second COMT – or alignment 1, 3, and 4 being three different COMT genes with alignment 2 being a common 3' end to 1 and 3. COMT 1, 3, and 4 all align most closely to *Rosa chinensis* (China Rose), with alignment 1 and 3 aligning to the same accession; but not each other; and alignment 4 aligning with a different accession. Alignment 2 most closely aligned with *Papaver somniferum* (Opium poppy); although it could be re-aligned discontiguously with either alignment 1 or 3.

COMT is a large multigene family with a number of isoforms. Three *Arabidopsis* COMTs exist in the TAIR database along with thirteen other COMT-like accessions. A maximum parsimony tree was created to determine the relationship between the COMT isoforms and the *Arabidopsis* COMT gene family (Figure 3-10). COMT alignment 4 (primer set 6) is most closely related to the *Arabidopsis* COMT and COMT-like family. COMT alignment 2 provides a 3'end for either alignment 1 or alignment 3 and was not use in the tree. COMT alignment 1 and 3 (corresponding to primer sets 1, 2, 4 and 5) are most closely related to *H. lupulus*

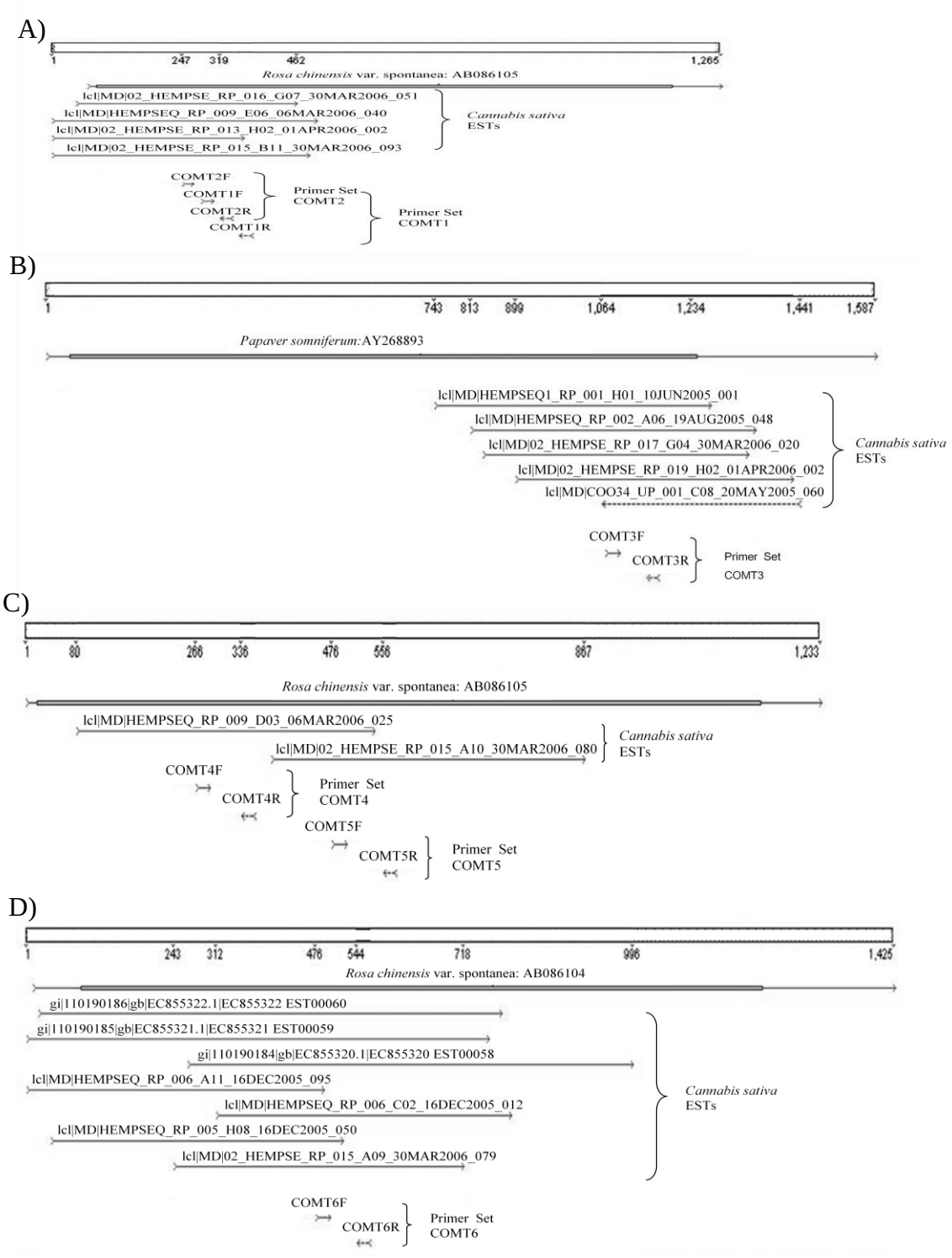


Figure 3-9: Four alignments of COMT ESTs
and the related qRT-PCR primer sets to their nearest Blast search neighbor.
A) alignment 1, B) alignment 2, C) alignment 3, and D) alignment 4.

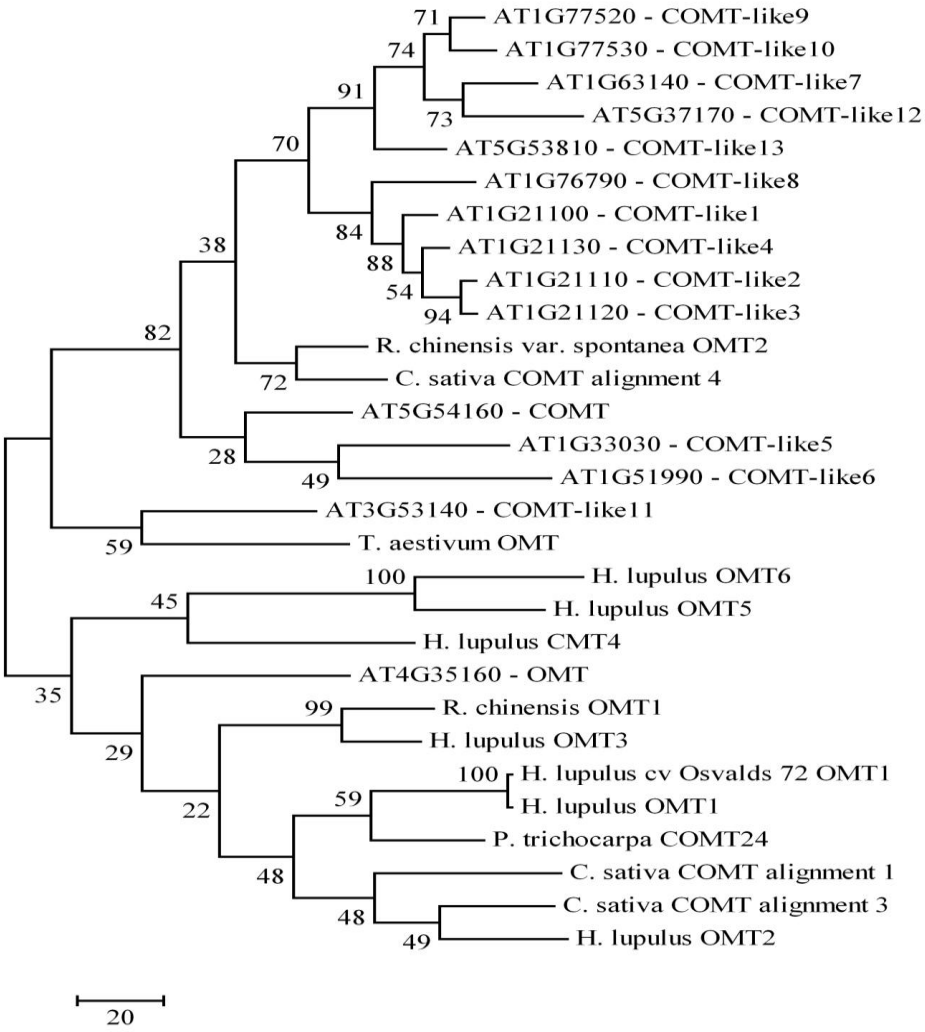


Figure 3-10: Maximum parsimony tree for COMT.
An alignment was made using CLUSTAL W, and a tree was made using the maximum parsimony method (Eck and Dayhoff 1966) of all the *Arabidopsis* COMT accessions from TAIR, three *Cannabis* COMT contigs (Alignment 1, 3, and 4), and the highest NCBI BLAST hits. 165 of 178 total nucleotide positions were parsimony informative. Phylogenetic analyses were conducted in MEGA4 (Tamura *et al.* 2007).

O-methyltransferase 2 (OMT2). *H. lupulus* OMT2 is active in the phenylpropanoid pathway of Hops. COMT primer set two (alignment 1) shows an increase in snap point fiber.

The qRT-PCR expression analysis of COMT primer set 2 (Figure 3-11) showed a significant increase in snap point fiber expression when compared to both snap point epidermis and xylem; and fiber above and below the snap point. The microarray results showed an increase in the fiber above the snap point and in the xylem. The selected example EST in the microarray results (02_HEMPSE_RP_016_G07_30MAR2006_051) showed significant differences; however, the other ESTs from alignment 1, which had microarray results, although they showed similar patterns, did not show patterns with significant differences (not shown). Repeats of the qRT-PCR with the same primer set on other samples (not shown) continued to show a similar significant expression for the snap point epidermis, fiber, xylem pattern.

Microarray results from ESTs corresponding to COMT primer set 4 (HEMPSEQ_RP_009_D03_06MAR2006_025) (not shown) only gave significant results in fiber for the PEX-S experiment, initial qRT-PCR trials did not give stable results. Also, microarray results from ESTs corresponding to COMT primer set 6 (not shown) showed no significant expression in the fiber above, at, and

below the snap point; but showed a significant increase in snap point xylem compared to epidermis and fiber.

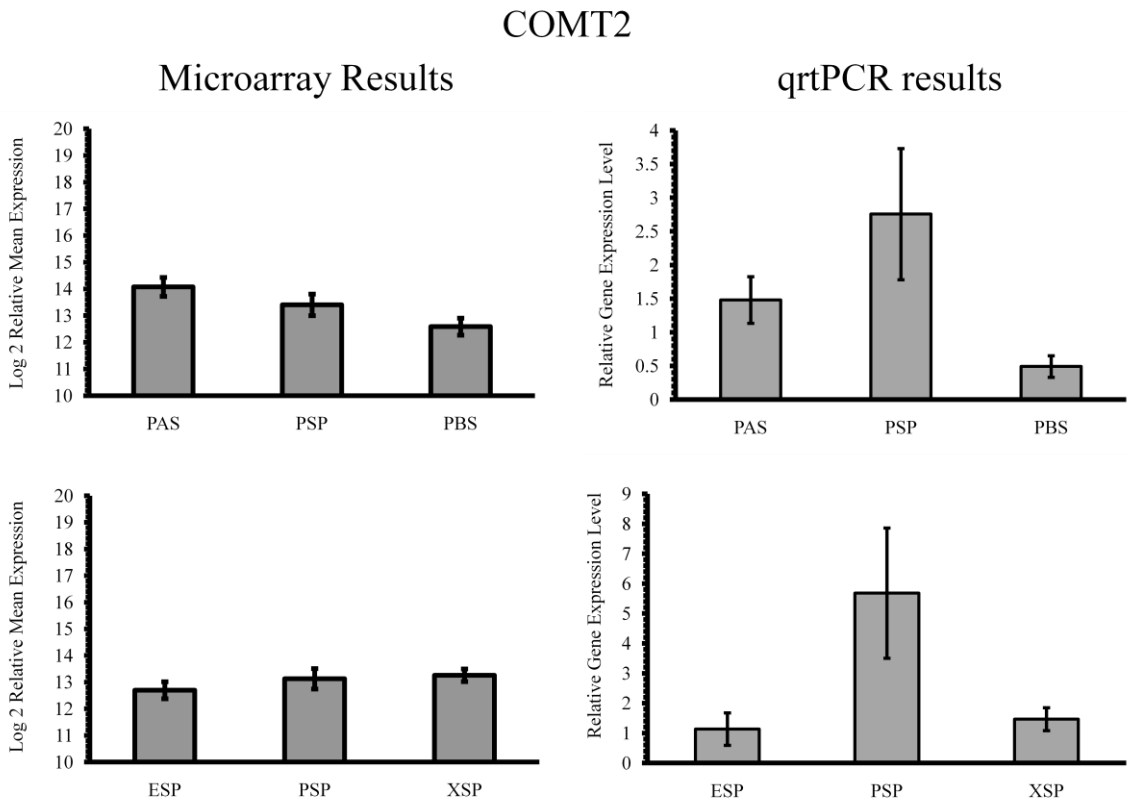
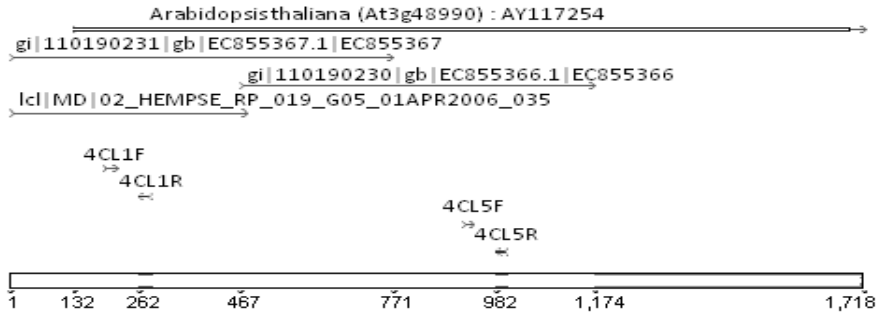


Figure 3-11: Caffeic acid O-methyltransferase normalized to tubulin qRT-PCR expression pattern from primer set comt2 compared to the microarray results for one of the ESTs from the first alignment (02_HEMPSE_RP_013_H02_01APR2006_002). All the microarray results for the ESTs in the first alignment show similar expression patterns. Samples were tested at 1/50 dilutions of the cDNA from 5 week old *Cannabis sativa* genotype Carmen. PAS – phloem above snap point, PSP – fiber at the snap point, PBS – fiber below the snap point, ESP – epidermis at the snap point, XSP – xylem at the snap point.

3.3.5 4-coumarate-CoA Ligase (4CL)

Seven 4-coumarate-CoA ligase (4CL) ESTs were found, 6 from the NCBI Genbank database, and one from the Deyholos lab library. These aligned into two separate groups (Figure 3-12). gi|110190231, gi|110190230|, and lcl|MD|02_HEMPSE_RP_019_G05_01APR2006_035 aligned most closely to a putative 4-coumarate-CoA ligase mRNA from *Arabidopsis thaliana*. The remaining ESTs aligned most closely to *Rubus idaeus* (red raspberry) 4-coumarate-coA ligase 1 mRNA.

A)



B)

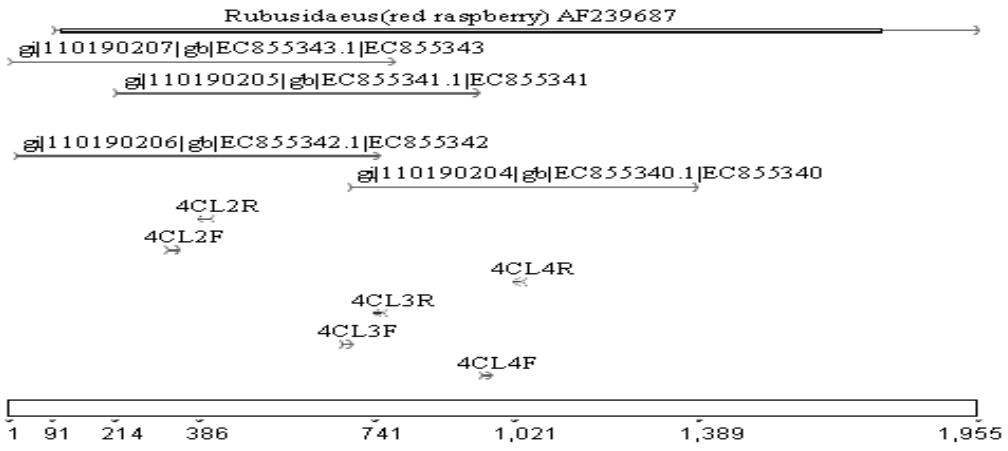


Figure 3-12: Alignment of *Cannabis sativa* ESTs for 4CL,
and the qRT-PCR primer sets designed for it, against the full length mRNA from nearest BLAST hits to the ESTs.

A tree was made to determine the relationship of the two 4CL alignments compared to the *Arabidopsis* isoforms, which represent the best characterized 4CL family (Figure 3-13). The *Arabidopsis* 4CL family has four 4CL isoforms that show measurable activity on phenylpropanoid pathway substrates; and seven remaining putative 4CLs did not show catalytic activity which were renamed 4CL-like genes (Costa *et al.* 2005). The tree; which contained all the At4CL accessions and contigs of the putative hemp 4CL, showed hemp 4CL1 is related most closely to At4CL8, a 4CL-like gene in *Arabidopsis* that has no known function but is highly expressed in young tissue (Costa *et al.* 2005). 4CL2 is related to the At4CL1/At4CL2 branch, which is highly active on phenylpropanoid pathway substrates, and is highly expressed in older *Arabidopsis* tissues (Costa *et al.* 2005).

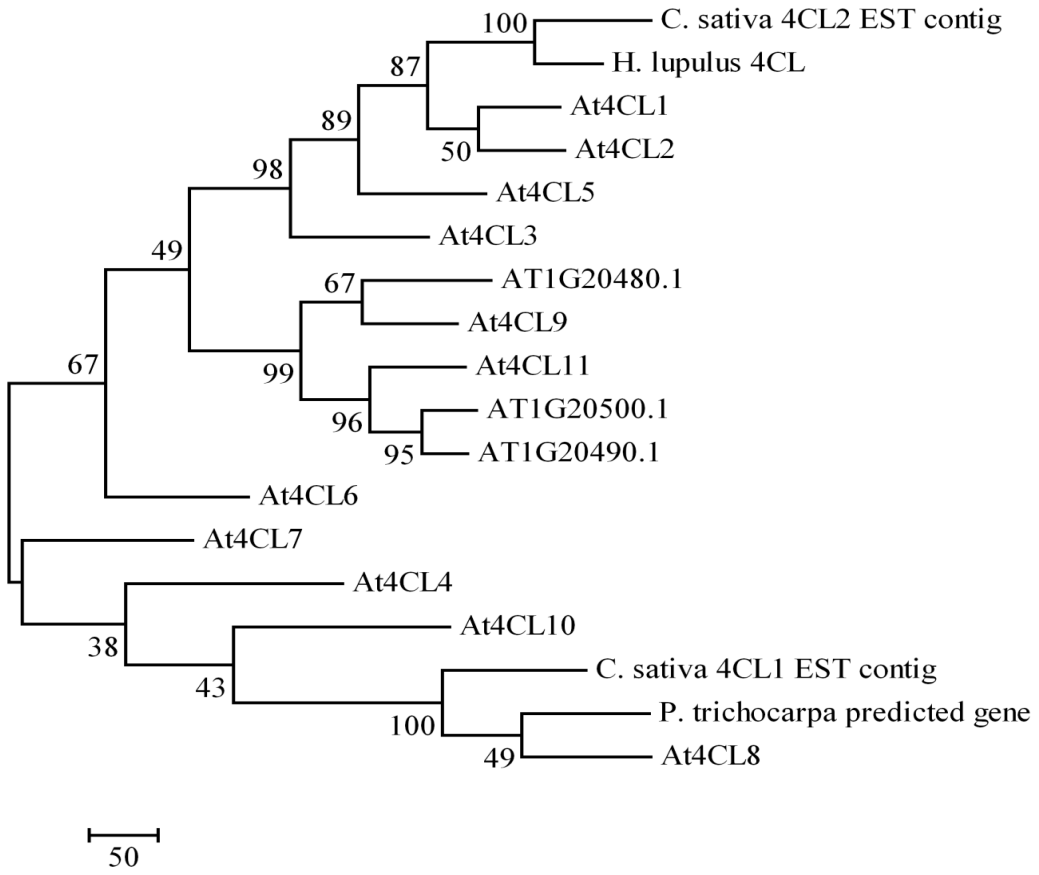


Figure 3-13: Maximum parsimony tree of 4CL.
An alignment was made using CLUSTAL W, and a tree was made using the maximum parsimony method (Eck and Dayhoff 1966) of all the *Arabidopsis* 4CL accessions from TAIR, the two putative *Cannabis* 4CL contigs, and the two highest NCBI BLAST hits to them. 496 of 582 total nucleotide positions were parsimony informative. Phylogenetic analyses were conducted in MEGA4 (Tamura *et al.* 2007).

Initial qRT-PCR tests confirm these alignments. Primer sets 4cl1 and 4cl5 showed similar expression patterns with different expression levels, indicating an increase in snap point phloem. The relative mean expression of 4cl1 in phloem was 21, while the relative mean expression of 4cl5 in phloem was 83. This difference is

likely due to subtle differences in primer binding but could be indicative of gene differences as well. Primer sets 4cl2 and 4cl4 showing a different expression pattern, with little snap point phloem expression. 4cl3 shows a third expression pattern suggesting there may be three different 4CL genes. Alternately, the alignment pairing it with 4cl2 and 4cl4; which show decreased expression in snap point fiber, suggests that 4cl3 may be hybridizing to multiple 4CL genes. 4cl3, 4cl2, and 4cl4 were not analyzed further.

Further qRT-PCR experiments done for the 4cl1 and 4cl5 primer sets showed similar expression patterns and levels (Figure 3-14) confirming the alignment (Figure 3-12). The qRT-PCR expression patterns were similar to the microarray expression pattern for the snap point epidermis-fiber-xylem experiments; with the snap point fiber showing a significant increase. But the above-at-below snap point experiments showed a difference between the two methods. The microarray experiment showed an increase in snap point fiber expression, but the qRT-PCR the expression above the snap point was significantly less than expression shown at or below the snap point. The expression below the snap point increased compared to the expression at the snap point, although this difference was not statistically significant.

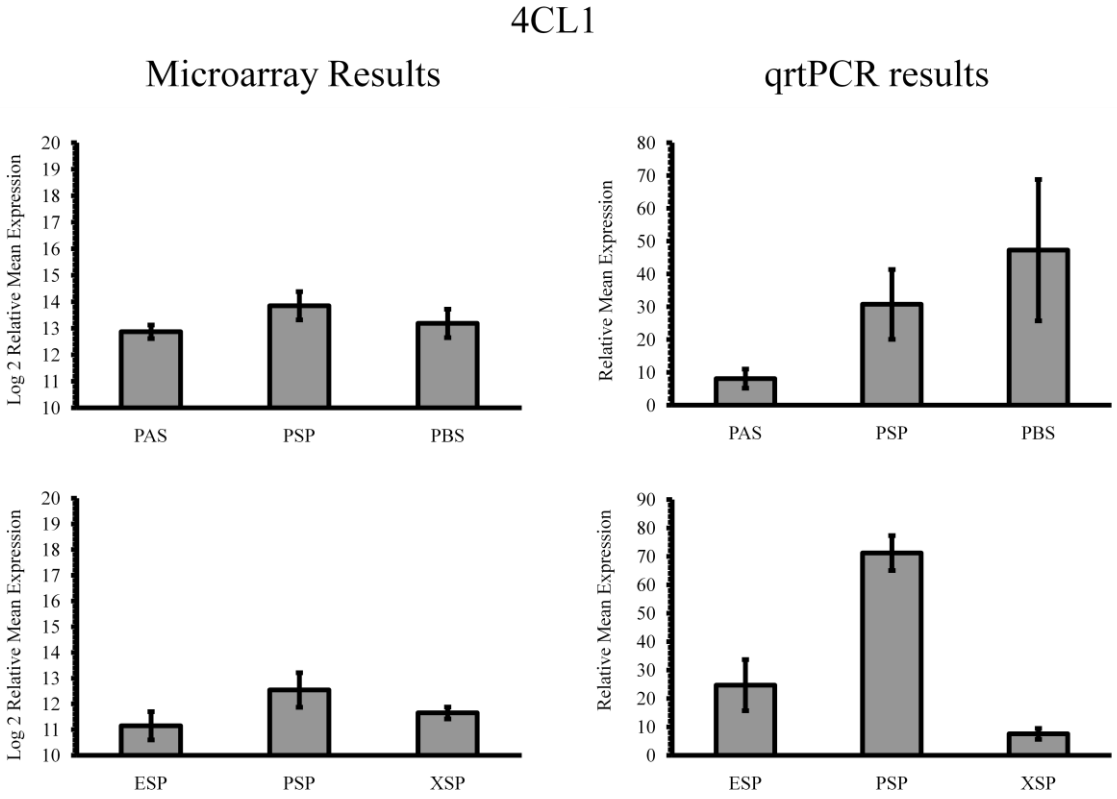


Figure 3-14: 4-coumarate-CoA ligase normalized to tubulin qRT-PCR expression pattern from primer set 4cl1 compared to the microarray results for the Deyholos Lab EST.
4cl5 primer set qRT-PCR expression pattern (not shown) was similar to the 4cl1 primer set. Samples were tested at 1/50 dilutions of the cDNA from 5 week old *Cannabis sativa* genotype Carmen. PAS – phloem above snap point, PSP – fiber at the snap point, PBS – fiber below the snap point, ESP – epidermis at the snap point, XSP – xylem at the snap point.

3.3.6 Cellulose Synthase (CESA)

Nine ESTs were found for cellulose synthase; 6 from NCBI genbank and three from the Deyholos library. They aligned into three possible contigs (Figure 3-15).

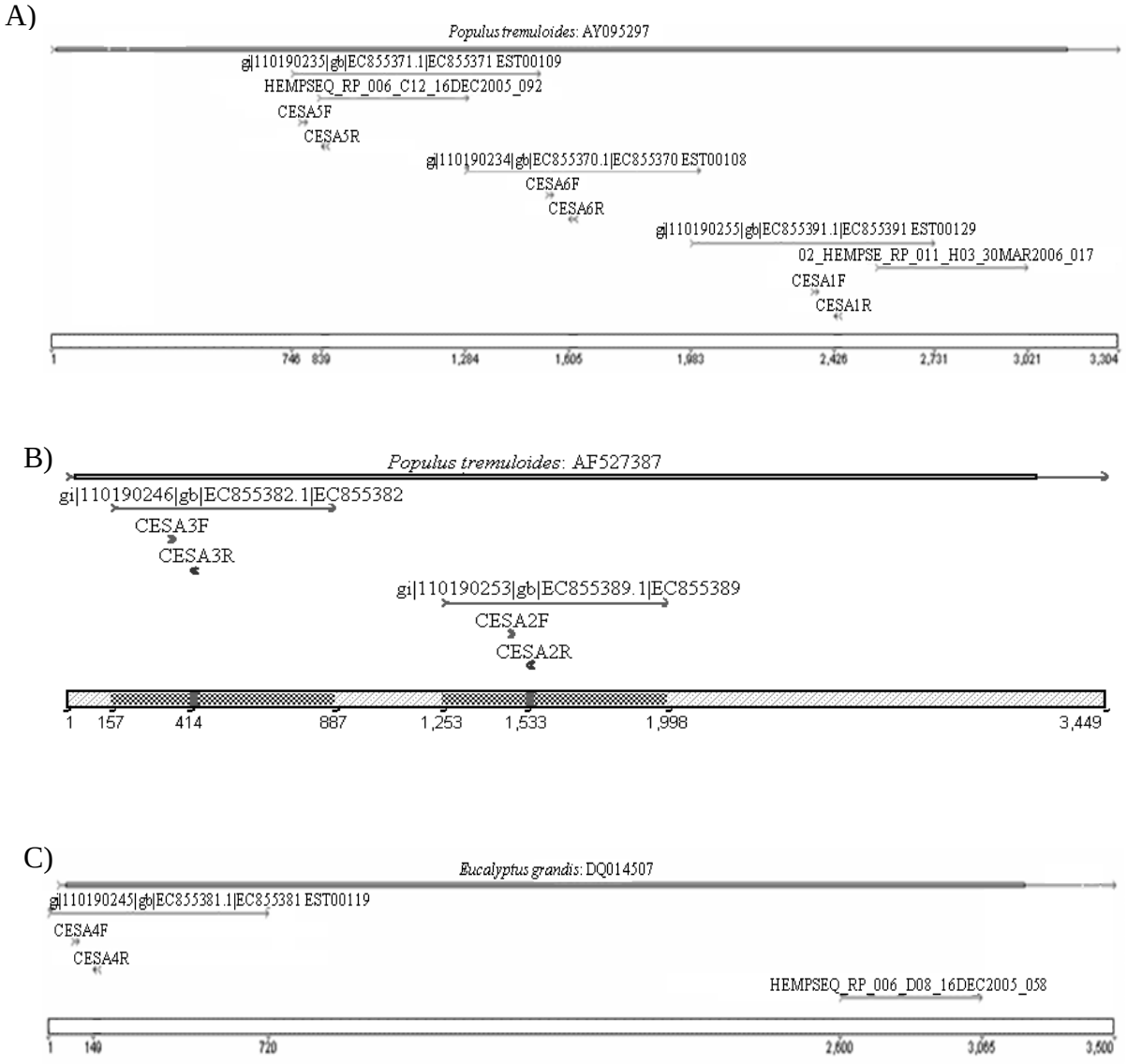


Figure 3-15: Alignment of *Cannabis sativa* ESTs for CESA, and the qRT-PCR primer sets designed for it, against the full length mRNA from nearest BLAST hits to the ESTs.

Initial qRT-PCR results show an increase CESA expression in snap point fiber compared to the stem above and below the snap point for primer sets cesa2, cesa3, and cesa4. Primer sets cesa1, cesa5, and cesa6 all showed different expression patterns, none of which gave any increase in CESA expression in the snap point fiber. Further qRT-PCR results for primer sets cesa2, cesa1, and cesa4 all showed increases in xylem for the PEX-S experiments, or phloem below the snap point for the SAB-P experiments (Figure 3-16, Figure 3-17, Figure 3-18).

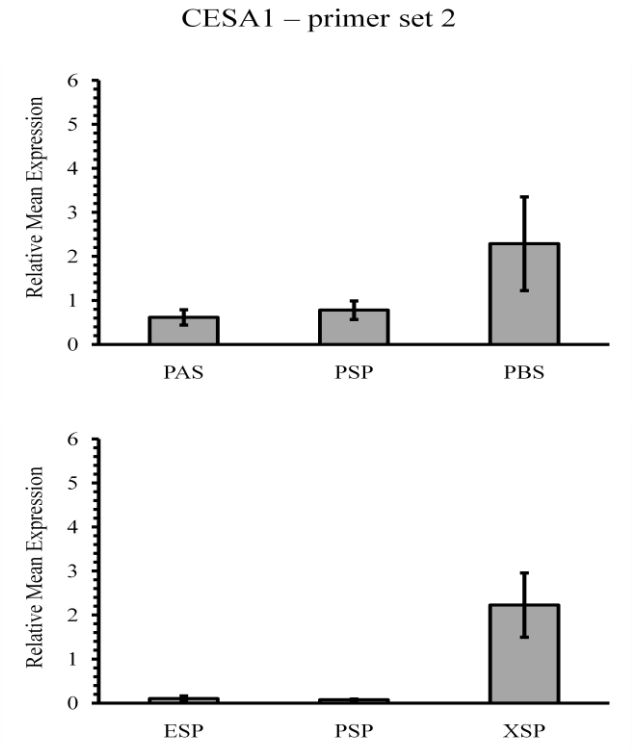


Figure 3-16: Cellulose synthase normalized to tubulin qRT-PCR expression pattern from primer set cesa2.
Samples were tested at 1/50 dilutions of the cDNA from 5 week old *Cannabis sativa* genotype Carmen. PAS – phloem above snap point, PSP – fiber at the snap point, PBS – fiber below the snap point, ESP – epidermis at the snap point, XSP – xylem at the snap point.

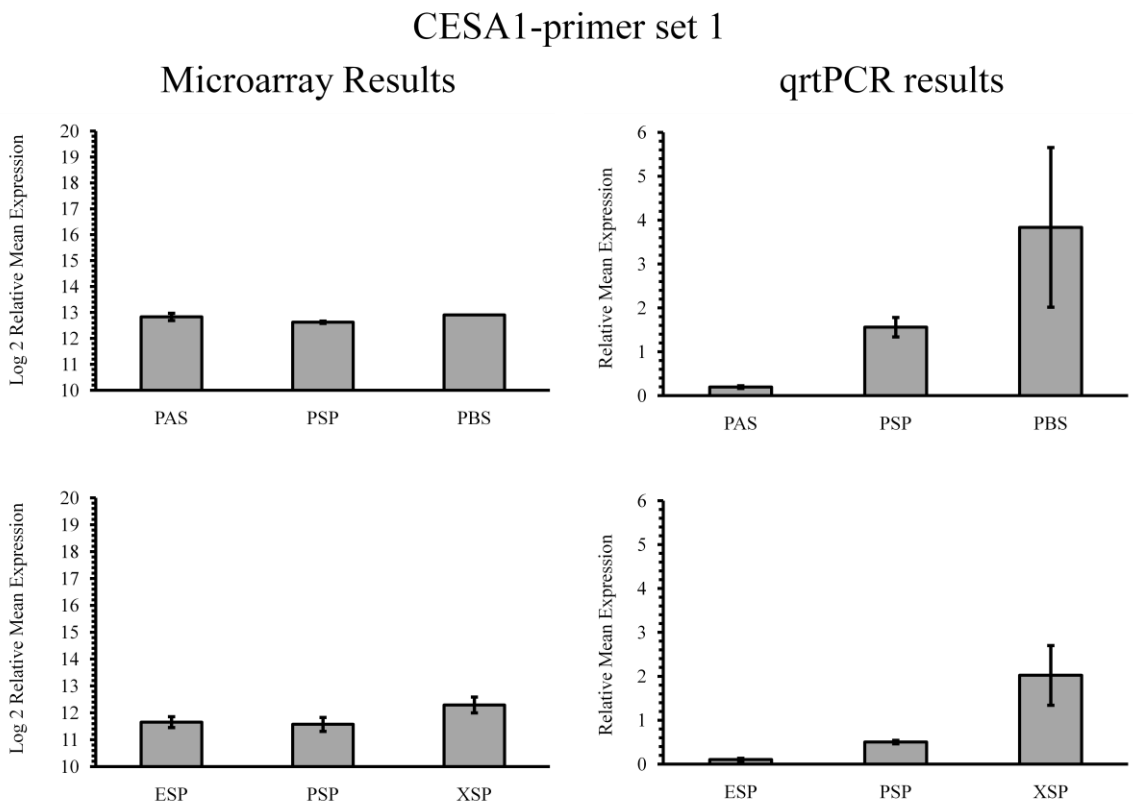


Figure 3-17: Cellulose synthase normalized to tubulin qRT-PCR expression pattern from primer set cesa1 compared to the microarray results for one of the ESTs from the first alignment (HEMPSEQ_RP_006_C12_16DEC2005_092). Both microarray results for the ESTs in the first alignment show similar expression patterns. Samples were tested at 1/50 dilutions of the cDNA from 5 week old *Cannabis sativa* genotype Carmen. PAS – phloem above snap point, PSP – fiber at the snap point, PBS – fiber below the snap point, ESP – epidermis at the snap point, XSP – xylem at the snap point.

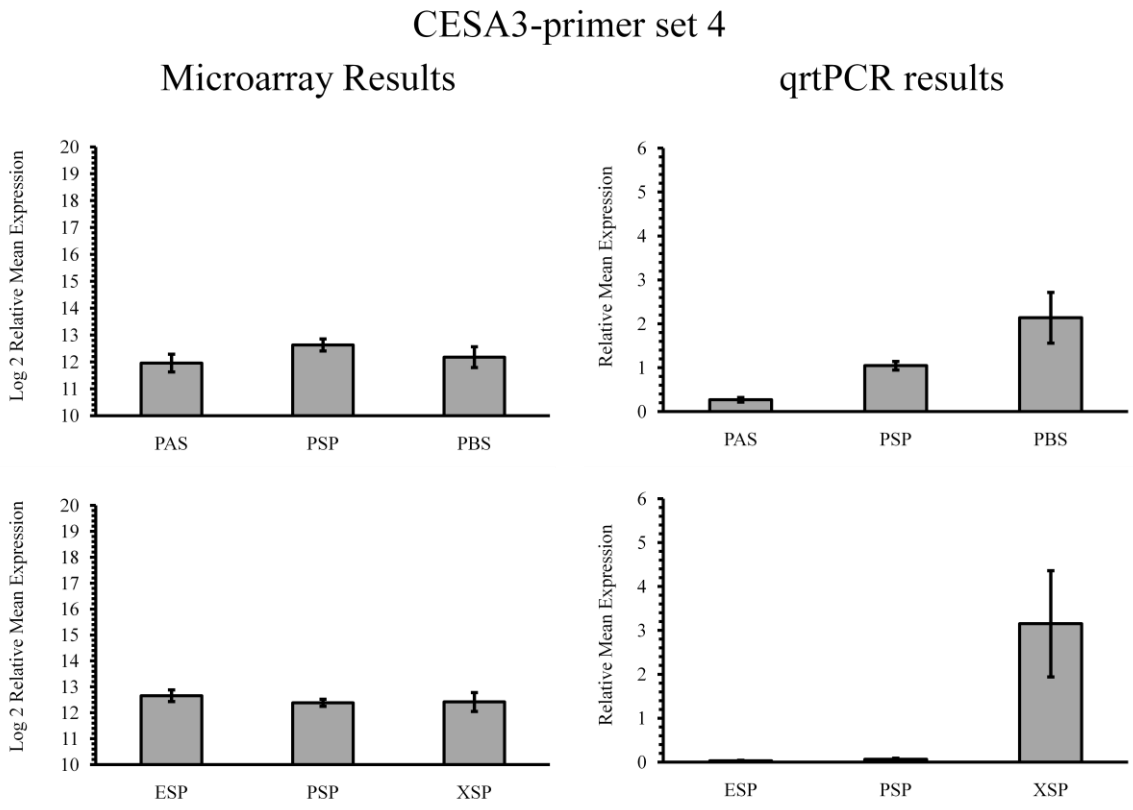


Figure 3-18: Cellulose synthase normalized to tubulin qRT-PCR expression pattern from primer set cesa4 compared to the microarray results the ESTs from the third alignment (HEMPSEQ_RP_006_D08_16DEC2005_058). Samples were tested at 1/50 dilutions of the cDNA from 5 week old *Cannabis sativa* genotype Carmen. PAS – phloem above snap point, PSP – fiber at the snap point, PBS – fiber below the snap point, ESP – epidermis at the snap point, XSP – xylem at the snap point.

3.3.7 Beta-D-galactosidase (BGAL)

Two beta-D-galactosidase ESTs were identified in the Deyholos lab library. They did not align to each other, even using their nearest BLAST neighbor as a guide. One set of qRT-PCR primers was made which aligned to HEMPSEQ_RP_007_F02_06FEB2006_006 (Figure 3-19).

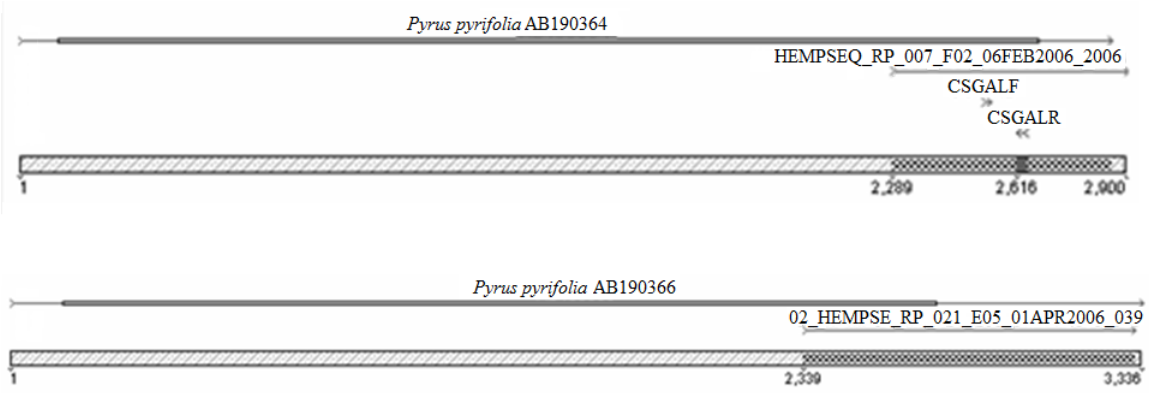


Figure 3-19: Alignment of *Cannabis sativa* ESTs for BGAL, and the qRT-PCR primer sets designed for it, against the full length mRNA from nearest BLASTn hits to the ESTs.

Microarray analyses show this gene to be significantly increased in snap point epidermis and in fiber tissue above and at the snap point (Figure 3-20). Both ESTs showed similar microarray results. qRT-PCR results support the microarray results showing an increased expression in epidermis and fiber. However, in the testing of fiber above, at, and below the snap point for HEMPSEQ_RP_007_F02_06FEB2006_006 the qRT-PCR and the microarray

results do not match, and the qRT-PCR shows a distinct significant increase in snap point fiber expression.

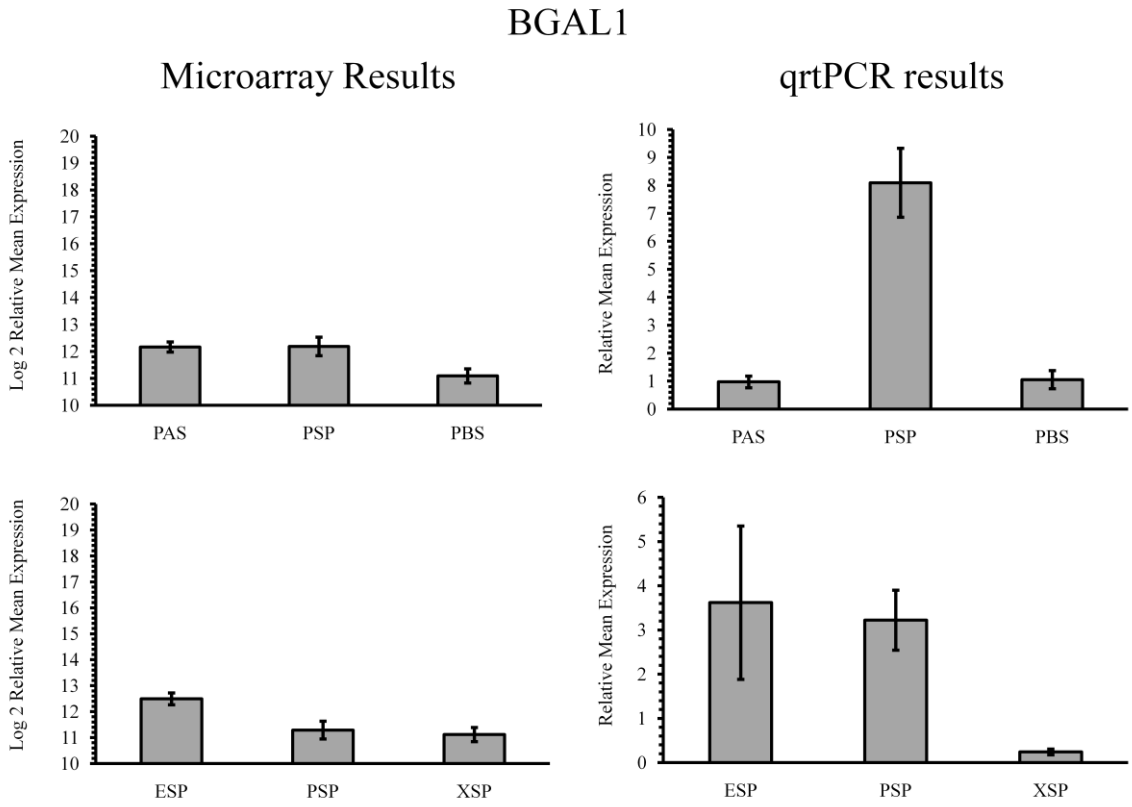


Figure 3-20: Beta-D-galactosidase normalized to tubulin qRT-PCR expression pattern from primer set BGAL compared to the microarray results from the EST: HEMPSEQ_RP_007_F02_06FEB2006_006. Samples were tested at 1/50 dilutions of the cDNA from 5 week old *Cannabis sativa* genotype Carmen. PAS – phloem above snap point, PSP – fiber at the snap point, PBS – fiber below the snap point, ESP – epidermis at the snap point, XSP – xylem at the snap point.

3.4 Discussion

Since full length mRNA for the phenylpropanoid pathway genes is not known and because *Cannabis* has a high level of natural heterozygosity, the determination of the identity of exact genes in the bast fiber phenylpropanoid pathway using ESTs is only a starting point. The alignments and the expression analysis suggests there are two PAL genes, one C4H gene, 2 CCR genes, two 4CL genes, one PCS3H gene, three COMT genes, one F5H gene, and one CCoAOMT gene. The expression analysis of the other cell wall genes of interest (BGAL and CESA) suggests there are two and three genes, respectively.

Three genes were only given a cursory examination PAL, PCS3H, and CCoAOMT. PCS3H is part of the shikamate pathway, and although its substrates can be used in the phenylpropanoid pathway, it is not expected to play a large role in it. In the case of PAL and CCoAOMT, down-regulation of either causes a down-regulation in G subunits; but in hemp the G subunits are low compared to other bast fibers such as flax (Crônier, Monties and Chabbert 2005). Therefore, other genes were considered better candidates for further study.

The genes that show an increase in expression in the snap point fiber both in the PEX-S experiment and the SAB-P experiment are COMT1 (primer set comt2), and BGAL. Those that show an increase in fiber at the snap point PEX-S experiment and an increase in the fiber below the snap point are F5H, and 4CL1;

and the CCR and C4H expression patterns are different from the other genes examined.

C4H shows an increase in snap point fiber when compared to the fiber above and below the snap point, but it does not show a significant difference in the different tissues; suggesting that it is involved to the same extent in all types of fiber production – bast (phloem fiber), wood (xylem fiber), and trichome resin (epidermis). This is unsurprising since C4H controls the second step in the phenylpropanoid pathway, and is active in the production of all the lignin polymer subunits and free phenylpropanoids.

4CL, which catalyses the steps leading to the different types of lignin (G, H, and S forms), is a gene family which has at least two members showing activity in the maturing hemp fiber. 4CL1 is highly expressed in the fiber and below the snap point, and 4CL2 is highly expressed below the snap point. 4CL1, most closely to the *Arabidopsis* At4CL8 a 4CL-like gene was shown to be highly expressed in young tissue (Costa *et al.* 2005), but its function has yet to be determined. 4CL2 is related to the *Arabidopsis* At4CL1/At4CL2 branch, which is highly active on phenylpropanoid pathway substrates (Costa *et al.* 2005). Down-regulation of 4CL has a number of species-specific effects, but results in an overall decrease in lignin content. The changes to the subunit forms depends on the species of plant examined, this could be due to differences in the various isoforms. 4CL down-

regulated transgenic aspen contains less lignin, more cellulose, and grows taller (Anterola and Lewis 2002). Suggesting 4CL1 and 4CL2 are valuable targets for fiber modification in *Cannabis*.

The two CCR genes gave two significantly different expression patterns. CCR2 showed an increase of expression in the snap point epidermis but its expression was at equal levels in the fiber above, below, and at the snap point. This may indicate that CCR2 is involved in trichome resin production, and is likely not a target for reducing lignin amounts. CCR1 expression was increased in the snap point xylem compared to fiber or epidermis and in fiber below the snap point compared to fiber above and below the snap; which suggests that it may be involved in maturing fibers. Experiments with CCR mutants in other plants tobacco and *Arabidopsis* show not only a decrease in the G subunit, but also a decrease in the cohesion of the cellulose microfibrils (Boerjan, Ralph and Baucher 2003); which suggests that CCR1 may be a good target in developing plants that are more easily processed.

One F5H alignment was determined from two *Cannabis* ESTs. It has an increased expression level in the snap point fiber compared to the epidermis and the fiber, and in the fiber below the snap point suggesting that it may be involved in maturation of the fiber. *Arabidopsis* F5H null mutants contain no S subunits (Boerjan, Ralph and Baucher 2003), and thus higher G subunits. G subunits are

involved in lignin branching in poplar (Stewart *et al.* 2009), so changes to F5H may change the lignin type rather than the lignin content of the fibers. This makes this a valuable target for altering the bast fibers in *Cannabis*.

COMT had four possible alignments, but one of which was a three prime end for two of the three remaining alignments. Expression analysis gave three possible patterns, one for each section of stem, with the second alignment showing an increase at the snap point fiber for both comparisons. Past studies have shown that reduction in COMT activity does reduce lignin but makes it harder to pulp because of a decrease in S unit and an increase in cross linking of the fibers (Boerjan, Ralph and Baucher 2003). COMT1 (primer set 2) may make a good target for altering lignin content but a simple over expression or under expression of the gene is not recommended if the intent is to improve the characteristics of hemp for pulping.

Of the two cell wall genes examined (CESA and BGAL), only BGAL has increased expression levels in snap point fiber for both qRT-PCR comparisons. CESA, for all three of the qRT-PCR sets showed an increase in snap point xylem, and in the bast fiber below the snap point. This is not unexpected since the maturing fiber walls are increasing in width and thus using more cellulose. BGAL, on the other hand, has been shown to be active in the loosening and remodeling of phloem fibers ((Roach and Deyholos 2007); (Gorshkova and

Morvan 2006)), and is likely another good candidate gene for altering fiber quality.

3.5 Conclusion

The best target genes for fiber alteration in hemp based on this study are COMT1, F5H, and BGAL. COMT1, which show increased expression in the phloem, including fiber at the snap point for both experimental sets (SAB-P and PEX-S) is the best target for manipulation in regards to clear expression level results. Because its disruption causes a decrease in S subunits and increase in lignin cross-linkages, a down-regulated strain may be of interest in composite production but not good for processes concerned with pulping or digestion. Both F5H and BGAL show an increase in expression in the phloem and fiber in the PEX-S experiment, but also an increase in phloem including fiber below the snap point in the SAB-P experiment, suggesting involvement in fiber maturation. F5H down-regulation has been shown to decrease S subunits and increase G subunits, so this is a gene of interest where the need to manipulate the S/G ratio is required such as composite production. BGAL is active in the loosening and remodeling of the phloem fibers during the intrusive growth phase, this has the potential to change the length and strength of the fibers and may be of interest in applications such as ease of harvest, and composite production.

Future work requires further sequence information especially for CCoAOMT, and PCS3H, such as the isolation of full length mRNA; and in some cases functional analysis of the activity of the specific isomers in the pathway, such as 4CL1. Future analysis of 4CL, because it has been shown to be an informative modification in other plants, such as *Arabidopsis* and aspen, would be of particular benefit. The 4CL isomer that shows increased expression in the snap point phloem, 4CL1, aligns most closely with an *Arabidopsis* 4CL-like gene with no known function, so analysis of its action in the lignin pathway must be determined before it can be considered a target for lignin manipulation.

Chapter 4

Development of a VIGS (Virally Induced Gene Silencing) System for Hemp

4 Development of a VIGS (Virally Induced Gene Silencing) System for Hemp

4.1 Introduction

To help infer gene function, it is useful to study plants with alterations in gene activity. Most commonly, these are obtained transgenically or through mutant screens. Unlike many other commercial crops and model species, a practical transformation and regeneration protocol has not yet been established for hemp, despite efforts in several laboratories. Progress towards transgenic hemp has been limited to the production of transformed cell cultures (Feeney and Punja 2003) from which plantlets could not be regenerated, and a report of (Mandolino and Ranalli 1999) the regeneration of a single plantlet from tissue culture, which was not transformed.

Forward genetics uses phenotypic screening to identify mutations that affect a biological process of interest. In hemp, this is an inefficient process, since the imperfect flowers and high out-crossing rate make it difficult to obtain homozygous individuals that may reveal phenotypes of recessive mutations. Furthermore, the map-based cloning or transgenic complementation that often follows forward genetics screens in other species is also difficult in hemp because of the lack of a genetic map or the ability to produce transgenic plants. In contrast, reverse genetics starts with a known gene sequence, and then tries to find the phenotype associated with mutations in this sequence. A reverse genetics

(TILLinG) platform is being developed for hemp at the Alberta Research Council (ARC). However, screening a population for sequence variants, then back-crossing and phenotyping are slow processes, particularly in hemp. Thus, a virally-induced gene silencing (VIGS) system may allow us to increase the speed of the current reverse genetics studies.

Virally induced gene silencing exploits the natural reaction of a plant to turn off the genes of a virus that is infecting it. This is a process involving post-transcriptional gene silencing (PTGS) also called RNA interference (RNAi) or silencing. When double stranded RNA (dsRNA), seen in many viruses, infects cells it triggers a ribonuclease, DICER-like ribonuclease, to digest it into small 21-24bp fragments. These fragments are incorporated into an RNA induced silencing complex (RISC) which then targets sequences homologous to the fragments for degradation. This RNA silencing spreads systemically through the plant, allowing it to overcome both viral dsRNA and DNA infections. Some viruses encode suppressors of the plant defense system, which helps them survive the plant's defenses causing a constant low level of infection in the plant that may not show visible symptoms.

If a host gene segment, usually between 150 and 500 base pairs, is inserted into a vector that the plant identifies as viral it will silence not only the virus but also the host gene; or, depending on how carefully the fragment is chosen, it may silence a

gene family. The VIGs vector is transferred into the plant usually mediated by *Agrobacterium tumefaciens*, or directly using liquor made up of engineered virus (Burch-Smith *et al.* 2006). Agrobacterium infects hemp (Feeney and Punja 2003) so a transient system of an Agrobacterium-mediated VIGs vector which induces silencing is desirable. Silencing of a gene usually happens 1-12 weeks post infection, depending on the virus and plant system. This allows researchers a fast and convenient way to examine the effects of a null mutant in a plant.

VIGS systems have some advantages over other strategies for generating loss-of-function phenotypes. Because silencing can be induced in seedlings and older plants, VIGS can be used to characterize genes that would otherwise be embryo lethal as genetic null mutants. It can also be less labour- and time-intensive than developing and screening thousands of mutant lines in hemp, and it can be used on plants that have no systems for regeneration and transformation. Furthermore, either specific genes or gene families can be targeted depending on the vector sequence. The system also has disadvantages. It is transient, thus making the study of long term developmental effects problematic. Viral symptoms seen in the plant can mask the phenotype of the gene. Infections may not always lead to a systemic reaction, especially in older plants, and a mutant plant line cannot be made from virally infected plants ((Burch-Smith *et al.* 2004), (Dinesh-Kumar *et al.* 2003), (Unver and Budak 2009)).

Many VIGS systems have been developed: tobacco mosaic virus (TMV)-for tobacco, potato virus X (PVX)-for *N. benthamiana* and *Arabidopsis*, Tobacco rattle virus (TRV)-for a large variety of hosts, brome mosaic virus (BMV)-for barley, rice and maize, poplar mosaic virus (PopMV-for poplar, and others. However, these are somewhat specific for each plant species, depending in part on the host range of the natural virus.

There are relatively few reports of viral infection in hemp and none of them have been characterized by modern genetic methods. Arabis mosaic virus, Cucumber mosaic virus, Alfalfa mosaic virus, Hemp mosaic virus and Hemp streak virus have all been reported in hemp (Spaar, Kleinhmpel and Fritzsche 1990). However, those isolates specifically infecting hemp are not available from any extant collection. Hemp streak virus and hemp mosaic virus, although reported (McPartland, Clarke and Watson 2000) were never characterized thoroughly, but numerous accessions for the other three virus from infections in other species of plant are held in the German Collection of Microorganisms and Cell Cultures (DSMZ).

The DSMZ collection has seven Arabis mosaic virus (ArMV) isolates which have a wide range of hosts. It causes necrotic or chlorotic lesions forming rings or mottles (ICTVdB Management 2006), and is known to cause stunting of plants. The lesions may disappear after infection leaving a stunted plant. ArMV genomic

material is contained in a hexagonal un-enveloped viral capsid around 25nm in diameter. Its genome is bipartite single-stranded positive sense RNA. The first segment is 9000 bp and the second is 5100 bp, and each is encapsulated separately. Some strains may have extra satellite RNA. Transmission of the virus usually occurs in nature by nematode-mediated infection, although mechanical inoculation, grafting, and infected seed are also reported to spread it (ICTVdB Management 2006). It is not spread by contact between host plants. The most common hosts used for propagation are *Nicotiana clevelandii*, and *Petunia x hybrid*.

Cucumber mosaic virus (CMV) has 33 isolates in the DSMZ collection. It is contained in an un-enveloped icosahedral capsid of approximately 29nm in diameter (ICTVdB Management 2006). It has a tripartite genome consisting of 4 or 5 segments of single-stranded RNA, which are divided between three separate capsids. It has a wide variety of hosts, and in nature is transmitted by members of the Aphididae family (aphids). It can also be transmitted via inoculation and sometimes through seeds. *Cucumis sativus*, *Nicotiana clevelandii*, *N. glutinosa*, and *N. tabacum* are the most common hosts for propagation.

There are 4 isolates of Alfalfa mosaic virus (AMV) in the DSMZ collection. AMV causes a variety of symptoms in different plants (ICTVdB Management 2006); for example *Nicotiana tabacum* develops chlorotic or necrotic lesions, and

Chenopodium amaranticolor develops chlorotic local lesions and systemic chlorotic or necrotic flecks. The common propagation hosts are *Nicotiana glutinosa* and *N. tabacum*. The viron is contained in a bacilliform or ellipsoidal un-enveloped capsule, the length of the capsule varies but the width is approximately 18nm. The genome is in 4 segments of linear, positive sense, single stranded RNA which are tripartite, either 8274 or 9155bp in size. AMV is transmitted by aphids, inoculation, grafting, and sometimes by seeds, and pollen; but not by contact between infected hosts.

To develop a suitable VIGs system a virus which infects *Cannabis* but shows few symptoms needs to be found. Accessions of AMV, ArMV, and CMV will be tested on a variety of *Cannabis* genotypes under a variety of inoculation conditions and from that a VIGs system will be developed.

4.2 Methods

4.2.1 Viral Propagation

Viral strains were received as freeze-dried, infected leaves from the German Collection of Microorganisms and Cell Cultures (DSMZ). The host species the strains were received in were not listed, however the plants each strain was isolated from was. Arabis Mosaic Virus (ArMV) PV-0232 (*Rubus* sp.) and PV-0045 (*Vitis vinifera*), Cucumber Mosaic Virus (CMV) PV-0453 (*Cucurbita pepo*) and PV-0037(*Nicotiana tabacum*), and Alfalfa Mosiac Virus (AMV) PV-0196 (*Solanum tuberosum*) were inoculated onto 6 leaf stage *Nicotiana benthamiana* by grinding the infected leaves in buffer in a 1:5 weight to volume ratio (ArMV – 50mM phosphate buffer pH 7.0 (Dijkstra and de Jager, Protocol 1 - Mechanical Inoculation of Plants 1998), AMV 50mM NaPO4 buffer pH 7.2 (Timmerman-Vaughan *et al.* 2001), CMV 20mM KPO4 pH 7.0 (Whitham *et al.* 2000)), and rubbing the inoculum onto leaves dusted with carborundum. Plants rubbed with buffer only and un-rubbed plants were used as controls. The rubbed leaves were rinsed with water after 10 minutes of exposure to inoculum, placed in a growth chamber and observed daily for visual appearance of viral symptoms. Four weeks after inoculation visually infected leaves were cut into approximately 1 cm square pieces and were placed over desiccant (Dry-rite) and stored for 1 week at 4ºC before being moved into room temperature storage (Dijkstra and de Jager, Protocol 21 - Dehydration 1998).

4.2.2 Development of Reverse Transcriptase (RT) PCR Method for Testing Virus Infection

RNA, for RT-PCR, was extracted from either fresh or dried leaf tissue using a Qiagen RNeasy kit: 1ug to 50ng total RNA, 0.5mM each coat protein primer or Oligo (dT)18; heated to 70ºC for 5 minutes then chilled on ice, 1mM dNTP, 20U Fermentas Ribolock Rnase Inhibitor, heated to 37ºC for 5 minutes then 20U Fermentas H minus M-MuLV Reverse Transcriptase was added and the solution was heated to 42ºC for one hour, then 70ºC for 10 minutes; 1 uL of the RT reaction was used in a PCR reaction containing 0.5 mM of each coat protein primer (Table 4-1) with the cycling conditions of: 95ºC 2 min; (95ºC 1min, 30sec gradient 50-70 ºC, 72C 1 min)30X; 72ºC 5 min. For both CMV and ArMV primer sets 65ºC was determined to be the optimal annealing temperature.

Sets of primers were designed based on coat proteins for viruses of the same species listed in Genbank. For multiple accessions of coat proteins one set of primers was designed for each accession, and tested on a small amount of the stock obtained from DSMZ.

Primer Name	Primer Sequence	Diagnostic Set Band Size	Coat protein Genbank accession
AMVCPF2	GCTGGTGGGAAAGC TGGTAAAC	483bp	NC_002025.1 GI:9626928
AMVCPR2	GGCTACGGCATAGG AATGCTTG		
ArMVF2	ACACTGTCTGTCCCT CATTGG	843bp	D10086.1 GI:221017
ArMVR2	CCTCGACCCTATCAC ATACTC		
CMVF2	AGAGTCTTGTCGCAG CAGCTTTCG	367bp	NC_001440.1 GI:9626472
CMVR2	ACTGATAAACCAGT ACCGGTGAGG		

Table 4-1: Coat protein primers used for RT-PCR confirmation of infection of *N. benthamiana* and *C. sativa* samples.

4.2.3 Inoculation via Carborundum

C. sativa genotype Carmen plants at the cotyledons, two-leaf, six-leaf, or 5 wk. stage were inoculated in a similar fashion to the propagation plants using small amounts of the viral inoculum from DSMZ or *N. benthamiana* propagated CMV PV-0453 rubbed onto leaves dusted with carborundum. To be certain that no asymptomatic infections were occurring all plants were tested; but to save time and money, fresh leaves from asymptomatic plants were pooled into pools of eight plants three weeks after inoculation. Any leaves of plants showing visual symptoms of infection were tested by RT-PCR individually as fresh samples and leaf samples were desiccated for storage. Un-inoculated samples were grown and tested alongside the inoculated plants as controls, and to check for potential insect transmission (Table 4-2), since hemp is prone to infection with thrips and

although the plants were treated with Orthene prior to inoculation there was still an underlying population in the growth chambers.

Inoculum	Two leaf stage hemp	Six leaf stage hemp	Five week old hemp	hemp Cotyledons
DSMZ CMV PV-0453	3 indiv.	3 indiv.	3 indiv.	3 indiv.
N. benthamiana CMV PV-0453	2 pools of 8 10 indiv.	2 pools of 8 10 indiv.	2 pools of 8 10 indiv.	2 pools of 8 10 indiv.
Carborundum + PO4 Buffer	1 pool of 8 1 indiv.	1 pool of 8 1 indiv.	1 pool of 8 1 indiv.	1 pool of 8 1 indiv.
No inoculation	1 indiv.	1 indiv.	1 indiv.	1 indiv.

Table 4-2: Number of individuals (indiv.) for each carborundum inoculation at each stage of development of *C. sativa* genotype Carmen.

A slightly altered version of this method was used to test a number of different varieties of hemp: Zolo11, Anka, UnikoB, Felina, Fedora, Camen, Finola, Kompolti, Silesa, Bilobrzeskie, and LKCSD.

4.2.4 Inoculation via Injection

30 plants of each of the nine varieties used above were inoculated via injection with propagated CMV PV-0453. The infected leaf matter was ground in a mortar and pestle to coarse fragments and 2mL of 20mM KPO4 pH 7.0 were added and the liquor placed in a 2mL syringe. For 1/3 of the plants of each genotype the Syringe was pressed gently to a 2week old leaf and some of the liquor was expelled (so that the leaf was unbroken). For 1/3 of the plants the leaf was injected with a 16-gauge needle. For the last 1/3 of the plants the stem was

injected with a 16-gauge needle. Six plants of each genotype were injected with water following a similar pattern (2 plants – w/o needle, 2 plants – w/needle to stem, 2 plants w/needle to leaf); and 2 plants of each genotype were left un-injected as controls. These plants were observed for symptoms over two weeks and then each genotype and treatment was pooled and tested via RT-PCR. Again, any plants showing symptoms were tested individually.

4.3 Results

As a first step towards identifying a virus that could be used as the basis for a VIGS system in hemp, freeze-dried leaves containing three viruses; AMV, CMV, and ArMV, were obtained from a stock center. Each of these viral species, but not these isolates, had previously been reported to infect hemp (McPartland, Clarke and Watson 2000), (Spaar, Kleinhmpel and Fritzsche 1990). For CMV and ArMV, two accessions were obtained, for AMV only one accession was available (Table 4-3).

Virus Abbreviation	Virus Name	Accession	Original Isolate Plant
AMV-0196	Alfalfa Mosaic Virus	PV-0196	*Solanum tuberosum* - potato
ArMV-0232	Arabis Mosaic Virus	PV-0232	*Rubus* sp. – raspberry
ArMV-0045	Arabis Mosaic Virus	PV-0045	*Vitis vinifera* - grape
CMV-0453	Cucumber Mosaic Virus	PV- 0453	*Cucurbita pepo* – zucchini
CMV-0037	Cucumber Mosaic Virus	PV-0037	*Nicotiana tabacum* – tobacco

Table 4-3: List of virus accessions and original isolate plants.

Because the stock center provides only a small amount of inoculum, it is necessary to propagate viruses on a susceptible host prior to infection trials on hemp. Two commonly used species for viral propagation are *Nicotiana benthamiana* and *Chenopodium album*. Initial infection via carborundum inoculation of *N. benthamiana* with AMV, CMV, and ArMV produced visually positive results for all five viral isolates. Infections in *Chenopodium album* (via carborundum), while producing a necrotic lesion at the site of inoculation, did not produce any systemic visual symptoms. In *N. benthamiana* symptomatic curled

inoculated leaves appeared 4 days after inoculation (leaves from mock inoculations (carborundum and buffer) were not curled). The inoculated leaves uncurled after 2 days, and new leaves developed a curled aspect with some mottling.

Fourteen days after inoculation the different strains showed different symptoms. CMV-0453-infected plants showed a mosaic pattern, and all the new leaves were curved and crinkled. CMV-0037-infected plants showed stunted growth, and a general lightening of new leaves caused by a mosaic of small spots; the old leaves were curled. New leaves on ArMV-0232-infected plants were larger than those of uninfected plants, had sunken veins with the start of vein lightening. ArMV-0045-infected plants had new leaves that were chlorotic along the veins, with some necrosis on the heavily infected leaves. AMV-0196-infected plants showed little in the way of symptoms aside from an increase in leaf tissue growth causing the edges of the leaf to look misshapen; but a single leaf from the AMV-0196-infected plant, not containing the inoculation site, developed a large chlorotic spot.

Twenty days post-inoculation CMV-0453 infected plants showed a systemic mosaic pattern. Plants infected with CMV-0037 showed a light mosaic on new leaves, but older leaves had recovered. ArMV-0232 infected plants developed exceptionally large leaves with a faint mosaic on all of them while ArMV-0045

leaves were chlorotic to necrotic, with the stem starting to show chlorosis. AMV-0196 infected plants showed chlorotic to necrotic lesions on some but not all of the leaves, and new leaves were very small, badly curled, and chlorotic. At this point, a selection of leaves were collected for RT-PCR and desiccated for storage. At 28 days post-inoculation a second collection of leaves was harvested and desiccated for storage.

To confirm the presence of viral RNA within the symptomatic leaf tissue, total RNA was extracted from three or four 1cm pieces of leaf tissue and attempted RT-PCR amplification of genes specific to each viral genome. As negative controls extracted RNA from leaves from mock-inoculated plants (carborundum and buffer), and a PCR water blank was used. As positive controls a small amount (a fragment of a leaf piece) was processed from the DSMZ sample. This PCR reaction was saved for future positive controls. CMV-0453 and the two types of ArMV from DSMZ were amplifiable with RT-PCR; but despite the symptoms described above, only CMV-0453 gave a positive test in *N. benthamiana* (Figure 4-1), all RT-PCR tests of *Chenopodium album* leaves gave negative results.

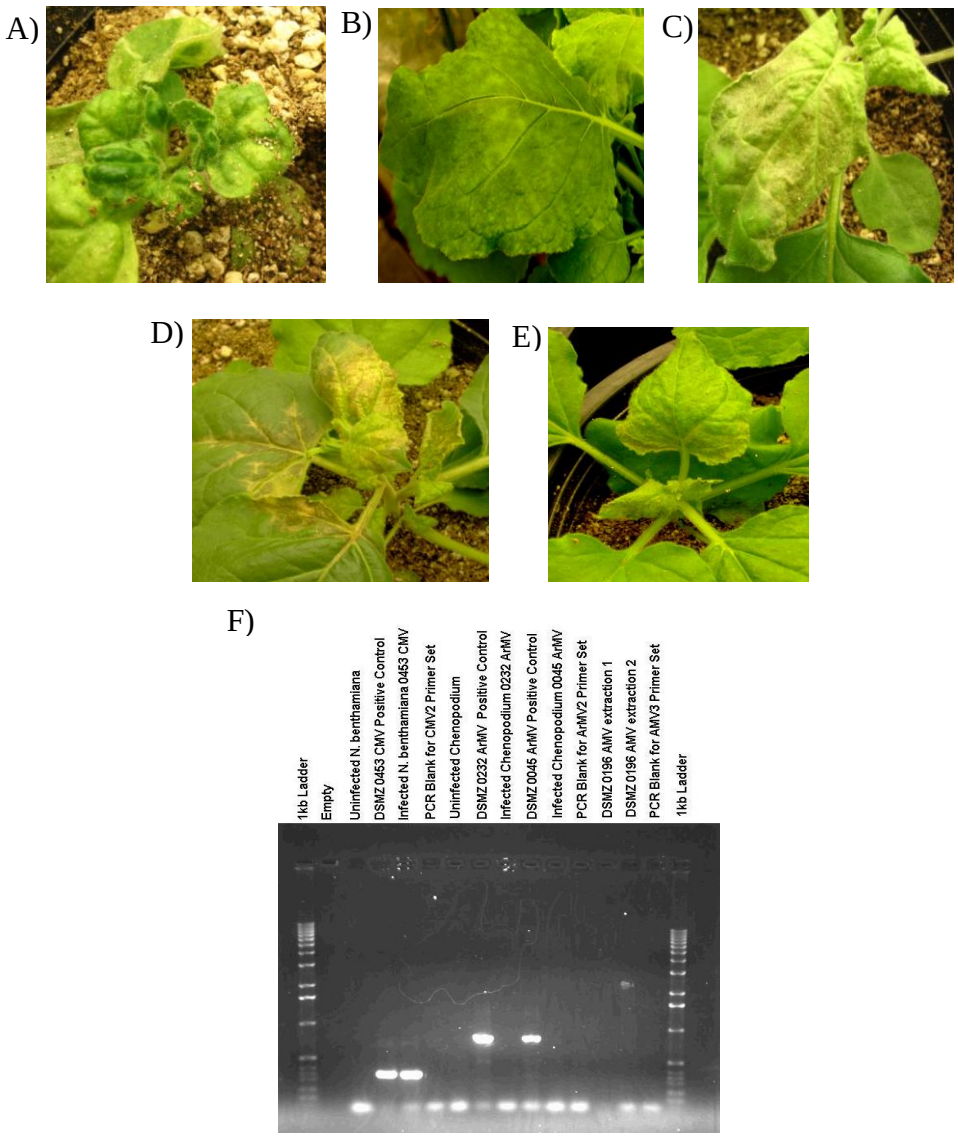


Figure 4-1: *N. benthamiana* showing visible signs of infection and RT-PCR results.
A) CMV PV-0453. B) ArMV PV-0232. C) AMV PV-0196. D) ArMV PV-0045. E) CMV PV-0037. F) Ethidium agarose gel of RT-PCR for viral infection. AMV and ArMV *N. benthamiana* RT-PCR tests (not shown) were negative for any bands.

To test the ability of the viruses to infect industrial hemp (*C. sativa* genotype Carmen), I inoculated plants using carborundum with sap from each of the five viral isolates that had been propagated by DSMZ. Three individual plants were inoculated at each of the following stages: cotyledon stage, 2 leaf stage, 6 leaf stage and 5 wk past germination (using different plants at each stage). Infected hemp plants from CMV-0453, CMV-0037, ArMV-0232, and AMV 0196 showed visible signs of systemic infection (Figure 4-2) and (Table 4-4) and leaves that were symptomatic but not inoculated were collected for RT-PCR. RT-PCR reactions of the hemp samples were negative indicating infection did not occur. Positive controls in most cases produced the correct PCR band. Positive controls included extracts of the DSMZ-derived plant tissues and sap samples from infected *N. benthamiana.* CMV-0453, ArMV-0233, and ArMV-0045 produced the correct size bands in the positive control; CMV-0037 and AMV-0196 produced no bands. Mock-inoculated hemp leaves, and a PCR water blank were used as negative controls, none of these negative controls produced bands.

Next, a larger sampling of 200 *C. sativa* plants infected with *N. benthamiana*-propagated CMV PV-0453 was tested by RT-PCR, to determine if the plants might show a level of susceptibility in the population that was not seen in the initial tests. Non-symptomatic plants were pooled into pools of eight individuals and systemic symptomatic plants were tested individually. *N. benthamiana*

propagated CMV PV-0453 used to infect the *C. sativa* as well as the original DSMZ sample were used as positive controls (Figure 4-3). 22 symptomatic plants and 20 pools were analyzed by RT-PCR; none showed bands indicating infection (results not shown). Controls used were the same as those used in previous experiment, positive controls produced bands, negative controls produced no bands.

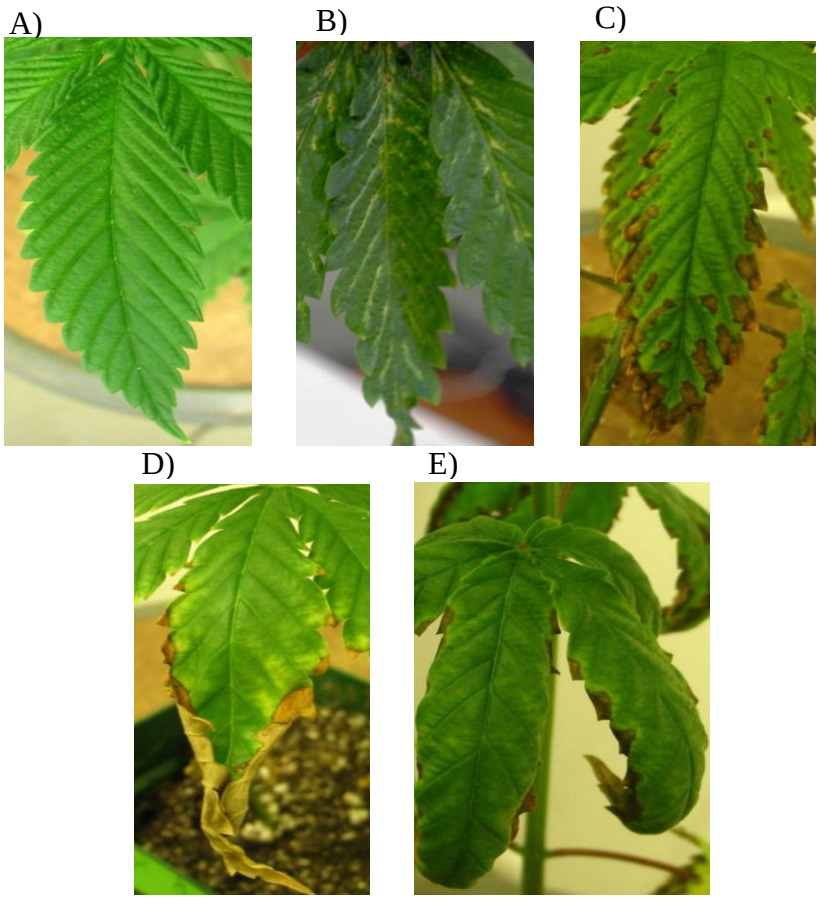


Figure 4-2: 5 wk. *C. sativa,* following inoculation showing viable signs of infection.
A) 5 wk. *C. sativa* uninfected. B) 5 wk. *C. sativa,* following inoculation with CMV PV-0453, C) 5 wk. *C. sativa,* following inoculation with ArMV PV-0232, D) 5 wk. *C. sativa*, following inoculation with AMV PV-0196. E) 5 wk. *C. sativa*, following inoculation with CMV PV-0037. Plants shown were inoculated at the 6-leaf stage of growth. Not shown: cotyledon, two, and six leaf stage infected plants showing similar symptoms to the larger plants, and ArMV PV-0045 which showed no symptoms.

Virus Name	Visual symptoms in *N. benthamiana*	Positive RT-PCR in *N. ben-thamiana*	Visual symptoms in *C. sativa*	Positive RT-PCR in *C. sativa*
CMV-0453	Leave curling Mosaic Leaf crinkling Stunting	Y	Leaf curling Mosaic Leaf chlorosis starting near the veins	N
CMV-0037	Mosaic on new leaves only	N	Leaf curling and mosaic	N
ArMV-0232	Mosaic Increased leaf growth	N	Leaf Necrosis starting at outer tips	N
ArMV-0045	Chlorosis of leaves and veins leading to necrosis	N	N	N
AMV-0196	Chlorosis or necrosis of the leaves	N	Leaf curling and necrosis at the leaf tips	N

Table 4-4: Table of viral infection results in *C. sativa*.
Y indicates that the RT-PCR reaction produced a band; N indicates that no bands were seen for the reaction or that no visual symptoms were observed. See Table 4-3 for full names and accessions for each virus.

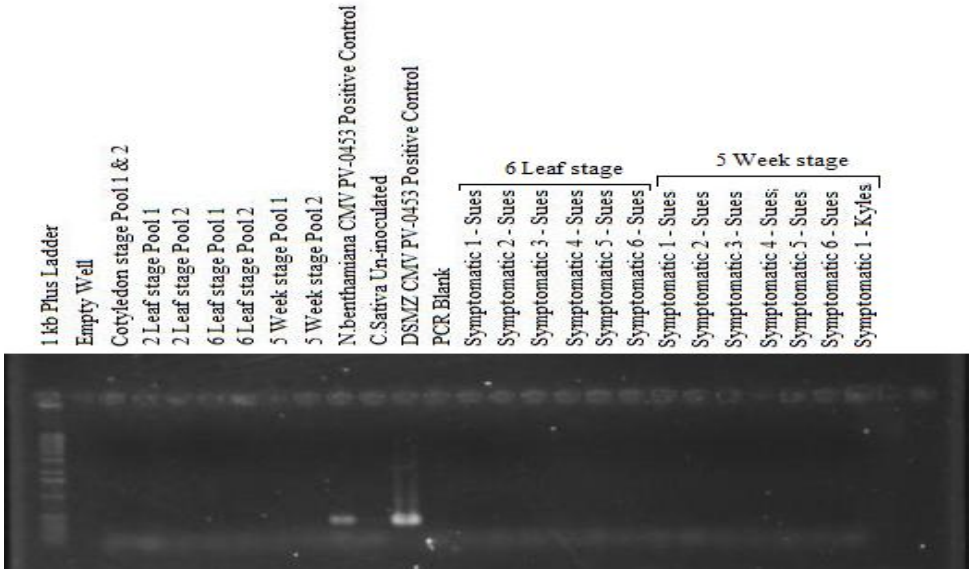


Figure 4-3: Ethidium agarose gel of RT-PCR test of *C. sativa* genotype Carmen leaf samples inoculated with *N. benthamiana* propagated CMV PV-0453. Positive controls of inoculum and DSMZ sample were run, as were negative controls of un-inoculated *C. sativa*, and a PCR blank.

Since viral resistance can be genotype dependent, I next attempted to infect a wide range of hemp germplasm. The varieties Zolo11 (Ukrainian genotype), Anka (Canadian genotype) UnikoB (Hungarian genotype), Felina (French genotype), Fedora (a cross between JUS 9 (Russian) and Fibrimon 21 (French)), Finola (of unknown origin), Kompolti (Hungarian genotype), and Silesa (of unknown origin) were inoculated with *N. benthamiana* propagated CMV PV-0453 (Table 4-5). Samples were pooled into 2 pools of 8 individuals and one pool of 8 un-inoculated plants for each genotype. Individuals showing systemic infections were processed separately. *N. benthamiana* was inoculated along with the *C. sativa*. Infection was determined by RT-PCR in *N. benthamiana* but not in any of the *C. sativa*; even though there were similar visual symptoms to previous trials (Figure 4-4).

Genotype	Number of visually symptomatic plants in the genotype tests/total inoculated
Anka (Canadian genotype)	0/36
UnikoB (Hungarian genotype),	1/35
Felina (French genotype),	0/36
Fedora (a cross between JUS 9 (Russian) and Fibrimon 21 (French)),	1/36
Finola (of unknown origin),	0/54
Kompolti (Hungarian genotype),	0/3
Silesa (of unknown origin)	3/54
Zolo11 (Ukrainian genotype),	1/17
Carmen (Canadian genotype)	1/53
LKCSD (directly selected from northern European land races)	8/8
Bialobrzeskie (Old Polish genotype registered)	7/7

Table 4-5: Number of symptomatic plants per total number of plants inoculated during the genotype test.
The Carmen and Silesia numbers do not include those inoculated in other tests.

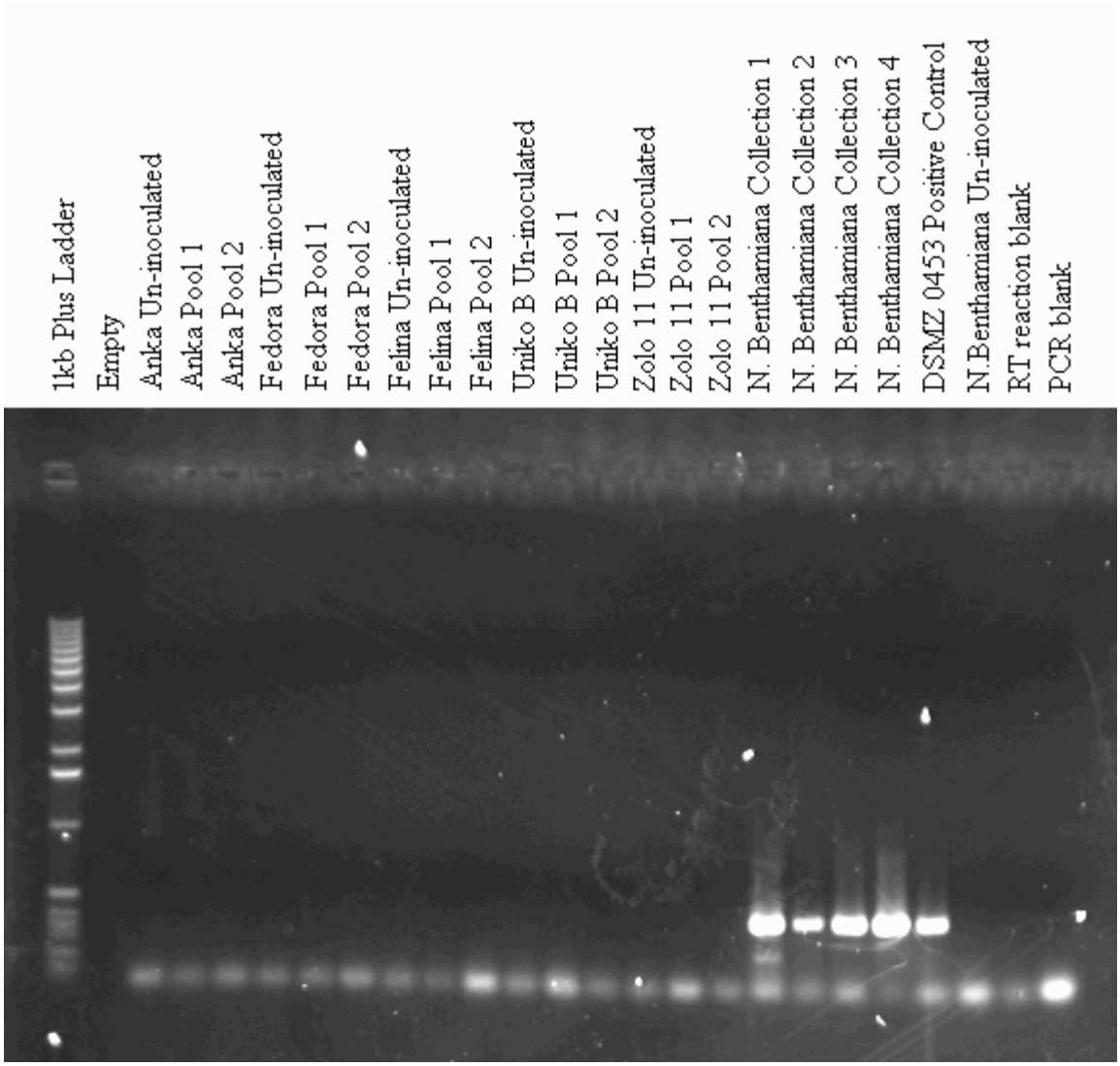


Figure 4-4: Pools of *C. sativa* varieties Anka, Fedora, Felina, Uniko B, and Zolo 11 inoculated with *N. benthamiana* propagated CMV PV-0453.
Along with *N. benthamiana* inoculated with the same *N. benthamiana* propagated CMV PV-0453 sap. Not shown: varieties Finola, Kompolti, and Silesa genotypes (all negative); and PCR of THC gene fragment used to check if RNA was amplifiable (all positive).

The varieties Anka and Silesia were then used to test different solutions and methods of inoculation. Each genotype at the 6 leaf stage was either inoculated on their leaves using Carborundum grit and water, or phosphate buffer; or via injection with a needle either to their stems or on their leaves using water or phosphate buffer. Leaf samples were collected once visual symptoms started appearing, and tested by RT-PCR. None of the samples showed positive results (Figure 4-5).

The varieties LKCSD (directly selected from northern European land races) and Bialobrzeskie (a Polish genotype registered in 1968), are two of the oldest varieties in the ARC collection, and were tested at the 6 lead stage using carborundum inoculations of *N. benthamiana* propagated CMV PV-0453. These both showed visual unambiguous symptoms of infection within four days of inoculation. Bialobrzeskie become completely chlorotic and died; and LKCSD showed a mottled and stunted appearance. RT-PCR tests again showed no positive results. This may be due to something within the hemp interfering with the RT-PCR results or with the replication of the virus RNA (Figure 4-6).

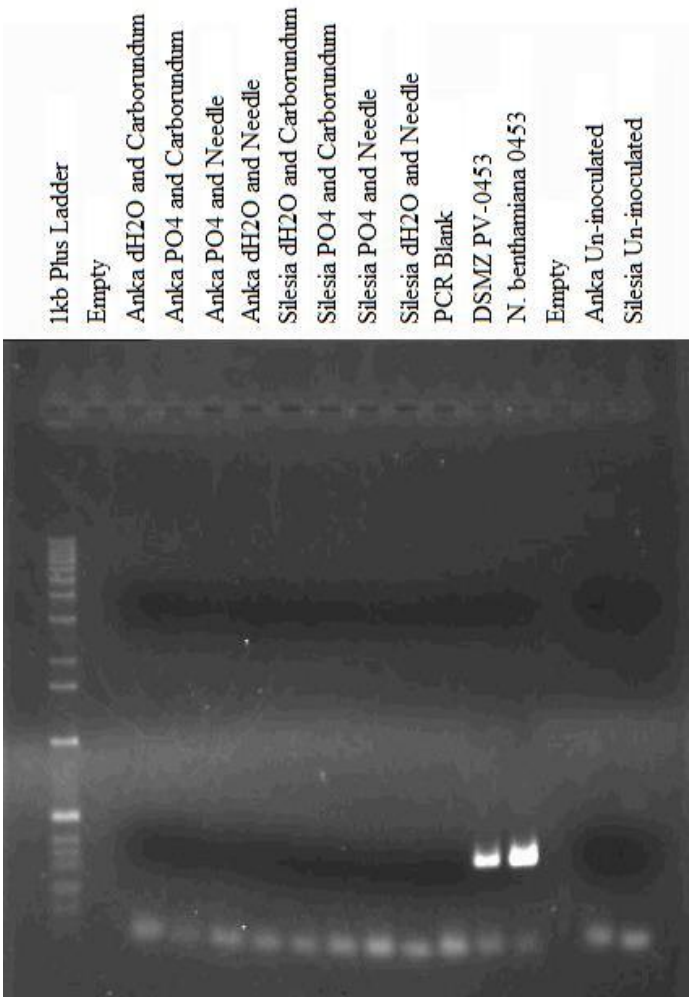


Figure 4-5: Ethidium agarose gel of RT-PCR results from pools of Anka or Silesia leaves inoculated by carborundum grit or by needle,
the various saps ground either in water or phosphate buffer. Not shown: Anka or Silesia stems inoculated by injection with a needle (also negative).

A) B) C)

D)

Figure 4-6: CMV-0453 inoculated LKCSD and Bialobrzeskie *Cannabis* varieties, A) Inoculated LKCSD B) Inoculated Bialobrzeskie C) Un-inoculated carborundum and phosphate buffer challenged control Bialobrzeskie D) Tray of test plants from left to right two rows of LKCSD inoculated plants, control carborundum and phosphate buffer challenged LKCSD (middle row), Bialobrzeskie inoculated plants, control carborundum and phosphate buffer challenged Bialobrzeskie.

4.4 Discussion

In an attempt to identify viral isolates that could infect *C. sativa*, I identified five isolates from five viruses that were able to produce visual symptoms of infection on *N. benthamiana*. The presence of viral RNA was confirmed by RT-PCR for CMV-0453 and ArMV-0232 and ArMV-0045 in the original innoculum obtained from the DSMZ stock center, but only CMV-0453 was subsequently confirmed in *N. benthamiana* that I had infected with DSMZ innoculum (Table 4-4). CMV-0037 and AMV-0196 probably had sequences too divergent from the available coat protein sequences for the RT-PCR to work properly, since positive results were not achieved in the DSMZ sap tests. ArMV-0232 and ArMV-0045 plants may have had an induced resistance that prevented the infection from either becoming established or systemic in the *N. benthamiana* plants I infected. Some alternate theories about why the RT-PCR reaction did not work are discussed below when considering why the RT-PCR did not work on any of the *Cannabis* samples.

Following inoculation of hemp plants of eleven different varieties, symptoms of infection were present in 8 of the varieties (Carmen, Silesa, Fedora, Zolo, Uniko, LKCSD, and Bialobrzeskie). LKCSD and Bialobrzeskie seemed most susceptible, while Anka, Felina, Finola, and Kompolti appear the least susceptible (Table 4-5). However, I was not able to detect viral RNA in any of the *C. sativa* tissues using RT-PCR. The inability to confirm viral infection in hemp by RT-PCR may have

numerous causes. First, none of the *C. sativa* may have been infected. The apparent symptoms seen might be due to environmental factors, insect damage, or a reaction of *C. sativa* to substances in the *N. benthamiana* leaves rather than to the virus propagating in the leaves. However, since un-inoculated and mock inoculated samples (carborundum and buffer), grown in parallel, did not show symptoms similar to inoculated plants, environmental or insect damage can largely be discounted. Futhermore, a reaction of *C. sativa* to substances in the *N. benthamiana* leaves is also unlikely since the inoculated plants showed systemic symptoms rather than just on the inoculated leaves. In individual cases inoculated plants did not show symptoms indicating that they were resistant to the virus or recovered from the inoculation quickly. This was particularly noticeable in the first set of genotype tests, where pools of non-symptomatic plants were tested.

Natural resistance in plants to viruses, called cross protection, can be triggered by infecting a plant with a virus that causes less severe symptoms in order to protect it from a virus that causes more severe symptoms (Stanford and Johnston 1985). If the seed stocks of these varieties were already carrying a virus they may have an increased resistance to infections by related viruses. A non-symptomatic hemp streak or hemp mosaic virus infection may be a possible source of resistance. Hemp mosaic virus was first described in 1949, and may be a strain of CMV or ArMV, but has not been characterized by modern techniques (McPartland, Clarke and Watson 2000). Hemp streak virus was first described in 1941 and caused

serious losses in Germany, Czechoslovakia, and Italy but has not been analyzed by modern methods (McPartland, Clarke and Watson 2000). Alternately, those varieties which have been bred after the 1940's may be from seed stock that has a natural resistance to viral infection; since hemp breeding has been traditional (as opposed to molecular) and the parentage of lines is not always clear, this information is not easily tracked.

Second, the virus or the host may be altering the viral coat protein gene (for which the primers are designed) once it infects *Cannabis* causing the RT-PCR reaction to fail. The inoculated leaves were not tested, as that could cause false positives in the RT-PCR reaction, so one other possibility is that *Cannabis* has only a local infection, and the symptoms seen throughout the plant is due to a cascade immune response (a severe resistance response) rather than the virus becoming systemic(Kang, Yeam and Jahn 2005).

It is also possible that the virus, at the time of sampling, is not producing large amounts of coat protein. Environmental changes, stress, light, heat, or a longer period before sampling can change the infectivity (Dijkstra and de Jager, Infectivity of virus in crude sap 1998).

There are other tests that could determine if the symptomatic plants are infected. ELISA and immuno-fluorescence tests (Lopez *et al.* 2003) usually react to coat

protein present in the plant tissue, though if the plants are not producing viral coat proteins this may produce similar results to the RT-PCR reactions. Microscopy to look for virion particles (Dijkstra and de Jager, Electron Microscope and Light Microscope Serology 1998) is another possibility, but again this requires that the virus is producing coat protein particles. Developing a set of RT-PCR primers that detect the movement proteins rather than coat protein is another possibility, as is the development of a nested RT-PCR system or use of qRT-PCR to improve sensitivity (Lopez *et al.* 2003).

Once the detection of viral infection can be achieved then the process of designing a VIGS system from the infecting virus can begin. Ideally, we need to decrease the symptomatic effects that damage the plant, and a silent infection is preferable. To do this only the genes necessary for infection are transformed into a binary Agrobacterium vector which is then used to infect the plant. The most developed system for this is the tobacco rattle virus (TRV)-VIGS vector system (Dinesh-Kumar *et al.* 2003). The viral RNA dependent RNA polymerase, the movement protein(s), a self cleaving ribozyme, and the coat protein are typically placed between the 2X35S promoter and a NOS terminator of binary vectors, dependent on the natural topology of the virus. In the case of the TRV-VIGS system the virus is bipartite, so the parts are placed into two binary vectors according to their viral configuration. Into one vector a multiple cloning site is placed, so that a target gene fragment can easily be inserted. Both of these vectors are then

infiltrated into the test plant; in *N. benthamiana*, the TRV-VIGS vectors are infiltrated when it is at the six leaf stage, and in tomatoes the vectors are infiltrated at the two leaf stage.

Chapter 5

Summary and Conclusions

5 Summary and Conclusions

In this thesis, I identified through microarray analysis a set of transcripts that were enriched in bast fiber-containing tissues within the vicinity of the snap point. These transcripts had increased abundance in this tissue compared to epidermis and xylem (PEX-S) from the same internode. I also identified in a separate set of experiments transcripts that were enriched in bast-fiber producing phloem at the snap-point as compared to tissues above the snap-point and below the snap-point (SAB-P). The genes that intersected in both studies were consistent with the restructuring of cell walls and potential changes in cell walls during intrusive growth: ATP Sulfurylase, xyloglucan endotransglycosylase, xyloglucan endotransglucosylase, S-adenosylmethionine decarboxylase, cytochrome P450, glutathione transferase, polyubiquitin, ubiquitin, arabinogalactan protein, fasciclin-like AGP 15, 11, 6, 5, and 2, DUF 250, dehydrin 1, aquaporin, stress-related - rubber elongation factor protein, and DUF 642. Phenylpropanoid genes (alcohol dehydrogenase, cinnamate-4-hydroxylase, cinnamoyl-CoA reductase, and aldehyde dehydrogenase), and wound-induced WI12, a protein which acts in cell wall reinforcement, were also up-regulated in the microarray data.

From the qRT-PCR of the available hemp phenylpropanoid ESTs plus BGAL and CESA, I determined alignments, phylogenetic trees and expression profiles for C4H, F5H, COMT, 4CL, BGAL, 2 isomers of CCR, and 3 isomers of CESA. Of these, COMT2, F5H, and BGAL appear to be the best targets for fiber alteration in hemp because they all show marked increases in expression at the snap point and are related to known genes affecting fiber development in other plant species. Two of them (COMT2, and F5H) are currently being used to screen mutant populations in through ARCs biofibers reverse genetics hemp TILLinG platform.

The development of a VIGS platform; using AMV, ArMV, or CMV, in *Cannabis* was hampered by the inability to confirm viral infection in hemp via RT-PCR. Nevertheless, the two hemp varieties LKCSD, and Bialobrzeskie seemed to display quite strong visual symptoms when inoculated with CMV-0453. Further inoculation and RT-PCR tests on Bialobrzeskie and LKCSD using all the viral strains discussed in this thesis will be tried at the University of Alberta, and from there a VIGS system will be developed in hemp.

Future work to be done, aside from that listed above is to continue qRT-PCR experiments on the remaining isomers from the list of current phenylpropanoid and secondary cell wall genes, and to add WI12, and the regulatory genes MYB, LIM and EIN3 to the list. Determining more sequence information for the ESTs of interest, isolating full length mRNA via the Rolling circle amplification,

currently underway at ARC, are required for more precise qRT-PCR results. Functional analysis of the proteins on substrates from the lignin pathway of some of the isomers with unclear functions, especially 4CL1, is important before proceeding with development of lines containing alterations to the particular protein.

References

6 References

Anterola, A.M., and N.G. Lewis. "Trends in lignin modification: a comprehensive analysis of the effects of genetic manipulations/mutations on lignification and vascular integrity." *Phytochemistry* 61, no. 3 (2002): 221-294.

Applied Biosystems. "Relative Standard Curve Method for Quantification." In *Real-Time PCR Systems Chemistry Guide*, 3-31 - 3-36. Applera Corporation, 2004.

Bate, N.J., J. Orr, W. Ni, A. Meromi, T. Nadler-Hassar, and *et al.* "Quantitative relationship between phenylalanine ammonia-lyase levels and phenylpropanoid accumulation in transgenic tobacco identifies a rate-determining step in natural product synthesis." *Proc. Natl. Acad. Sci. USA* 91 (1994): 7608-12.

Baucher, M., B. Chabbert, G. Pilate, J. Van Doorsselaere, M.T. Tollier, and *et al.* "Red xylem and higher liginin extractability by down-regulating a cinnamyl alcohol dehydrogenase in poplar (Populus tremula x P.alba)." *Plant Physiology* 112 (1996): 1479-90.

Baucher, M., B. Monties, M. Van Montagu, and W. Boerjan. "Biosynthesis and Genetic Engineering of Lignin." *Critical Reviews in Plant Sciences* 17, no. 2 (1998): 125–197.

Blake, A. W., S.E. Marcus, J.E. Copeland, R.S. Blackburn, and J.P. Knox. "In situ analysis of cell wall polymers associated with phloem fiber cells in stems of hemp, *Cannabis sativa* L." *Planta* 228, no. 1 (2008): 1–13.

Boerjan, W., J Ralph, and M. Baucher. "Lignin Biosynthesis." *Annu. Rev. Plant Biol.* 54 (2003): 519-546.

Bosca, I. "Genetic improvement: conventional approaches." In *Advances in hemp research*, edited by P. Ranalli, 153-184. New York: Binghamton, 1999.

Burch-Smith, T. M., J. C. Anderson, G. B. Martin, and S. P. Dinesh-Kumar. "Applications and advantages of virus-induced gene silencing for gene function studies in plants." *The Plant Journal* 39 (2004): 734-746.

Burch-Smith, T. M., M. Schiff, Y. Liu, and S.P. Dinesh-Kumar. "Efficient virus-induced gene silencing in *Arabidopsis*." *Plant Physiology* 142 (2006): 21-27.

Chabannes, M., *et al.* "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." *The Plant Journal* 28, no. 3 (2001): 271-282.

Chapple, C.C.S., T. Vogt, B.E. Ellis, and C.R. Somerville. "An *Arabidopsis* mutant defective in the general phenylpropanoid pathway." *The Plant Cell* 4 (1992): 1413-1423.

Chaudhary, N. "Industrial hemp production in Canada." *Government of Alberta Agricultural and Rual Development.* Edited by G. Atkinson. December 4, 2008. http://www1.agric.gov.ab.ca/$department/deptdocs.nsf/all/econ9631 (accessed August 12, 2009).

Chaudhary, N. "Industrial Hemp Production in Canada." *Agriculture and Rural Development. Alberta.ca.* 2005, Revised 2008. http://www1.agric.gov.ab.ca/$department/deptdocs.nsf/all/econ9631 (accessed August 9, 2009).

Chen, F., and R.A. Dixon. "Lignin modification improves fermentable sugar yields for biofuel production." *Nature Biotechnology* 25, no. 7 (2007): 759-761.

Clarke, R.C. "Botany of the genus *Cannabis*." In *Advances in Hemp Research*, edited by P. Ranalli, 1-19. New York: Binghamton, 1999.

Coffey, C. "Unpublished Results." 2008.

Costa, M.A., *et al.* "Characterization in vitro and in vivo of the putative multigene 4-coumarate:CoA ligase network in *Arabidopsis*: syringyl lignin and sinapate/sinapyl alcohol derivative formation." *Phytochemistry* 66 (2005): 2072-2091.

Crônier, D., B Monties, and B. Chabbert. "Structure and Chemical Composition of Bast Fibers Isolated from Developing Hemp Stem." *J. Agric. Food Chem.* 53, no. 21 (2005): 8279–8289.

de Meijer, E.P.M. "Fiber hemp cultivars: A survey of origin, ancestry, availability and brief agronomic characteristics." *J. Int. Hemp Assoc.* 2, no. 2 (1995): 66-73.

de Pauw, M.A., J.J. Vidmar, J. Collins, R.A. Bennett, and M.K. Deyholos. "Microarray analysis of bast fiber producing tissues of *Cannabis sativa* identifies transcripts associated with concerved and specialised processes of secondary wall development." *Functional Plant Biology* 34 (2007): 737-749.

Dear, M., and S.P. Koziel. "Unpublished Observations." 2006.

Dijkstra, J., and C.P. de Jager. "Electron Microscope and Light Microscope Serology." In *Practical Plant Virology: Protocols and Exercises*, by J. Dijkstra and C.P. de Jager, 377-400. Berlin: Springer - Verlag, 1998.

Dijkstra, J., and C.P. de Jager. "Infectivity of virus in crude sap." In *Practical Plant Virology: Protocols and Exercises*, by J. Dijkstra and C.P. de Jager, 99-111. Berlin: Springer-Verlag, 1998.

Dijkstra, J., and C.P. de Jager. "Protocol 1 - Mechanical Inoculation of Plants." In *Practical plant virology: protocols and exercises*, by J. Dijkstra and C.P. de Jager, 5-13. Berlin: Springer - Verlag, 1998.

Dijkstra, J., and C.P. de Jager. "Protocol 21 - Dehydration." In *Practical plant virology: protocols and exercises*, by J. Dijkstra and C.P. de Jager, 194-197. Berlin: Springer - Verlag, 1998.

Dinesh-Kumar, S.P., R. Anandalakshmi, R. Marathe, M. Schiff, and Y. Liu. *Virus-Induced Gene Silencing.* Vol. 236, in *Methods in Molecular Biology: Plant Functional Genomics*, edited by E. Grotewold, 287-293. Totowa, NJ: Humana Press, Inc., 2003.

Dunky, M. "Application of Adhesives: Adhesives in the wood industry." In *Handbook of adhesive technology*, edited by A. Pizzi and K. L. Mittal, 887-956. New York: Marcel Deckkeer, 2003.

Eck, R.V., and M.O. Dayhoff. *Atlas of Protein Sequence and Structure.* Maryland: National Biomedical Research Foundation, Silver Springs, 1966.

Feeney, M., and Z.K. Punja. "Tissue culture and argrobacterium-mediated transformation of hemp (*Cannabis sativa* L.)." *In Vitro Cellular and Developmental Biology – Plant* 39, no. 6 (2003): 578-585.

Felby, C., J. Hassingboe, and M. Lund. "Pilot-scale production of fiberboards made by laccase oxidized wood fibers: board properties and evidence for cross-linking of lignin." *Enzyme and Microbial Technology* 31 (2002): 736-741.

Felsenstein, J. "Confidence limits on phylogenies: An approach using the bootstrap." *Evolution* 39 (1985): 783-791.

Franke, R., C.M. McMichael, K. Meyer, A.M. Shirley, J.C. Cusumano, and C. Chapple. "Modified lignin in tobacco and poplar plants over-expressing the *Arabidopsis* gene encoding ferulate 5-hydroxylase." *Plant Journal* 22 (2000): 223-34.

Franke, R., M.R. Hemm, J.W. Denault, M.O. Ruegger, J.M. Humphreys, and C. Chapple. "Changes in secondary metabolism and deposition of an unusual lignin in the ref8 mutant of *Arabidopsis*." *Plant Journal* 30 (2002): 47-49.

Gene Codes Corporation, Inc. *Sequencher version 4.9.* Ann Arbour, MI, 1991-2009.

Gorshkova, T.A., and C. Morvan. "Secondary cell-wall assembly in flax phloem fibers: role of galactans." *Planta* 223 (2006): 149-158.

Guo, D., Chen F., J. Wheeler, and J. Selman S, *et al.* Winder. "Improvement of in-rumen digestability of alfalfa forage by genetic manipulation of lignin O-methyltransferases." *Transgenic Res.* 10 (2001): 457-64.

Guo, D., F. Chen, K. Inoue, J.W. Blount, and R.A. Dixon. "Downregulation of caffeic acid 3-O-methyltransferase and caffeoyl CoA 3-O-methyltransferase in transgenic alfalfa: impacts on lignin structure and implications for the biosynthesis of G and S lignin." *Plant Cell* 13 (2001): 73-88.

Health Canada. "Industrial Hemp Regulations, List Of Approved Cultivars For The 2007 Growing Season *Cannabis sativa* L." *Health Canada. Healthy*

Environments And Consumer Safety Branch. 2007. http://www.hc-sc.gc.ca/dhp-mps/substancontrol/hemp-chanvre/comm-licen/index-eng.php (accessed 2009).

Higuchi, T., T. Ito, T. Umezawa, T. Hibino, and D. Shibata. "Red-brown coloration of lignified tissues of transgenic plants with antisense CAD gene: wine-red lignin from coniferaldehyde." *J. Biotechnolol.* 37 (1994): 151-58.

Hu, W.J., S.A. Harding, J. Lung, J.L., Ralph, J. Popko, and *et al.* "Repression of lignin biosynthesis promotes cellulose accumulation and growth in transgenic trees." *Nature Biotechnology* 17 (1999): 808-12.

ICTVdB. *The Universal Virus Database, version 4.* http://www.ncbi.nlm.nih.gov/ICTVdb/ICTVdB/ (accessed 2009).

ICTVdB Management. "00.010.0.01.001. Alfalfa mosaic virus." *ICTVdB - The Universal Virus Database, version 4.* Edited by C. Büchen-Osmond. Columbia University, New York, USA. 2006. (accessed 2009).

ICTVdB Management. "00.010.0.04.001. Cucumber mosaic virus." *ICTVdB - The Universal Virus Database, version 4.* Edited by C. Büchen-Osmond. Columbia University, New York, USA. 2006. (accessed 2009).

ICTVdB Management. "00.018.0.03.002. Arabis mosaic virus." *ICTVdB - The Universal Virus Database, version 4.* Edited by C. Büchen-Osmond. Columbia University, New York, USA. 2006. (accessed 2009).

Jones, L., A.R. Ennos, and S.R. Tunner. "Cloning andcharacterization of irregular xylem4 (irx4): a severely lignin-deficient mutant of *Arabidopsis*." *Plant J.* 26 (2001): 205-16.

Jouanin, L., T. Goujon, V. de Nadai, M.T. Martin, I. Mila, and *et al.* "Lignification in transgenic poplars with extremely reduced caffeic acid O-methyltransferase activity." *Plant Physiology* 123 (2000): 1363-73.

Kajita, S., S. Hishiyama, Y Tomimura, Y. Katayama, and S. Omori. "Structural characterization of modified lignin in transgenic tobacco plants in which activity of 4-coumarate: coenzyme A ligase is depressed." *Plant Physiology* 114 (1997): 871-79.

Kang, B., I. Yeam, and M. M. Jahn. "Genetics of plant virus resistance." *Annu. Rev. Phytopathol.* 43 (2005): 581-621.

Lapierre, C., B. Pollet, M. Petit-Conil, G. Toval, J. Romero, and *et al.* "Structural alterations of lignins in transgenic poplars with depressed cinnamyl alcohol

dehydrogenase or caffeic acid O-methyltransferase activity have opposite impact on the efficiency of industrial Kraft pulping." *Plant Physiology* 119 (1999): 153-63.

Lee, D., K. Meyer, C. Chapple, and C.J. Douglas. "Antisensesuppression of 4-counarate: coenzyme A ligase activity in *Arabidopsis* leads to altered lignin subunit composition." *Plant Cell* 9 (1997): 1985-98.

Li, X., J. Weng, and C. Chapple. "Improvement of biomass through lignin modification." *The Plant Journal* 54 (2008): 569–581.

Lopez, M. M., *et al.* "Innovative tools for detection of plant pathogenic viruses and bacteria." *Int. Microbiol.* 6 (2003): 233-243.

Mandolino, G., and P. Ranalli. "Advances in biotechnological approaches for hemp breeding and industry." In *Advances in Hemp Research*, edited by P. Ranalli, 185-212. NY: Binghamton, 1999.

McPartland, J.M., R.C. Clarke, and D.P. Watson. *Hemp diseases and pests: management and biological control.* New York: CABI Publishing, 2000.

Meyermans, H., Morreel K., Lapierre C., B. Pollet, and A. et al De Bruyn. "Modification in ligninand accumulation of phenolic glucosides in poplar xylem upon down-regulation of caffeoyl-coenzyme A O-methyltransferase, an enzyme involved in lignin biosynthesis." *J. Biol. Chem.* 275 (2000): 36899-909.

Mukherjee, A., *et al.* "Results of molecular analysis of an archaeological hemp (*Cannabis sativa* L.) DNA sample from North West China." *Genet. resour. crop evol.* 55, no. 4 (2008): 481-485.

Mutjè, P., A. López, M.E. Vallejos, J.P. López, and F. Vilaseca. "Full exploitation of *Cannabis sativa* as reinforcement/filler of thermoplastic composite materials." *Composites: Part A* 38 (2007): 369–377.

Nagel, J., *et al.* "EST analysis of hop glandular trichomes identifies and O-methyltransferase that catalyzes the biosynthesis of xanthohumol." *The Plant Cell* 20 (2008): 186-200.

NCBI. *Entrez Gene.* Aug 21, 2009. http://www.ncbi.nlm.gov/sites/entrez?Db=gene (accessed August 2009).

Nei, M., and S. Kumar. *Molecular Evolution and Phylogenetics.* . New York: Oxford University Press, 2000.

Pejic, B.M., M.M. Kostic, P.D. Skundric, and J.Z. Praskalo. "The effects of hemicelluloses and lignin removal on water uptake behavior of hemp fibers." *Bioresour. technol.* 99, no. 15 (2008): 7152-7159.

Pijlman, F. T. A., S.M. Rigter, J. Hoek, H. M. J. Goldschmidt, and R. J. M. Niesink. "Strong increase in total delta-THC in *Cannabis* preparations sold in Dutch coffee shops." *Addiction Biology* 10, no. 2 (2005): 171 – 180.

Piquemal, J., C. Lapierre, K. Myton, A. O'Connell, Schuch W., and *et al.* "Down-regulation in cinnamoyl-CoA reductase induces significant changes of lignin profiles in transgenic tobacco plants." *Plant Journal* 13 (1998): 71-83.

QIGEN Group. *RNeasy Mini Handbook.* Fourth. Qiagen, 2006.

Raes, J., A. Rohde, J.H. Christensen, Y. Van de Peer, and W. Boerjan. "Genome-wide characterization of the lignifications toolbox in Arabidosis." *Plant Physiology* 133 (2003): 1051-1071.

Reijmers, T.H., C. Maliepaard, H.C. van den Broeck, R.W. Kessler, A.J. Toonen, and H. van der Voet. "Integrated statistical analysis of cDNA microarray and NIR

spectoscopic data applied to a hemp dataset." *Journal of bioinformatics and computational biology* 3, no. 4 (2005): 891-913.

Roach, M.J., and M.K. Deyholos. "Microarray analysis of flax (Linum usitatissimum L.) stems identifies transcripts enriched in fiber-bearing phloem tissues." *Mol.Genet. Genomics* 278 (2007).

Rouison, D., M. Sain, and M. Couturier. " Resin transfer molding of hemp fiber composites: optimization of the process and mechanical properties of the materials." *Composites Science and Technology* 66 (2006): 895–906.

Russo, E.B., *et al.* "Phytochemical and genetic analyses of ancient *Cannabis* from Central Asia." *Journal of experimental botany* 59, no. 15 (2008): 4171-82.

Salnikov, V.V., M.V. Ageeva, and T.A. Gorshkova. "Homofusion of golgi secretory vesicles in flax phloem fibers during fromtation of the gelatinous secondary cell wall." *Protoplasma* 233, no. 3-4 (2008): 269-273.

Sewalt, V.J.H., W Ni, J.W. Blount, H.G. Jung, S.A. Masoud, and *et al.* "Reduced lignin content and altered lignin composition in transgenic tobacco down-regulated in expression of L-phenylalanine ammonia-lyase or cinnamate-4-hydrolase." *Plant Physiology* 115 (1997): 41-50.

Slaski, J. "Personal Communications." 2008.

Sneader, W. "Legacy Of The Past." In *Drug discovery: a history*, by W. Sneader, 8-11. Hoboken, N.J.: Wiley, 2005.

Spaar, D., H. Kleinhmpel, and R. Fritzsche. In *Ol-und Faserpflansen*, by D. Spaar, H. Kleinhmpel and R. Fritzsche, 220. Berlin: VEB Deutscher Landwirtschaftsverlag, 1990.

Stanford, J.C., and S.A. Johnston. "The concept of parasite-derived resistance - deriving resistance genes from the parasite's own genome." *J. Theor. Biol.* 113 (1985): 395-405.

Stewart, J.J., T. Akiyama, C. Chapple, J. Ralph, and S.D. Mansfield. "The effects on lignin structure of overexpression of ferulate 5-hydroxylase in hybrid Poplar." *Plant Physiology* 150 (2009): 621-635.

Tamura, K., J. Dudley, M. Nei, and S. Kumar. "MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0." *Molecular Biology and Evolution* 24 (2007): 1596-1599.

Thielemans, W., E. Can, S. S. Morye, and R. P. Wool. "Novel Applications of Lignin in Composite Materials." *Journal of Applied Polymer Science* 83 (2002): 323–331.

Timmerman-Vaughan, G.M., *et al.* "Partial Resistance of Transgenic Peas to Alfalfa Mosaic Virus under Greenhouse and Field Conditions." *Crop Science* 41 (2001): 846-853.

Unver, T., and H. Budak. "Virus-induced gene silencing, a post transcriptional gene silencing method." *International journal of plant genomics,* 2009: 1-8.

van den Broeck, H.C., C. Maliepaard, M.J.M. Ebskamp, M.A.J. Toonen, and A.J. Koops. "Differential expression of genes involved in C1 metabolism and lignin biosynthesis in wooden core and bast tissues of fiber hemp (*Cannabis sativa* L.)." *Plant Science* 174, no. 2 (2008): 205-220.

Van Zant, K. L., T. III Webb, G. M. Peterson, and R. G. Baker. "Increased *Cannabis*/Humulus Pollen, an Indicator of European Agriculture in Iowa ." *Palynology* 3 (1979): 227-233.

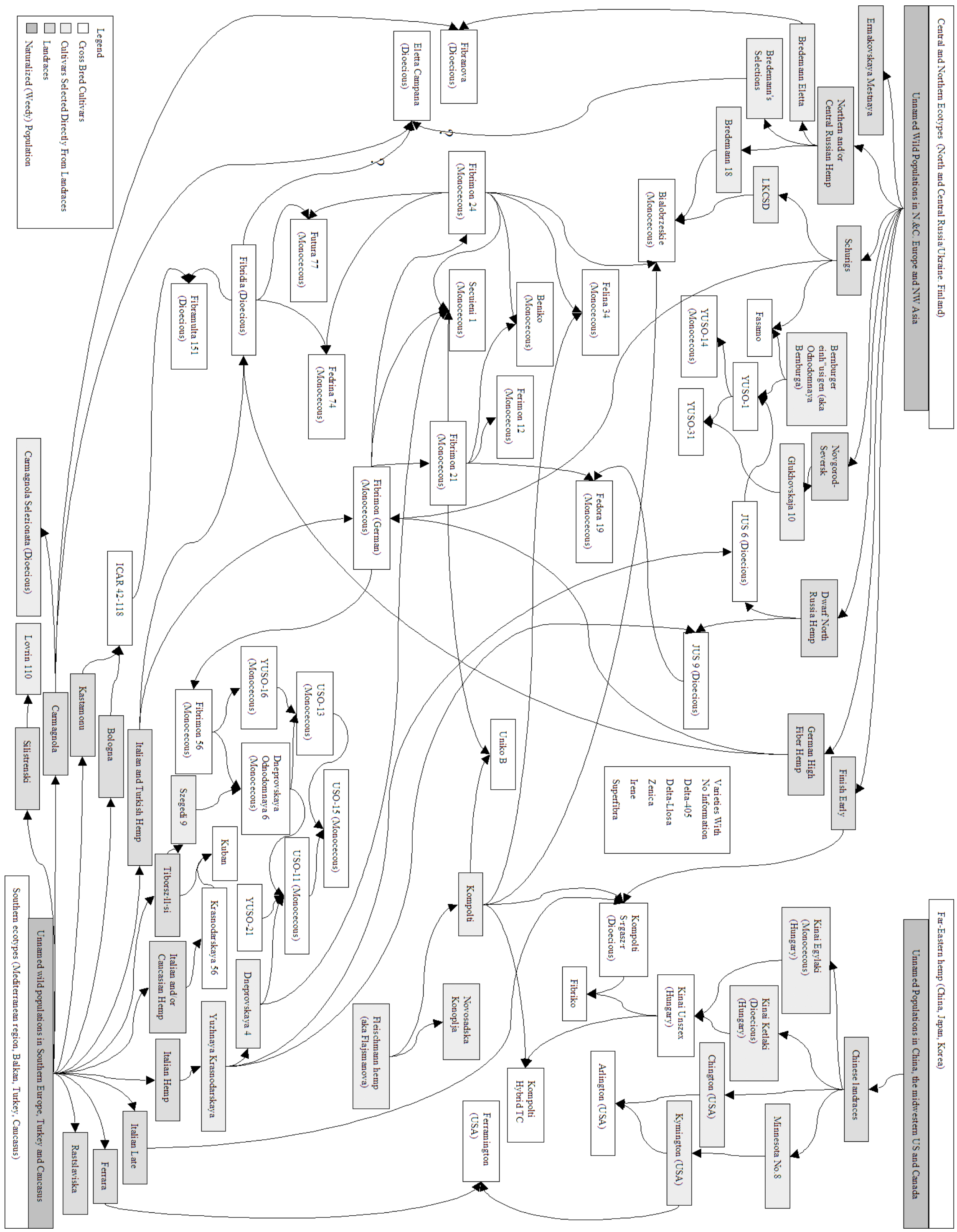
Voicu, M. "Unpublished results." 2009.

Whitham, S.A., R.J. Anderberg, S.T. Chisholm, and J.C. Carrington. "*Arabidopsis* RTM2 gene is necessary for specific restriction of tobacco etch virus and encodes an unusual small heat shock-like protein." *Plant Cell* 12, no. 4 (2000): 569-82.

Zhong, R., W.H. III Morrison, D.S. Himmelsbach, F.L. II Poole, and Z-H Ye. "Essential role of caffeoyl coenzyme A O-methyltransferase in lignin biosynthesis in woody poplar plants." *Plant Physiology* 124 (2000): 563-77.

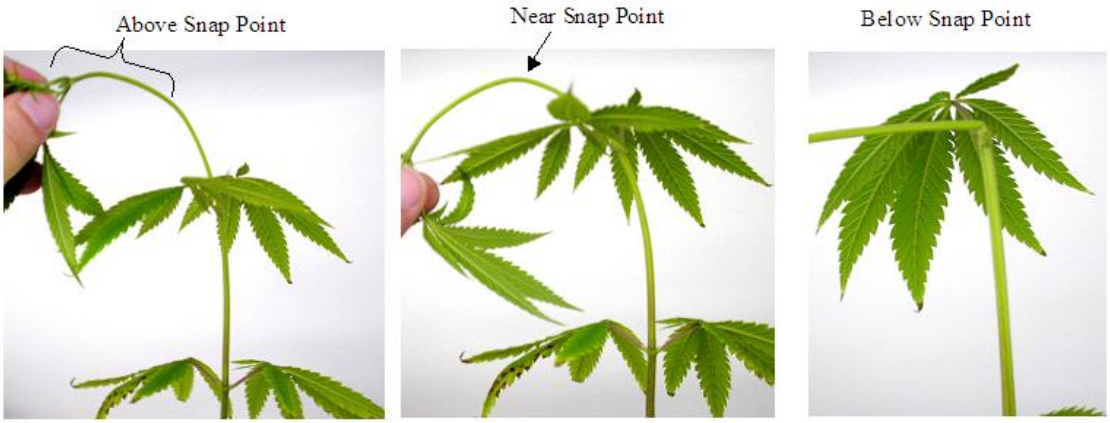
Appendices

Appendix 1 – European Fiber hemp Varieties



(Chart Based on Information from de Meijer, 1995)

Appendix 2 – Snap Point Determination



The flexibility of the plant as you bend the internode containing the snap point at different spots. The area of flexibility can also be felt as a change in the stiffness of the stem of plant. The nodes below the one containing the snap-point all tend to kink

Appendix 3 – Genes Found in Snap Point Epidermis, Phloem, or Xylem

Genes with Expression Increased In Snap Point Epidermis

Annotation	Contig Designation	Cluster Description
6-phosphogluconolactonase	Contig_060410_21	Cluster 11: ESP>PSP>XSP
ABC transporter	0	Cluster 11: ESP>PSP>XSP
acetyl-CoA synthetase	0	Cluster 11: ESP>PSP>XSP
acyltransferase	0	Cluster 11: ESP>PSP>XSP
adenosylhomocysteinase	0	Cluster 11: ESP>PSP>XSP
ADP,ATP carrier-like protein	0	Cluster 4: ESP>XSP>=PSP
allantoinase	0	Cluster 11: ESP>PSP>XSP
alpha-D-xylosidase	Contig_060410_135	Cluster 11: ESP>PSP>XSP
aluminum-induced protein	Contig_060410_11	Cluster 11: ESP>PSP>XSP
arabinogalactan protein	Contig_060410_217	Cluster 11: ESP>PSP>XSP
beta-glucosidase	Contig_060410_148	Cluster 11: ESP>PSP>XSP
BURP domain-containing protein	0	Cluster 4: ESP>XSP>=PSP
cell-wall protein P8	Contig_060410_117	Cluster 4: ESP>XSP>=PSP
chitinase	Contig_060410_189	Cluster 11: ESP>PSP>XSP
chlorophyll a/b binding protein	Contig_060410_118	Cluster 11: ESP>PSP>XSP
cytochrome P450	0	Cluster 11: ESP>PSP>XSP
dehydrogenase E1 alpha subunit	0	Cluster 11: ESP>PSP>XSP
dihydrolipoamide dehydrogenase	0	Cluster 11: ESP>PSP>XSP
DNA packaging protein of prophage CP-933X	Contig_060410_48	Cluster 11: ESP>PSP>XSP

Annotation	Contig Designation	Cluster Description
DnaJ protein	0	Cluster 11: ESP>PSP>XSP
endo-1,4-beta-D-glucanase	0	Cluster 11: ESP>PSP>XSP
fiber protein Fb34 [*Gossypium barbadense*]	Contig_060410_121	Cluster 11: ESP>PSP>XSP
glyceraldehyde-3-phosphate dehydrogenase	Contig_060410_164	Cluster 11: ESP>PSP>XSP
glyceraldehyde-3-phosphate dehydrogenase	Contig_060410_97	Cluster 11: ESP>PSP>XSP
glycosyl transferase	0	Cluster 11: ESP>PSP>XSP
histone H2A	0	Cluster 11: ESP>PSP>XSP
H-protein	Contig_060410_47	Cluster 11: ESP>PSP>XSP
ion membrane transport protein	0	Cluster 4: ESP>XSP>=PSP
Iron-sulfur-dependent L-serine dehydratase single chain	0	Cluster 11: ESP>PSP>XSP
lipid transfer protein	Contig_060410_101	Cluster 11: ESP>PSP>XSP
lipid transfer protein	Contig_060410_169	Cluster 11: ESP>PSP>XSP
mitochondrial processing peptidase alpha	Contig_060410_89	Cluster 11: ESP>PSP>XSP
non-specific lipid transfer protein	Contig_060410_193	Cluster 4: ESP>XSP>=PSP, Cluster 11: ESP>PSP>XSP
oligouridylate binding protein	0	Cluster 4: ESP>XSP>=PSP
pectinacetylesterase precursor	0	Cluster 11: ESP>PSP>XSP
pectinesterase	0	Cluster 11: ESP>PSP>XSP
peroxidase	Contig_060410_155	Cluster 4: ESP>XSP>=PSP
peroxidase	Contig_060410_225	Cluster 11: ESP>PSP>XSP

Annotation	Contig Designation	Cluster Description
peroxidase	Contig_060410_59	Cluster 11: ESP>PSP>XSP
peroxiredoxin	0	Cluster 11: ESP>PSP>XSP
photoassimilate-responsive protein PAR-1b-like protein	Contig_060410_128	Cluster 4: ESP>XSP>=PSP
photosystem 1 subunit 5	Contig_060410_173	Cluster 11: ESP>PSP>XSP
phytochelatin synthetase	Contig_060410_54	Cluster 4: ESP>XSP>=PSP
proline rich APG	Contig_060410_110	Cluster 4: ESP>XSP>=PSP
proline rich ARG-like	0	Cluster 4: ESP>XSP>=PSP
putative diphosphonucleotide phosphatase [*Oryza sativa* (japonica cultivar-group)]	0	Cluster 11: ESP>PSP>XSP
putative potassium transporter AtKT2p [*Arabidopsis thaliana*]	nd	Cluster 11: ESP>PSP>XSP
putative protein transport	0	Cluster 11: ESP>PSP>XSP
RD22	Contig_060410_171	Cluster 4: ESP>XSP>=PSP
RING finger protein	0	Cluster 11: ESP>PSP>XSP
senescence-associated protein 5	0	Cluster 4: ESP>XSP>=PSP
sterol C-24 reductase	Contig_060410_137	Cluster 11: ESP>PSP>XSP
syntaxin of plants 71 [*Arabidopsis thaliana*]	0	Cluster 11: ESP>PSP>XSP
TetR	0	Cluster 4: ESP>XSP>=PSP
thiamin biosynthetic enzyme	0	Cluster 11: ESP>PSP>XSP
transcription factor: KANADI GARP-like putative transcription factor KANADI4	0	Cluster 11: ESP>PSP>XSP

Annotation	Contig Designation	Cluster Description
auxin-repressed protein	Contig_060410_63	Cluster 11: ESP>PSP>XSP
xyloglucan endotransglucosylase hydrolase	Contig_060410_206	Cluster 11: ESP>PSP>XSP
zinc finger protein	Contig_060410_184	Cluster 11: ESP>PSP>XSP

Genes With Expression Increased In Snap Point Fiber/Phloem

Annotation	Contig Designation	Cluster Description
(1-4)-beta-mannan endohydrolase	Contig_060410_108	Cluster 9: PSP>XSP>ESP
4-coumarate-CoA ligase	0	Cluster 10: PSP>XSP=ESP
acid phosphatase	Contig_060410_163	Cluster 5: PSP>ESP=XSP
adenosine nucleotide translocator	0	Cluster 9: PSP>XSP>ESP
agglutinin	0	Cluster 9: PSP>XSP>ESP
aquaporin	0	Cluster 10: PSP>XSP=ESP
aquaporin	Contig_060410_143	Cluster 12: PSP>XSP>=ESP
arabinogalactan protein	Contig_060410_178	Cluster 10: PSP>XSP=ESP
arabinogalactan protein	Contig_060410_217	Cluster 5: PSP>ESP=XSP
ATP sulfurylase [*Brassica juncea*]	0	Cluster 12: PSP>XSP>=ESP
aux/IAA protein	Contig_060410_177	Cluster 1: PSP>ESP>XSP, Cluster 5: PSP>ESP=XSP
auxin growth promoter protein	Contig_060410_17	Cluster 5: PSP>ESP=XSP
auxin growth promoter protein	Contig_060410_18	Cluster 5: PSP>ESP=XSP
BAH (Bromo adjacent homology) domain-containing protein	0	Cluster 10: PSP>XSP=ESP
basic chitinase [*Nicotiana tabacum*]	Contig_060410_222	Cluster 1: PSP>ESP>XSP
beta-1,3 glucanase	0	Cluster 5: PSP>ESP=XSP
beta-D-galactosidase	0	Cluster 5: PSP>ESP=XSP
beta-ketoacyl-CoA-synthase	0	Cluster 5: PSP>ESP=XSP
caffeic acid O-methyltransferase	0	Cluster 12: PSP>XSP>=ESP, Cluster 5: PSP>ESP=XSP

Annotation	Contig Designation	Cluster Description
catalase	Contig_060410_214	Cluster 1: PSP>ESP>XSP
cinnamin acid 4-hydroxylase (C4H)	Contig_060410_45	Cluster 12: PSP>XSP>=ESP, Cluster 10: PSP>XSP=ESP
cop-coated membrane protein	0	Cluster 9: PSP>XSP>ESP
cryptochrome 1	0	Cluster 1: PSP>ESP>XSP
cystein protease	Contig_060410_125	Cluster 12: PSP>XSP>=ESP
Cytochrome c, class I	Contig_060410_104	Cluster 10: PSP>XSP=ESP
cytochrome P450	0	Cluster 5: PSP>ESP=XSP, Cluster 10: PSP>XSP=ESP
cytosolic 6-phosphogluconate dehydrogenase	0	Cluster 10: PSP>XSP=ESP
dehydrin 1	Contig_060410_42	Cluster 9: PSP>XSP>ESP
dhn1	Contig_060410_172	Cluster 10: PSP>XSP=ESP
DnaJ protein	0	Cluster 5: PSP>ESP=XSP, Cluster 10: PSP>XSP=ESP
DnaJ-like protein	0	Cluster 12: PSP>XSP>=ESP
EIN3-binding F-box protein	0	Cluster 10: PSP>XSP=ESP
EIN3-like protein	0	Cluster 5: PSP>ESP=XSP
elongation factor 1-alpha	Contig_060410_218	Cluster 12: PSP>XSP>=ESP
endo-xyloglucan transferase	Contig_060410_221	Cluster 1: PSP>ESP>XSP
ERD4 protein	0	Cluster 5: PSP>ESP=XSP
expansin 3	0	Cluster 5: PSP>ESP=XSP
fasciclin-like AGP	Contig_060410_107	Cluster 10: PSP>XSP=ESP
fasciclin-like AGP	Contig_060410_159	Cluster 10: PSP>XSP=ESP

Annotation	Contig Designation	Cluster Description
fasciclin-like AGP	Contig_060410_201	Cluster 10: PSP>XSP=ESP, Cluster 12: PSP>XSP>=ESP
ferritin	0	Cluster 12: PSP>XSP>=ESP
fiber protein Fb19 [*Gossypium barbadense*]	0	Cluster 10: PSP>XSP=ESP
flagellar P-ring protein	0	Cluster 1: PSP>ESP>XSP
glucose-6-phosphate/phosphate-translocator	0	Cluster 5: PSP>ESP=XSP
glutathione transferase	0	Cluster 1: PSP>ESP>XSP, Cluster 5: PSP>ESP=XSP
glyceraldehyde-3-phosphate dehydrogenase	Contig_060410_97	Cluster 10: PSP>XSP=ESP
glycosyl hydrolase family 5 protein / cellulase family protein [*Arabidopsis thaliana*]	Contig_060410_192	Cluster 10: PSP>XSP=ESP
heat shock protein	nd	Cluster 12: PSP>XSP>=ESP
hsc70	0	Cluster 10: PSP>XSP=ESP
LEA protein	0	Cluster 12: PSP>XSP>=ESP
lipid transfer like protein [*Vigna unguiculata*]	nd	Cluster 10: PSP>XSP=ESP
mitochondrial F1 ATP synthase beta subunit [*Arabidopsis thaliana*]	nd	Cluster 12: PSP>XSP>=ESP, Cluster 10: PSP>XSP=ESP
monodehydroascorbate reductase	0	Cluster 9: PSP>XSP>ESP, Cluster 10: PSP>XSP=ESP
NADH dehydrogenase	0	Cluster 12: PSP>XSP>=ESP
NAM-like TF	0	Cluster 9: PSP>XSP>ESP

Annotation	Contig Designation	Cluster Description
pectin acetylesterase	Contig_060410_197	Cluster 9: PSP>XSP>ESP, Cluster 10: PSP>XSP=ESP, Cluster 5: PSP>ESP=XSP
pectinacetylesterase precursor	Contig_060410_197	Cluster 5: PSP>ESP=XSP
pectinesterase	0	Cluster 5: PSP>ESP=XSP
peroxidase	Contig_060410_103	Cluster 5: PSP>ESP=XSP
phosphoribosylanthranilate transferase	0	Cluster 12: PSP>XSP>=ESP
polyubiquitin	0	Cluster 10: PSP>XSP=ESP
polyubiquitin	Contig_060410_150	Cluster 9: PSP>XSP>ESP
polyubiquitin	Contig_060410_208	Cluster 9: PSP>XSP>ESP, Cluster 12: PSP>XSP>=ESP
Proline rich protein	0	Cluster 9: PSP>XSP>ESP
P-type H+-ATPase	0	Cluster 12: PSP>XSP>=ESP
putative steroid membrane binding protein [*Oryza sativa* (japonica cultivar-group)]	0	Cluster 10: PSP>XSP=ESP
pyrophosphatase	0	Cluster 9: PSP>XSP>ESP, Cluster 10: PSP>XSP=ESP
quinone oxidoreductase	0	Cluster 10: PSP>XSP=ESP
S-adenosyl-L-methionine synthetase	Contig_060410_116	Cluster 5: PSP>ESP=XSP
S-adenosylmethionine decarboxylase	Contig_060410_120	Cluster 10: PSP>XSP=ESP, Cluster 12: PSP>XSP>=ESP
S-adenosylmethionine decarboxylase	Contig_060410_190	Cluster 10: PSP>XSP=ESP, Cluster 12: PSP>XSP>=ESP
stress-related	0	Cluster 5: PSP>ESP=XSP,

Annotation	Contig Designation	Cluster Description
		Cluster 10: PSP>XSP=ESP
sulfate adenylyltransferase	Contig_060410_119	Cluster 10: PSP>XSP=ESP, Cluster 9: PSP>XSP>ESP
TDP-glucose-4,6-dehydratase	Contig_060410_4	Cluster 10: PSP>XSP=ESP
TetR	0	Cluster 5: PSP>ESP=XSP
UBC E2	Contig_060410_46	Cluster 12: PSP>XSP>=ESP
ubiquitin carrier	0	Cluster 12: PSP>XSP>=ESP
ubiquitin homolog	Contig_060410_58	Cluster 9: PSP>XSP>ESP, Cluster 10: PSP>XSP=ESP
UDP-glycosyltransferase	0	Cluster 10: PSP>XSP=ESP
uircase	0	Cluster 5: PSP>ESP=XSP
xyloglucan endotransglycosylase	0	Cluster 9: PSP>XSP>ESP, Cluster 5: PSP>ESP=XSP
xyloglucan endotransglycosylase	Contig_060410_206	Cluster 1: PSP>ESP>XSP
zinc-finger protein	Contig_060410_184	Cluster 9: PSP>XSP>ESP
zinc-finger protein 1	0	Cluster 5: PSP>ESP=XSP

Genes With Expression Increased In Snap Point Xylem/Core

Annotation	Contig Designation	Cluster Description
(1-4)-beta-mannan endohydrolase	Contig_060410_108	Cluster 7: XSP=PSP>=ESP
(1-4)-beta-mannan endohydrolase	Contig_060410_162	Cluster 2: XSP>ESP>PSP
21kd polypeptide	Contig_060410_224	Cluster 7: XSP=PSP>=ESP
40S ribosomal protein	0	Cluster 7: XSP=PSP>=ESP
ABC transporter	0	Cluster 2: XSP>ESP>PSP, Cluster 7: XSP=PSP>=ESP
actin-depolymerizing factor 2	Contig_060410_202	Cluster 6: XSP>PSP>ESP, Cluster 2: XSP>ESP>PSP, Cluster 8: XSP=PSP>=ESP
adenosine nucleotide translocator	0	Cluster 7: XSP=PSP>=ESP
adenosylhomocysteinase [*Arabidopsis thaliana*]	nd	Cluster 6: XSP>PSP>ESP
Adenosylhomocysteinase 2 (S-adenosyl-L-homocysteine hydrolase 1) (SAH hydrolase 2) (AdoHcyase 2)	0	Cluster 8: XSP=PSP>=ESP
alcohol dehydrogenase	0	Cluster 2: XSP>ESP>PSP, Cluster 7: XSP=PSP>=ESP
alpha tubulin 1 [*Pseudotsuga menziesii* var. menziesii]	0	Cluster 2: XSP>ESP>PSP
alpha-expansin 1 [*Populus tremula* x *Populus tremuloides*]	nd	Cluster 6: XSP>PSP>ESP
amino acid permease	0	Cluster 7: XSP=PSP>=ESP
annexin 1 (?)	Contig_060410_24	Cluster 6: XSP>PSP>ESP
aquaporin	Contig_060410_143	Cluster 7: XSP=PSP>=ESP, Cluster 8: XSP=PSP>=ESP

Annotation	Contig Designation	Cluster Description
aquaporin	Contig_060410_167	Cluster 8: XSP=PSP>=ESP
aquaporin	Contig_060410_212	Cluster 6: XSP>PSP>ESP, Cluster 8: XSP=PSP>=ESP
aquaporin	Contig_060410_27	Cluster 3: XSP>PSP=ESP
aquaporin	Contig_060410_7	Cluster 6: XSP>PSP>ESP, Cluster 8: XSP=PSP>=ESP
arabinogalactan protein	0	Cluster 3: XSP>PSP=ESP
arabinogalactan protein	Contig_060410_217	Cluster 6: XSP>PSP>ESP
ascorbate oxidase	0	Cluster 6: XSP>PSP>ESP
ascorbate peroxidase	0	Cluster 7: XSP=PSP>=ESP
ATP-dependent RNA helicase	0	Cluster 6: XSP>PSP>ESP
beta-tubulin 1	Contig_060410_115	Cluster 2: XSP>ESP>PSP, Cluster 8: XSP=PSP>=ESP
blue copper protein	Contig_060410_176	Cluster 3: XSP>PSP=ESP
caffeic acid O-methyltransferase	Contig_060410_170	Cluster 6: XSP>PSP>ESP, Cluster 7: XSP=PSP>=ESP
caffeic acid O-methyltransferase	Contig_060410_199	Cluster 6: XSP>PSP>ESP, Cluster 2: XSP>ESP>PSP, Cluster 7: XSP=PSP>=ESP
caffeic O-methyltransferase	Contig_060410_168	Cluster 3: XSP>PSP=ESP
catalase	Contig_060410_214	Cluster 7: XSP=PSP>=ESP
CBS1	0	Cluster 2: XSP>ESP>PSP
CCAAT-binding transcription factor subunit	0	Cluster 8: XSP=PSP>=ESP
cellulose synthase	0	Cluster 2: XSP>ESP>PSP

Annotation	Contig Designation	Cluster Description
chitinase	Contig_060410_198	Cluster 3: XSP>PSP=ESP
chlorophyll a /b binding protein [*Beta vulgaris*]	Contig_060410_142	Cluster 2: XSP>ESP>PSP
chlorophyll a/b binding protein	Contig_060410_220	Cluster 6: XSP>PSP>ESP, Cluster 7: XSP=PSP>=ESP
chlorophyll a/b binding protein	Contig_060410_226	Cluster 7: XSP=PSP>=ESP
cobalamine-independent methionine synthase [*Solenostemon scutellarioides*]	0	Cluster 2: XSP>ESP>PSP, Cluster 6: XSP>PSP>ESP
cysteine proteinase RD19A	0	Cluster 7: XSP=PSP>=ESP
cytochrome b5 isoform	0	Cluster 3: XSP>PSP=ESP
disulphide isomerase PDI	0	Cluster 7: XSP=PSP>=ESP
elongation factor 1-alpha	Contig_060410_216	Cluster 8: XSP=PSP>=ESP
elongation factor 1-alpha	Contig_060410_218	Cluster 6: XSP>PSP>ESP, Cluster 7: XSP=PSP>=ESP, Cluster 8: XSP=PSP>=ESP
endo-1,4-beta glucanase	0	Cluster 3: XSP>PSP=ESP
senescence-associated protein	Contig_060410_196	Cluster 6: XSP>PSP>ESP, Cluster 8: XSP=PSP>=ESP
F1 ATPase	0	Cluster 2: XSP>ESP>PSP
fasciclin-like AGP	0	Cluster 3: XSP>PSP=ESP
fatty acid 9-hydroperoxide lyase	0	Cluster 8: XSP=PSP>=ESP
ferulate 5-hydroxylase	0	Cluster 8: XSP=PSP>=ESP
fiber protein Fb34 [*Gossypium barbadense*]	Contig_060410_121	Cluster 2: XSP>ESP>PSP
fis1 [*Linum*]	Contig_060410_111	Cluster 7: XSP=PSP>=ESP

Annotation	Contig Designation	Cluster Description
fructose 1,6-bisphosphate aldolase	Contig_060410_154	Cluster 6: XSP>PSP>ESP
fructose-bisphosphate aldolase	0	Cluster 8: XSP=PSP>=ESP
fructose-bisphosphate aldolase	Contig_060410_188	Cluster 6: XSP>PSP>ESP
FVE	0	Cluster 2: XSP>ESP>PSP
gamma-tocopherol methyltransferase	0	Cluster 2: XSP>ESP>PSP
GDP-D-mannose-4,6-dehydratase	0	Cluster 6: XSP>PSP>ESP, Cluster 8: XSP=PSP>=ESP
glutamate decarboxylase	0	Cluster 2: XSP>ESP>PSP
glyceraldehyde 3-phosphate dehydrogenase	Contig_060410_165	Cluster 6: XSP>PSP>ESP, Cluster 8: XSP=PSP>=ESP
glycine decarboxylase complex H-protein	0	Cluster 2: XSP>ESP>PSP
glycosyltransferase	0	Cluster 3: XSP>PSP=ESP, Cluster 2: XSP>ESP>PSP
hemolysin activator protein	Contig_060410_25	Cluster 2: XSP>ESP>PSP
hsp70	0	Cluster 7: XSP=PSP>=ESP
hydroxymethyltransferase	Contig_060410_16	Cluster 2: XSP>ESP>PSP
HyPRP	Contig_060410_64	Cluster 3: XSP>PSP=ESP
immunophilin	0	Cluster 2: XSP>ESP>PSP
kinase: protein kinase	0	Cluster 2: XSP>ESP>PSP
kinase: receptor like kinase	Contig_060410_75	Cluster 8: XSP=PSP>=ESP
leucine-rich reporter protein	0	Cluster 7: XSP=PSP>=ESP
L-lactate dehydrogenase	0	Cluster 8: XSP=PSP>=ESP

Annotation	Contig Designation	Cluster Description
mannan endo-1,4-beta-mannosidase	0	Cluster 7: XSP=PSP>=ESP
mannosyltransferase family protein	0	Cluster 6: XSP>PSP>ESP
metallothionein-like protein [*Pyrus pyrifolia*]	0	Cluster 8: XSP=PSP>=ESP
methionine synthase	Contig_060410_175	Cluster 2: XSP>ESP>PSP, Cluster 3: XSP>PSP=ESP
methionine synthase [*Zea mays*]	nd	Cluster 6: XSP>PSP>ESP, Cluster 8: XSP=PSP>=ESP
methylenetetrahydrofolate reductase	0	Cluster 8: XSP=PSP>=ESP
mitochondrial F1 ATP synthase beta subunit [*Arabidopsis thaliana*]	0	Cluster 2: XSP>ESP>PSP, Cluster 3: XSP>PSP=ESP, Cluster 6: XSP>PSP>ESP, Cluster 7: XSP=PSP>=ESP, Cluster 8: XSP=PSP>=ESP
myo-inositol 1-phosphate synthase	Contig_060410_109	Cluster 2: XSP>ESP>PSP
NADH:cytochrome b5 reductase	Contig_060410_73	Cluster 8: XSP=PSP>=ESP
NADP-isocitrate dehydrogenase	0	Cluster 8: XSP=PSP>=ESP
nicotinate-nucleotide adenylyltransferase	0	Cluster 3: XSP>PSP=ESP
nodulin	0	Cluster 3: XSP>PSP=ESP
nodulin	Contig_060410_124	Cluster 2: XSP>ESP>PSP, Cluster 6: XSP>PSP>ESP
nodulin 26	0	Cluster 2: XSP>ESP>PSP
nucleoid DNA-binding	Contig_060410_32	Cluster 6: XSP>PSP>ESP

Annotation	Contig Designation	Cluster Description
nucleotide-sugar dehydratase	Contig_060410_146	Cluster 2: XSP>ESP>PSP, Cluster 3: XSP>PSP=ESP
oxoglutarate malate translocator [*Solanum tuberosum*]	0	Cluster 2: XSP>ESP>PSP
peroxidase	Contig_060410_136	Cluster 3: XSP>PSP=ESP
peroxidase	Contig_060410_225	Cluster 8: XSP=PSP>=ESP
phosphoglycerate kinase, cytosolic	0	Cluster 2: XSP>ESP>PSP
photosystem II oxygen-evolving complex protein 3	0	Cluster 2: XSP>ESP>PSP
photosystem II type I chlorophyll a/b/ binding protein	Contig_060410_227	Cluster 6: XSP>PSP>ESP
phytochelatin synthetase	Contig_060410_54	Cluster 3: XSP>PSP=ESP
plastid division precursor	0	Cluster 2: XSP>ESP>PSP
PPase	Contig_060410_50	Cluster 2: XSP>ESP>PSP
profilin	Contig_060410_191	Cluster 8: XSP=PSP>=ESP
proline rich protein	Contig_060410_204	Cluster 2: XSP>ESP>PSP
protease: subtilisin	Contig_060410_62	Cluster 2: XSP>ESP>PSP, Cluster 6: XSP>PSP>ESP
proteasome regulatory particle triple-A ATPase subunit4 [*Oryza sativa* (japonica cultivar-group)]	0	Cluster 3: XSP>PSP=ESP
protein kinase [*Trifolium repens*]	nd	Cluster 7: XSP=PSP>=ESP
proton-inorganic pyrophosphatase	Contig_060410_140	Cluster 2: XSP>ESP>PSP
Pto kinase inhibitor	0	Cluster 7: XSP=PSP>=ESP
putative Argonaute (AGO1)	nd	Cluster 2: XSP>ESP>PSP,

Annotation	Contig Designation	Cluster Description
protein [*Arabidopsis thaliana*]		Cluster 7: XSP=PSP>=ESP
ribosomal protein L28-like [*Oryza sativa* (japonica cultivar-group)]	0	Cluster 7: XSP=PSP>=ESP
RING finger	0	Cluster 7: XSP=PSP>=ESP
S-adenosyl-L-homocysteine hydrolase	Contig_060410_151	Cluster 2: XSP>ESP>PSP, Cluster 8: XSP=PSP>=ESP
S-adenosyl-L-methionine synthetase	Contig_060410_116	Cluster 6: XSP>PSP>ESP, Cluster 3: XSP>PSP=ESP, Cluster 8: XSP=PSP>=ESP, Cluster 7: XSP=PSP>=ESP
senescence-associated protein 5	0	Cluster 2: XSP>ESP>PSP
STYLOSA	0	Cluster 6: XSP>PSP>ESP, Cluster 8: XSP=PSP>=ESP
sucrose synthase	Contig_060410_74	Cluster 6: XSP>PSP>ESP, Cluster 3: XSP>PSP=ESP
TCTP protein [Fragaria x ananassa]	nd	Cluster 7: XSP=PSP>=ESP
thionin	Contig_060410_156	Cluster 2: XSP>ESP>PSP
translationally controlled tumor protein	Contig_060410_224	Cluster 2: XSP>ESP>PSP, Cluster 8: XSP=PSP>=ESP
tubulin	Contig_060410_183	Cluster 2: XSP>ESP>PSP
tubulin alpha chain	Contig_060410_183	Cluster 3: XSP>PSP=ESP
tubulin beta-1	0	Cluster 3: XSP>PSP=ESP
tubulin beta-1	Contig_060410_145	Cluster 2: XSP>ESP>PSP
ubiquitin-specific protease 12	0	Cluster 7: XSP=PSP>=ESP
UDP-glycosyltransferase	0	Cluster 7: XSP=PSP>=ESP

Annotation	Contig Designation	Cluster Description
UMP/CMP kinase	Contig_060410_39	Cluster 6: XSP>PSP>ESP
xyloglucan endotransglycosylase/hydrolase 16 protein	Contig_060410_85	Cluster 2: XSP>ESP>PSP
zinc-finger protein	Contig_060410_184	Cluster 7: XSP=PSP>=ESP

Appendix 4 - Genes Found In Phloem Above, At, And Below The Snap Point

Genes With Expression Increased In Phloem Above Snap Point

Annotation	Contig Designation	Cluster Description
(1-4)-beta-mannan endohydrolase	Contig_060410_162	Cluster 11: PAS>=PBS>PSP
(1-4)-beta-mannan endohydrolase	Contig_060410_108	Cluster 12:PAS>=PSP>PBS
22kD-protein of PSII [*Spinacia oleracea*]	nd	Cluster 10: PAS>PSP>=PBS
23kDa polypeptide of the oxygen-evolving complex of photosystem II	Contig_060410_95	Cluster 10: PAS>PSP>=PBS, Cluster 11: PAS>=PBS>PSP
26S proteasome regulatory particle triple-A ATPase subunit4 [*Oryza sativa* (japonica cultivar-group)]	0	Cluster 10: PAS>PSP>=PBS
ABC transporter	0	Cluster 3: PAS>PSP>PBS
ABI3-interactin protein	0	Cluster 10: PAS>PSP>=PBS
acetoacyl-CoA-thiolase	0	Cluster 11: PAS>=PBS>PSP
acid phosphatase	Contig_060410_163	Cluster 3: PAS>PSP>PBS, Cluster 10: PAS>PSP>=PBS
alcohol dehydrogenase	0	Cluster 10: PAS>PSP>=PBS, Cluster 12:PAS>=PSP>PBS
alpha tubulin 1 [*Pseudotsuga menziesii* var. menziesii]	0	Cluster 11: PAS>=PBS>PSP
alpha-D-xylosidase	Contig_060410_135	Cluster 10: PAS>PSP>=PBS
aluminum-induced protein	Contig_060410_131	Cluster 12:PAS>=PSP>PBS
amino acid selective channel	0	Cluster 10: PAS>PSP>=PBS, Cluster

Annotation	Contig Designation	Cluster Description
protein		12:PAS>=PSP>PBS
arylsulfatase regulator	Contig_060410_149	Cluster 12:PAS>=PSP>PBS
ATP synthase	0	Cluster 10: PAS>PSP>=PBS
ATP:pyruvate phosphotransferase	0	Cluster 11: PAS>=PBS>PSP
ATPase	0	Cluster 12:PAS>=PSP>PBS
ATPase subunit B	0	Cluster 12:PAS>=PSP>PBS
auxin growth promoter protein	Contig_060410_17	Cluster 12:PAS>=PSP>PBS
auxin response factor	0	Cluster 11: PAS>=PBS>PSP
Bet1/Sft1-like SNARE AtBS14a [*Arabidopsis thaliana*]	0	Cluster 11: PAS>=PBS>PSP
beta-1,3-glucanase	0	Cluster 3: PAS>PSP>PBS
beta-D-galactosidase	0	Cluster 3: PAS>PSP>PBS
beta-N-acetylglucosaminidase-like	0	Cluster 12:PAS>=PSP>PBS
beta-tubulin 1	Contig_060410_115	Cluster 10: PAS>PSP>=PBS
blue copper protein	0	Cluster 11: PAS>=PBS>PSP
brassinosteroid biosynthetic protein LKB	Contig_060410_137	Cluster 10: PAS>PSP>=PBS, Cluster 11: PAS>=PBS>PSP
brassinosteroid receptor	0	Cluster 12:PAS>=PSP>PBS
caffeic acid O-methyltransferase	Contig_060410_170	Cluster 3: PAS>PSP>PBS, Cluster 10: PAS>PSP>=PBS
caffeic acid O-methyltransferase	Contig_060410_199	Cluster 3: PAS>PSP>PBS
carboxypeptidase type III	Contig_060410_23	Cluster 3: PAS>PSP>PBS

Annotation	Contig Designation	Cluster Description
cationic amino acid transporter	0	Cluster 12:PAS>=PSP>PBS
cellulose synthase	0	Cluster 12:PAS>=PSP>PBS
cell-wall protein P8	Contig_060410_117	Cluster 12:PAS>=PSP>PBS
chlorophyll a/b binding preprotein (AA - 32 to 231) [*Glycine max*]	0	Cluster 12:PAS>=PSP>PBS
chlorophyll a/b binding protein	0	Cluster 12:PAS>=PSP>PBS
chlorophyll a/b binding protein	Contig_060410_118	Cluster 10: PAS>PSP>=PBS
chlorophyll a/b binding protein	Contig_060410_142	Cluster 11: PAS>=PBS>PSP
chlorophyll a/b binding protein	Contig_060410_174	Cluster 10: PAS>PSP>=PBS, Cluster 12:PAS>=PSP>PBS
chlorophyll a/b binding protein	Contig_060410_207	Cluster 10: PAS>PSP>=PBS, Cluster 11: PAS>=PBS>PSP
chlorophyll a/b binding protein	Contig_060410_215	Cluster 10: PAS>PSP>=PBS, Cluster 11: PAS>=PBS>PSP, Cluster 12:PAS>=PSP>PBS
chlorophyll a/b binding protein	Contig_060410_220	Cluster 10: PAS>PSP>=PBS, Cluster 12:PAS>=PSP>PBS
chlorophyll a/b binding protein	Contig_060410_226	Cluster 10: PAS>PSP>=PBS, Cluster 11: PAS>=PBS>PSP, Cluster 12:PAS>=PSP>PBS
chlorophyll a/b binding protein type II [*Glycine max*]	0	Cluster 12:PAS>=PSP>PBS
chloroplast ferredoxin I	0	Cluster 10: PAS>PSP>=PBS

Annotation	Contig Designation	Cluster Description
cinnamoyl-CoA reductase	0	Cluster 3: PAS>PSP>PBS, Cluster 11: PAS>=PBS>PSP
cinnamyl-alcohol dehydrogenase	0	Cluster 12:PAS>=PSP>PBS
conserved hypothetical protein	0	Cluster 10: PAS>PSP>=PBS
CP29 Type I (26 kD) CP29 polypeptide	Contig_060410_207	Cluster 11: PAS>=PBS>PSP
dehydrogenase E1 alpha subunit	0	Cluster 11: PAS>=PBS>PSP
delta 12 oleic acid desaturase FAD2	0	Cluster 10: PAS>PSP>=PBS
desiccation-related protein	0	Cluster 10: PAS>PSP>=PBS
DNA packaging protein of prophage CP-933X	Contig_060410_48	Cluster 12:PAS>=PSP>PBS
dTDP-glucose 4-6-dehydratase	0	Cluster 11: PAS>=PBS>PSP
elongation factor 1-alpha	Contig_060410_218	Cluster 10: PAS>PSP>=PBS, Cluster 11: PAS>=PBS>PSP
embryo-defective 1025 [*Arabidopsis thaliana*]	nd	Cluster 10: PAS>PSP>=PBS, Cluster 12:PAS>=PSP>PBS
endo-1,4-beta-D-glucanase	0	Cluster 12:PAS>=PSP>PBS
endo-1,4-beta-glucanase	0	Cluster 12:PAS>=PSP>PBS
endosperm specific protein-like [*Arabidopsis thaliana*]	0	Cluster 10: PAS>PSP>=PBS
fasciclin-like arabinogalactan	Contig_060410_185	Cluster 3: PAS>PSP>PBS
fatty acid 9-hydroperoxide lyase	0	Cluster 12:PAS>=PSP>PBS
fiber protein Fb34 [*Gossypium barbadense*]	Contig_060410_121	Cluster 10: PAS>PSP>=PBS

Annotation	Contig Designation	Cluster Description
Flagellar L-ring protein	Contig_060410_26	Cluster 11: PAS>=PBS>PSP
galactinol synthase	0	Cluster 12:PAS>=PSP>PBS
GDSL family of lipolytic enzymes	0	Cluster 12:PAS>=PSP>PBS
GDSL-motif lipase/acylhydrolase	0	Cluster 12:PAS>=PSP>PBS
glutathione peroxidase	0	Cluster 12:PAS>=PSP>PBS
glyceraldehyde-3-phosphate dehydrogenase	Contig_060410_164	Cluster 3: PAS>PSP>PBS
glyceraldehyde-3-phosphate dehydrogenase	Contig_060410_165	Cluster 11: PAS>=PBS>PSP
glyceraldehyde-3-phosphate dehydrogenase	Contig_060410_29	Cluster 12:PAS>=PSP>PBS
glycerophosphoryl diester phosphodiesterase	0	Cluster 12:PAS>=PSP>PBS
glycine-rich protein-like [*Oryza sativa* (japonica cultivar-group)]	0	Cluster 12:PAS>=PSP>PBS
glycolate oxidase	0	Cluster 12:PAS>=PSP>PBS
glycosyl hydrolase family 5 protein / cellulase family protein [*Arabidopsis thaliana*]	Contig_060410_192	Cluster 3: PAS>PSP>PBS
glycosyltransferase [*Triticum aestivum*]	Contig_060410_102	Cluster 11: PAS>=PBS>PSP
granule-bound starch synthase	Contig_060410_51	Cluster 10: PAS>PSP>=PBS
grr1	0	Cluster 11: PAS>=PBS>PSP
hemolysin activator protein	Contig_060410_25	Cluster 12:PAS>=PSP>PBS
histidine-containing phosphotransfer protein 2	0	Cluster 12:PAS>=PSP>PBS

Annotation	Contig Designation	Cluster Description
histone H3	0	Cluster 3: PAS>PSP>PBS, Cluster 10: PAS>PSP>=PBS, Cluster 11: PAS>=PBS>PSP
H-protein	Contig_060410_47	Cluster 12:PAS>=PSP>PBS
hydrolase	0	Cluster 10: PAS>PSP>=PBS, Cluster 12:PAS>=PSP>PBS
importin alpha 1	Contig_060410_81	Cluster 11: PAS>=PBS>PSP
kinase: Pti1	0	Cluster 12:PAS>=PSP>PBS
kinase: receptor like kinase	Contig_060410_75	Cluster 11: PAS>=PBS>PSP
kinase: receptor like kinase	0	Cluster 10: PAS>PSP>=PBS, Cluster 12:PAS>=PSP>PBS
kinase: receptor protein-like	Contig_060410_79	Cluster 10: PAS>PSP>=PBS
laccase	0	Cluster 3: PAS>PSP>PBS
leucine zipper-containing protein	0	Cluster 10: PAS>PSP>=PBS
leucine zipper-containing protein	Contig_060410_35	Cluster 11: PAS>=PBS>PSP
light harvesting chlorophyll A/B binding protein [*Prunus persica*]	nd	Cluster 12:PAS>=PSP>PBS
Lissencephaly type-1-like homology motif; CTLH, C-terminal to LisH motif; Nitrous oxide reductase, N-terminal; WD40-like; Quinonprotein alcohol dehydrogenase-like	0	Cluster 12:PAS>=PSP>PBS
L-sorbosone dehydrogenase	0	Cluster 11: PAS>=PBS>PSP
major facilitator family transporter	0	Cluster 12:PAS>=PSP>PBS

Annotation	Contig Designation	Cluster Description
mandelonitrile lyase	0	Cluster 10: PAS>PSP>=PBS
meloidogyne-induced giant cell protein	0	Cluster 10: PAS>PSP>=PBS
metallophosphatase	Contig_060410_106	Cluster 3: PAS>PSP>PBS
metallophosphatase	Contig_060410_158	Cluster 3: PAS>PSP>PBS
methionine sulfoxide reductase	0	Cluster 12:PAS>=PSP>PBS
methionine synthase	0	Cluster 10: PAS>PSP>=PBS, Cluster 12:PAS>=PSP>PBS
methionine synthase	Contig_060410_175	Cluster 11: PAS>=PBS>PSP, Cluster 12:PAS>=PSP>PBS
methyltransferase	0	Cluster 12:PAS>=PSP>PBS
mitochondrial F1 ATP synthase beta subunit [*Arabidopsis thaliana*]	0	Cluster 3: PAS>PSP>PBS, Cluster 11: PAS>=PBS>PSP, Cluster 12:PAS>=PSP>PBS
mitochondrial processing peptidase alpha	Contig_060410_89	Cluster 12:PAS>=PSP>PBS
myo-inositol 1-phosphate synthase	Contig_060410_109	Cluster 10: PAS>PSP>=PBS, Cluster 11: PAS>=PBS>PSP
myo-inositol oxygenase	Contig_060410_2	Cluster 3: PAS>PSP>PBS
nodulin 26	0	Cluster 10: PAS>PSP>=PBS, Cluster 12:PAS>=PSP>PBS
nodulin-like	0	Cluster 12:PAS>=PSP>PBS
non-specific lipid transfer protein	Contig_060410_193	Cluster 3: PAS>PSP>PBS, Cluster 11: PAS>=PBS>PSP

Annotation	Contig Designation	Cluster Description
Orf122 like protein	0	Cluster 12:PAS>=PSP>PBS
outer envelope protein	0	Cluster 11: PAS>=PBS>PSP, Cluster 12:PAS>=PSP>PBS
pectinacetylesterase precursor	Contig_060410_197	Cluster 12:PAS>=PSP>PBS
pectinesterase	0	Cluster 10: PAS>PSP>=PBS, Cluster 12:PAS>=PSP>PBS
peroxidase	Contig_060410_225	Cluster 3: PAS>PSP>PBS
peroxidase	0	Cluster 11: PAS>=PBS>PSP
peroxidase	Contig_060410_59	Cluster 11: PAS>=PBS>PSP
peroxidase	Contig_060410_225	Cluster 12:PAS>=PSP>PBS
peroxiredoxin	0	Cluster 10: PAS>PSP>=PBS, Cluster 12:PAS>=PSP>PBS
phenylalanine ammonia-lyase2	0	Cluster 12:PAS>=PSP>PBS
phosphatase: purple acid	Contig_060410_106	Cluster 3: PAS>PSP>PBS, Cluster 10: PAS>PSP>=PBS
phosphoglycerate kinase	0	Cluster 11: PAS>=PBS>PSP, Cluster 12:PAS>=PSP>PBS
phosphoglycerate kinase, cytosolic	0	Cluster 11: PAS>=PBS>PSP
photoassimilate-responsive protein PAR-1b-like protein	Contig_060410_128	Cluster 12:PAS>=PSP>PBS
photosystem 1 subunit 5	Contig_060410_173	Cluster 12:PAS>=PSP>PBS
photosystem I chain XI precursor	Contig_060410_203	Cluster 12:PAS>=PSP>PBS

Annotation	Contig Designation	Cluster Description
photosystem II	Contig_060410_96	Cluster 12:PAS>=PSP>PBS
photosystem II type I chlorophyll a/b/ binding protein	Contig_060410_227	Cluster 12:PAS>=PSP>PBS
photosystem II: 33 kDa polypeptide of water-oxidizing complex of photosystem	Contig_060410_213	Cluster 10: PAS>PSP>=PBS, Cluster 11: PAS>=PBS>PSP
photosytem I subunit III precursor	Contig_060410_10	Cluster 12:PAS>=PSP>PBS
photosystem oxygen evolving complex 33 kDa photosystem II protein	Contig_060410_213	Cluster 10: PAS>PSP>=PBS, Cluster 12:PAS>=PSP>PBS
phytochelatin synthetase	Contig_060410_54	Cluster 10: PAS>PSP>=PBS, Cluster 11: PAS>=PBS>PSP
pollen-specific protein	Contig_060410_129	Cluster 12:PAS>=PSP>PBS
polygalacturonase	0	Cluster 11: PAS>=PBS>PSP
polygalacturonase inhibitor-like	Contig_060410_112	Cluster 3: PAS>PSP>PBS, Cluster 12:PAS>=PSP>PBS
polygalacturonase isoenzyme 1 beta subunit	nd	Cluster 12:PAS>=PSP>PBS
polypeptide 15 precursor	nd	Cluster 10: PAS>PSP>=PBS
polyubiquitin [*Populus tremula* x *Populus tremuloides*]	nd	Cluster 3: PAS>PSP>PBS, Cluster 10: PAS>PSP>=PBS
proline rich protein	0	Cluster 12:PAS>=PSP>PBS, Cluster 11: PAS>=PBS>PSP
PSI-H subunit [*Brassica rapa*]	nd	Cluster 12:PAS>=PSP>PBS
putative diphosphonucleotide phosphatase [*Oryza sativa* (japonica cultivar-group)]	0	Cluster 11: PAS>=PBS>PSP

Annotation	Contig Designation	Cluster Description
putative myo-inositol oxygenase [*Oryza sativa* (japonica cultivar-group)]	Contig_060410_223	Cluster 3: PAS>PSP>PBS, Cluster 10: PAS>PSP>=PBS
putative potassium transporter AtKT2p [*Arabidopsis thaliana*]	nd	Cluster 12:PAS>=PSP>PBS
putative protein transport	0	Cluster 12:PAS>=PSP>PBS
putative steroid membrane binding protein [*Oryza sativa* (japonica cultivar-group)]	nd	Cluster 10: PAS>PSP>=PBS
RD22	Contig_060410_171	Cluster 12:PAS>=PSP>PBS
ribosomal protein S4	0	Cluster 11: PAS>=PBS>PSP
RSH3	0	Cluster 3: PAS>PSP>PBS
secretory peroxidase [*Avicennia marina*]	nd	Cluster 12:PAS>=PSP>PBS
serine/threonine kinase	0	Cluster 11: PAS>=PBS>PSP
serine/threonine protein phosphatase	0	Cluster 10: PAS>PSP>=PBS
sterol C-24 reductase	Contig_060410_137	Cluster 10: PAS>PSP>=PBS
stress related protein, putative [*Arabidopsis thaliana*]"	0	Cluster 12:PAS>=PSP>PBS
sucrose synthase	0	Cluster 10: PAS>PSP>=PBS, Cluster 11: PAS>=PBS>PSP
thaumatin-like protein	0	Cluster 10: PAS>PSP>=PBS
Thaumatin-like protein precursor	0	Cluster 10: PAS>PSP>=PBS
thionin	Contig_060410_99	Cluster 3: PAS>PSP>PBS
transcription factor lim1	Contig_060410_182	Cluster 11: PAS>=PBS>PSP

Annotation	Contig Designation	Cluster Description
translation elongation factor 1-gamma	0	Cluster 12:PAS>=PSP>PBS
trehalose synthase	0	Cluster 12:PAS>=PSP>PBS
tubulin	Contig_060410_183	Cluster 11: PAS>=PBS>PSP
tubulin alpha chain	Contig_060410_183	Cluster 11: PAS>=PBS>PSP
tubulin beta-1	Contig_060410_145	Cluster 11: PAS>=PBS>PSP
tumor differentially expressed protein 1	0	Cluster 11: PAS>=PBS>PSP
ubiquitin activating enzyme	0	Cluster 10: PAS>PSP>=PBS
ultraviolet-B-repressible protein	Contig_060410_141	Cluster 12:PAS>=PSP>PBS
vacuolar sortin receptor	0	Cluster 12:PAS>=PSP>PBS
xyloglucan endotransglycosylase	0	Cluster 11: PAS>=PBS>PSP, Cluster 12:PAS>=PSP>PBS
xyloglucan endotransglycosylase/hydrolase 16 protein	Contig_060410_85	Cluster 10: PAS>PSP>=PBS, Cluster 11: PAS>=PBS>PSP
xyloglucan fucosyltransferase	Contig_060410_53	Cluster 11: PAS>=PBS>PSP

Genes With Expression Increased In Fiber/Phloem At Snap Point

Annotation	Contig Designation	Cluster Description
(1-4)-beta-mannan endohydrolase	Contig_060410_108	Cluster 8: PSP>PBS>PAS
16kDa membrane protein	0	Cluster 7: PSP>=PAS>PBS
4-coumarate-CoA ligase	0	Cluster 8: PSP>PBS>PAS, Cluster 1: PSP>PBS>=PAS
ABC transporter	0	Cluster 8: PSP>PBS>PAS
acetoacyl-CoA-thiolase	0	Cluster 8: PSP>PBS>PAS
acyltransferase-like protein [*Gossypium hirsutum*]	nd	Cluster 8: PSP>PBS>PAS
adenosylhomocysteinase	0	Cluster 8: PSP>PBS>PAS
alcohol dehydrogenase	0	Cluster 1: PSP>PBS>=PAS
alginate biosynthesis protein	0	Cluster 1: PSP>PBS>=PAS
aluminum-induced protein	Contig_060410_131	Cluster 8: PSP>PBS>PAS
aquaporin	0	Cluster 1: PSP>PBS>=PAS
aquaporin	Contig_060410_143	Cluster 1: PSP>PBS>=PAS
aquaporin	Contig_060410_167	Cluster 1: PSP>PBS>=PAS
aquaporin	Contig_060410_212	Cluster 8: PSP>PBS>PAS
arabinogalactan protein [*Gossypium hirsutum*]	0	Cluster 8: PSP>PBS>PAS
ATP sulfurylase	Contig_060410_119	Cluster 8: PSP>PBS>PAS
ATP sulfyltransferaase	Contig_060410_119	Cluster 8: PSP>PBS>PAS
aux/IAA protein	Contig_060410_177	Cluster 4: PSP>PAS>PBS, Cluster 7: PSP>=PAS>PBS
auxin growth promoter protein	Contig_060410_18	Cluster 7: PSP>=PAS>PBS

Annotation	Contig Designation	Cluster Description
basic chitinase [*Nicotiana tabacum*]	Contig_060410_222	Cluster 4: PSP>PAS>PBS, Cluster 7: PSP>=PAS>PBS
beta-1,3 glucanase	0	Cluster 1: PSP>PBS>=PAS
beta-D-galactosidase	0	Cluster 7: PSP>=PAS>PBS
beta-ketoacyl-CoA-synthase	0	Cluster 1: PSP>PBS>=PAS
BURP domain-containing protein	0	Cluster 7: PSP>=PAS>PBS
catalase	Contig_060410_214	Cluster 4: PSP>PAS>PBS, Cluster 7: PSP>=PAS>PBS
cell-wall protein P8	Contig_060410_117	Cluster 4: PSP>PAS>PBS
chitinase	Contig_060410_189	Cluster 1: PSP>PBS>=PAS, Cluster 7: PSP>=PAS>PBS
chlorophyll a/b binding protein	Contig_060410_207	Cluster 1: PSP>PBS>=PAS
cinnamin acid 4-hydroxylase (C4H)	Contig_060410_45	Cluster 4: PSP>PAS>PBS
cobalamine-independent methionine synthase [*Solenostemon scutellarioides*]	0	Cluster 7: PSP>=PAS>PBS
cullin 3	0	Cluster 7: PSP>=PAS>PBS
cystein protease	Contig_060410_125	Cluster 1: PSP>PBS>=PAS
cytochrome P450	0	Cluster 8: PSP>PBS>PAS
cytosolic 6-phosphogluconate dehydrogenase	0	Cluster 1: PSP>PBS>=PAS
dehydrin 1	Contig_060410_42	Cluster 4: PSP>PAS>PBS
dhn1	Contig_060410_172	Cluster 4: PSP>PAS>PBS
dhn1 [*Populus euramericana*]	0	Cluster 1: PSP>PBS>=PAS, Cluster 8: PSP>PBS>PAS

Annotation	Contig Designation	Cluster Description
DNA packaging protein of prophage CP-933X	Contig_060410_48	Cluster 1: PSP>PBS>=PAS
DnaJ-like protein	0	Cluster 8: PSP>PBS>PAS
EIN3-binding F-box protein	0	Cluster 1: PSP>PBS>=PAS
elongation factor 1-alpha	Contig_060410_218	Cluster 1: PSP>PBS>=PAS
endo-1,4-beta glucanase	Contig_060410_19	Cluster 7: PSP>=PAS>PBS
endo-1,4-beta-glucanase	0	Cluster 1: PSP>PBS>=PAS
endosperm specific protein	Contig_060410_185	Cluster 7: PSP>=PAS>PBS
endo-xyloglucan transferase	Contig_060410_221	Cluster 4: PSP>PAS>PBS
ERD4 protein	0	Cluster 1: PSP>PBS>=PAS
expansin 3	0	Cluster 7: PSP>=PAS>PBS
F1 ATPase alpha subunit	0	Cluster 7: PSP>=PAS>PBS
fasciclin-like arabinogalactan	0	Cluster 7: PSP>=PAS>PBS
fasciclin-like arabinogalactan	Contig_060410_185	Cluster 7: PSP>=PAS>PBS
ferulate 5-hydroxylase	0	Cluster 8: PSP>PBS>PAS
Flagellar L-ring protein	Contig_060410_26	Cluster 8: PSP>PBS>PAS
GDSL-motif lipase/hydrolase	0	Cluster 1: PSP>PBS>=PAS
geranylgeranylated protein	0	Cluster 4: PSP>PAS>PBS
glucose-6-phosphate/phosphate-tranlocator	0	Cluster 8: PSP>PBS>PAS
glutathione transferase	0	Cluster 8: PSP>PBS>PAS
glyceraldehyde-3-phosphate dehydrogenase	Contig_060410_97	Cluster 1: PSP>PBS>=PAS

Annotation	Contig Designation	Cluster Description
glycosyl transferase	0	Cluster 7: PSP>=PAS>PBS
hemolysin activator protein	Contig_060410_25	Cluster 1: PSP>PBS>=PAS
hsc70	0	Cluster 1: PSP>PBS>=PAS
hypersensitive-induced reaction protein	0	Cluster 1: PSP>PBS>=PAS
leucine rich repeat protein	Contig_060410_160	Cluster 7: PSP>=PAS>PBS
lipid transfer protein	Contig_060410_91	Cluster 1: PSP>PBS>=PAS, Cluster 7: PSP>=PAS>PBS
Lissencephaly type-1-like homology motif; CTLH, C-terminal to LisH motif; Nitrous oxide reductase, N-terminal; WD40-like; Quinonprotein alcohol dehydrogenase-like	0	Cluster 1: PSP>PBS>=PAS
LTI6A	0	Cluster 8: PSP>PBS>PAS
lysosomal Pro-X carboxypeptidase	0	Cluster 8: PSP>PBS>PAS
mandelonitrile lyase	0	Cluster 1: PSP>PBS>=PAS
mannan endo-1,4-beta-mannosidase	0	Cluster 8: PSP>PBS>PAS, Cluster 7: PSP>=PAS>PBS
mannosyltransferase family protein	0	Cluster 1: PSP>PBS>=PAS
methionine synthase	Contig_060410_175	Cluster 7: PSP>=PAS>PBS
mitochondrial F1 ATP synthase beta subunit [*Arabidopsis thaliana*]	0	Cluster 1: PSP>PBS>=PAS, Cluster 7: PSP>=PAS>PBS
mitochondrial processing peptidase alpha	Contig_060410_89	Cluster 1: PSP>PBS>=PAS

Annotation	Contig Designation	Cluster Description
monodehydroascorbate reductase	0	Cluster 1: PSP>PBS>=PAS
MtN4 [*Medicago truncatula*]	0	Cluster 8: PSP>PBS>PAS
NAM-like TF	0	Cluster 8: PSP>PBS>PAS
nodulin	Contig_060410_124	Cluster 7: PSP>=PAS>PBS
non-specific lipid transfer protein	Contig_060410_193	Cluster 7: PSP>=PAS>PBS
PAS	0	Cluster 1: PSP>PBS>=PAS
pectin acetylesterase	Contig_060410_197	Cluster 8: PSP>PBS>PAS, Cluster 4: PSP>PAS>PBS, Cluster 7: PSP>=PAS>PBS, Cluster 1: PSP>PBS>=PAS
pectinacetylesterase precursor	Contig_060410_197	Cluster 4: PSP>PAS>PBS, Cluster 1: PSP>PBS>=PAS
pectinesterase	Contig_060410_139	Cluster 1: PSP>PBS>=PAS
pectinesterase	0	Cluster 8: PSP>PBS>PAS, Cluster 1: PSP>PBS>=PAS
pectin-glucuronyltransferase	0	Cluster 1: PSP>PBS>=PAS
peroxidase	Contig_060410_155	Cluster 1: PSP>PBS>=PAS
peroxidase	Contig_060410_225	Cluster 1: PSP>PBS>=PAS, Cluster 7: PSP>=PAS>PBS
peroxidase	Contig_060410_103	Cluster 8: PSP>PBS>PAS
peroxidase	Contig_060410_114	Cluster 8: PSP>PBS>PAS, Cluster 1: PSP>PBS>=PAS
pollen-specific protein	Contig_060410_129	Cluster 1: PSP>PBS>=PAS
polygalacturonase inhibitor-like	Contig_060410_112	Cluster 7: PSP>=PAS>PBS
polygalacturonase inhibitor-like	Contig_060410_160	Cluster 7: PSP>=PAS>PBS

Annotation	Contig Designation	Cluster Description
polyubiquitin	0	Cluster 8: PSP>PBS>PAS
polyubiquitin	Contig_060410_150	Cluster 8: PSP>PBS>PAS
polyubiquitin	Contig_060410_208	Cluster 8: PSP>PBS>PAS
proline rich APG	Contig_060410_110	Cluster 1: PSP>PBS>=PAS
proline rich protein	0	Cluster 8: PSP>PBS>PAS, Cluster 7: PSP>=PAS>PBS
proline-rich protein	Contig_060410_204	Cluster 1: PSP>PBS>=PAS
P-type H+-ATPase	0	Cluster 1: PSP>PBS>=PAS
pumilio domain-containing protein PPD1	0	Cluster 7: PSP>=PAS>PBS
putative peroxidase [*Arabidopsis thaliana*]	0	Cluster 8: PSP>PBS>PAS
putative ripening-related protein [*Vitis vinifera*]	0	Cluster 1: PSP>PBS>=PAS
putative steroid membrane binding protein [*Oryza sativa* (japonica cultivar-group)]	0	Cluster 1: PSP>PBS>=PAS
pyrophosphatase	0	Cluster 8: PSP>PBS>PAS, Cluster 1: PSP>PBS>=PAS
quinone oxidoreductase	0	Cluster 8: PSP>PBS>PAS
ribosomal protein L17	0	Cluster 4: PSP>PAS>PBS, Cluster 7: PSP>=PAS>PBS
ribulose-1,5-bisphosphate carboxylase small subunit rbcS1 [*Glycine max*]"	nd	Cluster 7: PSP>=PAS>PBS
ribulose-bisphosphate carboxylase	Contig_060410_205	Cluster 7: PSP>=PAS>PBS

Annotation	Contig Designation	Cluster Description
rubisco SSU	Contig_060410_205	Cluster 7: PSP>=PAS>PBS
S-adenosyl-L-methionine synthetase	Contig_060410_153	Cluster 1: PSP>PBS>=PAS
S-adenosylmethionine decarboxylase	Contig_060410_190	Cluster 8: PSP>PBS>PAS
stress-related	0	Cluster 8: PSP>PBS>PAS, Cluster 1: PSP>PBS>=PAS
sulfate adenylyltransferase	Contig_060410_119	Cluster 8: PSP>PBS>PAS
syntaxin of plants 71 [*Arabidopsis thaliana*]	0	Cluster 1: PSP>PBS>=PAS
thionin	Contig_060410_156	Cluster 7: PSP>=PAS>PBS
thionin	Contig_060410_99	Cluster 7: PSP>=PAS>PBS
Thioredoxin-like	0	Cluster 4: PSP>PAS>PBS
Transcriptional Regulator, IclR family	0	Cluster 7: PSP>=PAS>PBS
transcriptional regulator, LysR family	0	Cluster 7: PSP>=PAS>PBS
transporter	0	Cluster 7: PSP>=PAS>PBS
ubiquitin	Contig_060410_208	Cluster 8: PSP>PBS>PAS
ubiquitin homolog	Contig_060410_58	Cluster 8: PSP>PBS>PAS, Cluster 1: PSP>PBS>=PAS
UDP-glycosyltransferase	0	Cluster 8: PSP>PBS>PAS
xyloglucan endotransglucosylase	Contig_060410_221	Cluster 4: PSP>PAS>PBS
xyloglucan endotransglycosylase	Contig_060410_206	Cluster 4: PSP>PAS>PBS
xyloglucan endotransglycosylase	0	Cluster 8: PSP>PBS>PAS

Annotation	Contig Designation	Cluster Description
xyloglucan fucosyltransferase	Contig_060410_53	Cluster 8: PSP>PBS>PAS
zinc finger protein	0	Cluster 1: PSP>PBS>=PAS, Cluster 7: PSP>=PAS>PBS
zinc transporter protein ZIP2	0	Cluster 8: PSP>PBS>PAS

Genes With Expression Increased In Fiber/Phloem Below Snap Point

Annotation	Contig Designation	Cluster Description
(1-4)-beta-mannan endohydrolase	Contig_060410_108	Cluster 5: PBS>PAS>=PSP
20S proteasome subunit PBA1	0	Cluster 6: PBS>=PSP>PAS
21kd polypeptide	Contig_060410_224	Cluster 5: PBS>PAS>=PSP
26S proteasome regulatory particle non-ATPase subunit8 [*Oryza sativa* (japonica cultivar-group)]	0	Cluster 6: PBS>=PSP>PAS
26S proteasome subunit	0	Cluster 6: PBS>=PSP>PAS
40S ribosomal protein	Contig_060410_22	Cluster 9: PBS>PSP>=PAS
41 kD chloroplast nucleoid DNA binding protein	Contig_060410_132	Cluster 6: PBS>=PSP>PAS
4-alpha-methyl-sterol C4-methyl-oxidase	0	Cluster 6: PBS>=PSP>PAS
60S ribosomal protein	0	Cluster 9: PBS>PSP>=PAS, Cluster 6: PBS>=PSP>PAS, Cluster 5: PBS>PAS>=PSP
60S ribosomal protein L19	Contig_060410_6	Cluster 9: PBS>PSP>=PAS
ABC transporter	0	Cluster 6: PBS>=PSP>PAS, Cluster 5: PBS>PAS>=PSP
ABC transporter	Contig_060410_78	Cluster 6: PBS>=PSP>PAS, Cluster

Annotation	Contig Designation	Cluster Description
		5: PBS>PAS>=PSP
acetyl-CoA synthetase	0	Cluster 5: PBS>PAS>=PSP
actin	Contig_060410_147	Cluster 5: PBS>PAS>=PSP
actin-depolymerizing factor	0	Cluster 2: PBS>PSP>PAS, Cluster 9: PBS>PSP>=PAS
actin-depolymerizing factor 2	Contig_060410_202	Cluster 6: PBS>=PSP>PAS
adenosine nucleotide translocator	0	Cluster 9: PBS>PSP>=PAS
adenosylhomocysteinase [*Arabidopsis thaliana*]	nd	Cluster 5: PBS>PAS>=PSP
Adenosylhomocysteinase 2 (S-adenosyl-L-homocysteine hydrolase 1) (SAH hydrolase 2) (AdoHcyase 2)	0	Cluster 5: PBS>PAS>=PSP
ADP ribosylation factor 1 GTPase activating protein	0	Cluster 5: PBS>PAS>=PSP, Cluster 6: PBS>=PSP>PAS
ADP,ATP carrier-like protein	0	Cluster 9: PBS>PSP>=PAS
ADP-ribosylation factor	0	Cluster 6: PBS>=PSP>PAS
aldolase	Contig_060410_154	Cluster 5: PBS>PAS>=PSP
alpha-expansin 1 [*Populus tremula* x *Populus tremuloides*]	nd	Cluster 5: PBS>PAS>=PSP
aluminum-induced protein	Contig_060410_11	Cluster 5:

Annotation	Contig Designation	Cluster Description
		PBS>PAS>=PSP, Cluster 6: PBS>=PSP>PAS
aluminum-induced protein	Contig_060410_131	Cluster 6: PBS>=PSP>PAS
aquaporin	0	Cluster 6: PBS>=PSP>PAS
aquaporin	Contig_060410_143	Cluster 6: PBS>=PSP>PAS
aquaporin	Contig_060410_167	Cluster 6: PBS>=PSP>PAS
aquaporin	Contig_060410_212	Cluster 6: PBS>=PSP>PAS
aquaporin	Contig_060410_7	Cluster 6: PBS>=PSP>PAS
aquaporin	Contig_060410_80	Cluster 6: PBS>=PSP>PAS
aquaporin	Contig_060410_27	Cluster 9: PBS>PSP>=PAS, Cluster 6: PBS>=PSP>PAS
arabinogalactan protein	Contig_060410_178	Cluster 2: PBS>PSP>PAS
arabinogalactan protein	Contig_060410_159	Cluster 9: PBS>PSP>=PAS
arabinogalactan protein	Contig_060410_217	Cluster 2: PBS>PSP>PAS, Cluster 5: PBS>PAS>=PSP, Cluster 9: PBS>PSP>=PAS, Cluster 6: PBS>=PSP>PAS
arylsulfatase regulator	Contig_060410_149	Cluster 9: PBS>PSP>=PAS

Annotation	Contig Designation	Cluster Description
ascorbate peroxidase	0	Cluster 5: PBS>PAS>=PSP
ATP-dependent RNA helicase	0	Cluster 6: PBS>=PSP>PAS
beta subunit of ATP synthase	0	Cluster 5: PBS>PAS>=PSP
beta-expansin	0	Cluster 2: PBS>PSP>PAS, Cluster 5: PBS>PAS>=PSP
beta-glucosidase	Contig_060410_148	Cluster 5: PBS>PAS>=PSP, Cluster 9: PBS>PSP>=PAS
beta-tubulin 1	Contig_060410_115	Cluster 6: PBS>=PSP>PAS
bZIP protein BZ2	0	Cluster 6: PBS>=PSP>PAS
caffeic O-methyltransferase	Contig_060410_168	Cluster 5: PBS>PAS>=PSP, Cluster 9: PBS>PSP>=PAS
calmodulin	0	Cluster 6: PBS>=PSP>PAS
catalase	Contig_060410_214	Cluster 6: PBS>=PSP>PAS
catalase	Contig_060410_93	Cluster 6: PBS>=PSP>PAS
CCAAT-binding transcription factor subunit	0	Cluster 9: PBS>PSP>=PAS, Cluster 5: PBS>PAS>=PSP
CCR4-associated factor	0	Cluster 9: PBS>PSP>=PAS

Annotation	Contig Designation	Cluster Description
cellulase	Contig_060410_192	Cluster 2: PBS>PSP>PAS, Cluster 6: PBS>=PSP>PAS, Cluster 9: PBS>PSP>=PAS, Cluster 5: PBS>PAS>=PSP
chitinase	Contig_060410_198	Cluster 5: PBS>PAS>=PSP, Cluster 6: PBS>=PSP>PAS
chlorophyll a/b binding protein	Contig_060410_227	Cluster 5: PBS>PAS>=PSP
contains similarity to transcriptional activators such as RA-like and MYC-like regulatory R proteins [*Arabidopsis thaliana*]	0	Cluster 6: PBS>=PSP>PAS
cullin 3	0	Cluster 5: PBS>PAS>=PSP
cyclophilin	Contig_060410_144	Cluster 9: PBS>PSP>=PAS, Cluster 5: PBS>PAS>=PSP
cystein proteinase	0	Cluster 5: PBS>PAS>=PSP, Cluster 6: PBS>=PSP>PAS
cystein proteinase	Contig_060410_166	Cluster 5: PBS>PAS>=PSP, Cluster 6: PBS>=PSP>PAS
cysteine proteinase RD19A	0	Cluster 5: PBS>PAS>=PSP
cytochrome b5 isoform	0	Cluster 9: PBS>PSP>=PAS
cytochrome c oxidase subunit 6b	0	Cluster 5: PBS>PAS>=PSP, Cluster

Annotation	Contig Designation	Cluster Description
		6: PBS>=PSP>PAS
cytochrome P450	0	Cluster 6: PBS>=PSP>PAS
desacetoxyvindoline 4-hydroxylase	0	Cluster 2: PBS>PSP>PAS
dicarboxylate/tricarboxylate carrier	0	Cluster 9: PBS>PSP>=PAS
DNA-binding bromodomain-containing protein [*Arabidopsis thaliana*]	0	Cluster 5: PBS>PAS>=PSP
DnaJ protein	0	Cluster 5: PBS>PAS>=PSP, Cluster 6: PBS>=PSP>PAS
EIN3-like protein	0	Cluster 6: PBS>=PSP>PAS
elongation factor 1-alpha	Contig_060410_216	Cluster 5: PBS>PAS>=PSP
elongation factor 1-alpha	0	Cluster 9: PBS>PSP>=PAS
elongation factor 1-alpha	Contig_060410_218	Cluster 9: PBS>PSP>=PAS, Cluster 5: PBS>PAS>=PSP
elongation factor 2 [*Lithospermum erythrorhizon*]	0	Cluster 9: PBS>PSP>=PAS
endo-1,4-beta glucanase	0	Cluster 5: PBS>PAS>=PSP
endomembrane protein EMP70 precursor isolog	0	Cluster 5: PBS>PAS>=PSP
endosomal protein	0	Cluster 6: PBS>=PSP>PAS

Annotation	Contig Designation	Cluster Description
senescence-associated protein	Contig_060410_196	Cluster 9: PBS>PSP>=PAS, Cluster 5: PBS>PAS>=PSP
ER lumen protein retaining receptor	0	Cluster 9: PBS>PSP>=PAS
ethylene overproducer like	0	Cluster 2: PBS>PSP>PAS
eugenol O-methyltransferase	Contig_060410_83	Cluster 2: PBS>PSP>PAS, Cluster 9: PBS>PSP>=PAS
eukaryotic elongation factor 1A [*Salsola komarovii*]	nd	Cluster 5: PBS>PAS>=PSP
fasciclin-like AGP	0	Cluster 2: PBS>PSP>PAS, Cluster 6: PBS>=PSP>PAS
fasciclin-like AGP	Contig_060410_107	Cluster 6: PBS>=PSP>PAS, Cluster 2: PBS>PSP>PAS
fasciclin-like AGP	Contig_060410_201	Cluster 9: PBS>PSP>=PAS, Cluster 2: PBS>PSP>PAS
fasciclin-like AGP 11 [*Populus alba* x *Populus tremula*]	0	Cluster 2: PBS>PSP>PAS
fasciclin-like AGP 15 [*Populus alba* x *Populus tremula*]	nd	Cluster 2: PBS>PSP>PAS
ferritin	0	Cluster 9: PBS>PSP>=PAS
fiber protein	0	Cluster 9: PBS>PSP>=PAS
fiber protein Fb19 [*Gossypium barbadense*]	0	Cluster 6: PBS>=PSP>PAS

Annotation	Contig Designation	Cluster Description
fiber protein Fb27	0	Cluster 5: PBS>PAS>=PSP
fiber protein Fb34 [*Gossypium barbadense*]	Contig_060410_121	Cluster 5: PBS>PAS>=PSP
fis1 [*Linum*]	Contig_060410_111	Cluster 2: PBS>PSP>PAS
fis1 [*Linum*]	Contig_060410_62	Cluster 5: PBS>PAS>=PSP
Flagellar biosynthesis protein FlhA	0	Cluster 9: PBS>PSP>=PAS
flagellar P-ring protein	0	Cluster 5: PBS>PAS>=PSP
fructokinase	Contig_060410_126	Cluster 6: PBS>=PSP>PAS
fructose-bisphosphate aldolase	0	Cluster 2: PBS>PSP>PAS
fructose-bisphosphate aldolase	Contig_060410_188	Cluster 6: PBS>=PSP>PAS
FVE	0	Cluster 6: PBS>=PSP>PAS
gamma-tocopherol methyltransferase	0	Cluster 6: PBS>=PSP>PAS
GDP-D-mannose-4,6-dehydratase	0	Cluster 6: PBS>=PSP>PAS
glucose-6-phosphate/phosphate-tranlocator-like	0	Cluster 5: PBS>PAS>=PSP
glutamine synthetase	Contig_060410_92	Cluster 5: PBS>PAS>=PSP
glyceraldehyde 3-phosphate dehydrogenase	Contig_060410_165	Cluster 5: PBS>PAS>=PSP

Annotation	Contig Designation	Cluster Description
glycine-rich protein	Contig_060410_60	Cluster 9: PBS>PSP>=PAS
glycoside hydrolase family 19 protein [*Arabidopsis thaliana*]	0	Cluster 2: PBS>PSP>PAS, Cluster 9: PBS>PSP>=PAS
glycosyl hydrolase family 5 protein / cellulase family protein [*Arabidopsis thaliana*]	Contig_060410_192	Cluster 2: PBS>PSP>PAS
glycosyltransferase [*Triticum aestivum*]	nd	Cluster 5: PBS>PAS>=PSP
GTP binding protein	0	Cluster 6: PBS>=PSP>PAS
H-protein	Contig_060410_47	Cluster 5: PBS>PAS>=PSP
hsp70	0	Cluster 6: PBS>=PSP>PAS
hypersensitive-induced reaction protein	0	Cluster 6: PBS>=PSP>PAS
HyPRP	Contig_060410_64	Cluster 9: PBS>PSP>=PAS, Cluster 6: PBS>=PSP>PAS
IgE-dependent histamine-releasing factor homolog - rice	nd	Cluster 5: PBS>PAS>=PSP
immunophilin	0	Cluster 5: PBS>PAS>=PSP
importin alpha 1	Contig_060410_81	Cluster 9: PBS>PSP>=PAS, Cluster 5: PBS>PAS>=PSP
initiation factor eIFA	0	Cluster 5: PBS>PAS>=PSP

Annotation	Contig Designation	Cluster Description
inorganic pyrophosphatase	0	Cluster 6: PBS>=PSP>PAS
ion membrane transport protein	0	Cluster 6: PBS>=PSP>PAS
ketol-acid reductoisomerase	0	Cluster 5: PBS>PAS>=PSP
kinase: inflorescence and root apices receptor-like	0	Cluster 6: PBS>=PSP>PAS
kinase: protein kinase	0	Cluster 2: PBS>PSP>PAS
kinase: receptor like kinase	Contig_060410_75	Cluster 9: PBS>PSP>=PAS, Cluster 5: PBS>PAS>=PSP
kinase: receptor protein-like	Contig_060410_79	Cluster 2: PBS>PSP>PAS
kinase: RIO 2	0	Cluster 9: PBS>PSP>=PAS
laccase	0	Cluster 5: PBS>PAS>=PSP
LEA protein	0	Cluster 2: PBS>PSP>PAS, Cluster 6: PBS>=PSP>PAS
leucine rich repeat protein	Contig_060410_160	Cluster 9: PBS>PSP>=PAS
leucine-rich reporter protein	0	Cluster 6: PBS>=PSP>PAS
leucyl aminopeptidase	0	Cluster 9: PBS>PSP>=PAS
lipid transfer like protein [*Vigna unguiculata*]	nd	Cluster 2: PBS>PSP>PAS
lipid transfer protein	Contig_060410_41	Cluster 6:

Annotation	Contig Designation	Cluster Description
		PBS>=PSP>PAS
lipid transfer protein	0	Cluster 9: PBS>PSP>=PAS, Cluster 5: PBS>PAS>=PSP
lipid transfer protein	Contig_060410_101	Cluster 9: PBS>PSP>=PAS, Cluster 6: PBS>=PSP>PAS
lipoxygenase	0	Cluster 5: PBS>PAS>=PSP
meloidogyne-induced giant cell protein	0	Cluster 5: PBS>PAS>=PSP
metallothionein-like protein [*Pyrus pyrifolia*]	0	Cluster 2: PBS>PSP>PAS
mitochondrial F1 ATP synthase beta subunit [*Arabidopsis thaliana*]	0	Cluster 2: PBS>PSP>PAS, Cluster 5: PBS>PAS>=PSP, Cluster 6: PBS>=PSP>PAS, Cluster 9: PBS>PSP>=PAS
NADH:cytochrome b5 reductase	Contig_060410_73	Cluster 9: PBS>PSP>=PAS
NADP specific isocitrate dehydrogenase	0	Cluster 6: PBS>=PSP>PAS
nicotinate-nucleotide adenylyltransferase	0	Cluster 9: PBS>PSP>=PAS
nodulin	Contig_060410_124	Cluster 5: PBS>PAS>=PSP, Cluster 9: PBS>PSP>=PAS
nodulin-like	0	Cluster 6: PBS>=PSP>PAS
nucleoid DNA-binding protein	Contig_060410_132	Cluster 2: PBS>PSP>PAS,

Annotation	Contig Designation	Cluster Description
cnd41		Cluster 9: PBS>PSP>=PAS
nucleotide-sugar dehydratase	Contig_060410_146	Cluster 5: PBS>PAS>=PSP
outer membrane ferripyoverdine receptor	0	Cluster 9: PBS>PSP>=PAS
oxoglutarate malate translocator [*Solanum tuberosum*]	0	Cluster 5: PBS>PAS>=PSP
p23 co-chaperone	0	Cluster 9: PBS>PSP>=PAS
PAS	0	Cluster 9: PBS>PSP>=PAS
permease	0	Cluster 6: PBS>=PSP>PAS
peroxidase	0	Cluster 5: PBS>PAS>=PSP
peroxidase	Contig_060410_59	Cluster 5: PBS>PAS>=PSP
peroxidase	Contig_060410_225	Cluster 9: PBS>PSP>=PAS
pfkB-type carbohydrate kinase	Contig_060410_126	Cluster 9: PBS>PSP>=PAS
P-glycoprotein	Contig_060410_78	Cluster 5: PBS>PAS>=PSP
Phosphoglucomutase, alpha-D-glucose phosphate-specific	0	Cluster 5: PBS>PAS>=PSP, Cluster 6: PBS>=PSP>PAS
phosphoribosylanthranilate transferase	0	Cluster 6: PBS>=PSP>PAS

Annotation	Contig Designation	Cluster Description
photoassimilate-responsive protein PAR-1b-like protein	Contig_060410_128	Cluster 2: PBS>PSP>PAS, Cluster 5: PBS>PAS>=PSP
photosystem 1 subunit 5	Contig_060410_173	Cluster 5: PBS>PAS>=PSP
photosystem I chain IV	Contig_060410_15	Cluster 6: PBS>=PSP>PAS
photosystem I subunit II	0	Cluster 6: PBS>=PSP>PAS
photosystem II type I chlorophyll a/b/ binding protein	Contig_060410_227	Cluster 9: PBS>PSP>=PAS
phytochrome associated protein 2	Contig_060410_123	Cluster 6: PBS>=PSP>PAS
phytochrome-associated protein 1	0	Cluster 9: PBS>PSP>=PAS
phytosulfokine precursor 3_2	0	Cluster 9: PBS>PSP>=PAS, Cluster 5: PBS>PAS>=PSP
polyubiquitin	0	Cluster 6: PBS>=PSP>PAS
polyubiquitin	Contig_060410_208	Cluster 6: PBS>=PSP>PAS
potassium-efflux system	0	Cluster 2: PBS>PSP>PAS
PPase	Contig_060410_50	Cluster 5: PBS>PAS>=PSP
P-protein	0	Cluster 5: PBS>PAS>=PSP
profilin	Contig_060410_191	Cluster 6: PBS>=PSP>PAS

Annotation	Contig Designation	Cluster Description
profilin-like protein [*Humulus scandens*]	nd	Cluster 5: PBS>PAS>=PSP
proline rich ARG-like	0	Cluster 5: PBS>PAS>=PSP
protease: subtilisin	0	Cluster 5: PBS>PAS>=PSP
protease: subtilisin	Contig_060410_62	Cluster 5: PBS>PAS>=PSP
protein kinase [*Trifolium repens*]	nd	Cluster 6: PBS>=PSP>PAS
proton-inorganic pyrophosphatase	Contig_060410_140	Cluster 5: PBS>PAS>=PSP
Pto kinase inhibitor	0	Cluster 6: PBS>=PSP>PAS
putative auxin-regulated protein [*Oryza sativa* (japonica cultivar-group)]	0	Cluster 6: PBS>=PSP>PAS
putative immunophilin [*Hordeum vulgare* subsp. vulgare]	0	Cluster 5: PBS>PAS>=PSP
pyrroline-5-carboxylate synthetase	0	Cluster 5: PBS>PAS>=PSP
Ribonucleoside-diphosphate reductase	0	Cluster 2: PBS>PSP>PAS
ribosomal protein L28-like [*Oryza sativa* (japonica cultivar-group)]	0	Cluster 9: PBS>PSP>=PAS
ribosomal protein L3	0	Cluster 9: PBS>PSP>=PAS
ribosomal protein L9	0	Cluster 9: PBS>PSP>=PAS

Annotation	Contig Designation	Cluster Description
ribosomal protein S14	0	Cluster 5: PBS>PAS>=PSP
ribosomal protein S4	0	Cluster 9: PBS>PSP>=PAS
RNA helicase	0	Cluster 5: PBS>PAS>=PSP
RNA recognition motif (RRM)-containing protein	0	Cluster 6: PBS>=PSP>PAS
rubisco SSU	0	Cluster 9: PBS>PSP>=PAS
S-adenosyl-L-homocysteine hydrolase	Contig_060410_151	Cluster 9: PBS>PSP>=PAS, Cluster 5: PBS>PAS>=PSP
S-adenosyl-L-methionine synthetase	Contig_060410_116	Cluster 6: PBS>=PSP>PAS
S-adenosylmethionine decarboxylase	Contig_060410_120	Cluster 6: PBS>=PSP>PAS
S-adenosylmethionine decarboxylase	Contig_060410_190	Cluster 6: PBS>=PSP>PAS
S-adenosylmethionine synthase	Contig_060410_116	Cluster 6: PBS>=PSP>PAS
SEC61 alpha subunit	0	Cluster 6: PBS>=PSP>PAS
small RAS-like GTP-binding protein	Contig_060410_8	Cluster 5: PBS>PAS>=PSP
Spermidine synthase	0	Cluster 5: PBS>PAS>=PSP
STYLOSA	0	Cluster 2: PBS>PSP>PAS
succinate dehydrogenase	0	Cluster 6:

Annotation	Contig Designation	Cluster Description
		PBS>=PSP>PAS
TCTP protein [Fragaria x ananassa]	nd	Cluster 5: PBS>PAS>=PSP, Cluster 6: PBS>=PSP>PAS
TED4	Contig_060410_41	Cluster 6: PBS>=PSP>PAS
TetR	0	Cluster 6: PBS>=PSP>PAS
thaizole biosynthetic enzyme	Contig_060410_122	Cluster 6: PBS>=PSP>PAS
Thi1 protein [*Arabidopsis thaliana*]	nd	Cluster 9: PBS>PSP>=PAS, Cluster 5: PBS>PAS>=PSP
translation initiation factor 4A	Contig_060410_186	Cluster 9: PBS>PSP>=PAS, Cluster 5: PBS>PAS>=PSP
translation elongation factor 1A	Contig_060410_216	Cluster 5: PBS>PAS>=PSP, Cluster 9: PBS>PSP>=PAS
translation elongation factor 1A	Contig_060410_218	Cluster 5: PBS>PAS>=PSP, Cluster 9: PBS>PSP>=PAS
translation initiation factor	Contig_060410_186	Cluster 5: PBS>PAS>=PSP
translation initiation factor 5	0	Cluster 5: PBS>PAS>=PSP
translational initiation factor eIF1	0	Cluster 6: PBS>=PSP>PAS
translationally controlled tumor protein	Contig_060410_224	Cluster 9: PBS>PSP>=PAS, Cluster 6: PBS>=PSP>PAS,

Annotation	Contig Designation	Cluster Description
		Cluster 5: PBS>PAS>=PSP
tubulin beta-1	Contig_060410_145	Cluster 9: PBS>PSP>=PAS
UBC E2	Contig_060410_46	Cluster 6: PBS>=PSP>PAS
UBC7	Contig_060410_72	Cluster 5: PBS>PAS>=PSP
ubiquitin	Contig_060410_208	Cluster 6: PBS>=PSP>PAS
ubiquitin activating enzyme	0	Cluster 5: PBS>PAS>=PSP, Cluster 6: PBS>=PSP>PAS
ubiquitin carrier	0	Cluster 6: PBS>=PSP>PAS
ubiquitin carrier protein	0	Cluster 5: PBS>PAS>=PSP
ubiquitinating enzyme	0	Cluster 5: PBS>PAS>=PSP, Cluster 9: PBS>PSP>=PAS
vacuolar processing enzyme	0	Cluster 6: PBS>=PSP>PAS
xyloglucan endotransglycosylase	0	Cluster 9: PBS>PSP>=PAS
zinc finger protein	Contig_060410_184	Cluster 6: PBS>=PSP>PAS
zinc-finger protein 1	0	Cluster 6: PBS>=PSP>PAS

Appendix 5 – NCBI Conserved Domain Database and Entrez Gene References for Function of Genes Found in Microarray Analysis

Nearest NCBI GI_ACCESSION (Protein database)*	Gene Name	Conserved Domain Reference	Entrez Gene Reference	Functional Class
15076515	Mitochondrial Processing Peptidase E Alpha Subunit	COG0612: PqqL		Amino acid & Protein Metabolism
28804499	Polyubiquitin - GI28804499	cd01803: Ubiquitin		Amino acid & Protein Metabolism
7862066	Polyubiquitin - GI7862066	cd01803: Ubiquitin, cl:00155		Amino acid & Protein Metabolism
21592581	Ubiquitin	cd01803: Ubiquitin		Amino acid & Protein Metabolism
471162	Vacuolar Processing Enzyme Precursor	cl02159: Peptidase_C13		Amino acid & Protein Metabolism
10177010	Alcohol Dehydrogenase	COG1062: AdhC	834230	Glycolysis & Respiration
77540210	Glyceraldehyde-3-Phosphate Dehydrogenase A Subunit	PRK07403: PRK07403	837848	Glycolysis & Respiration
7329666	(1-4)-Beta-Mannan Endohydrolase-Like Protein	pfam00150: Cellulase	831699	Glycosylhydrolases
51536480	AT1G13130 like	pfam00150		Glycosylhydrolases
34016875	Chitinase-like protein	cd00325: chitinase glyco_hydro_19		Glycosylhydrolases
2244732	Endo-Xyloglucan Transferase	cd02176: GH16_XET		Glycosylhydrolases
1431629	Pectinacetylesterase Precursor	pfam03283:PAE	827683	Glycosylhydrolases
732913	Pectinesterase-3 Precursor	cl04375: PMEI		Glycosylhydrolases
42795466	Xyloglucan Endotransglucosylase-Hydrolase XTH7	cd02176: GH16_XET	836702	Glycosylhydrolases
1008904	Xyloglucan Endotransglycosylase	cd02176: GH16_XET		Glycosylhydrolases
950299	Xyloglucan Endotransglycosylase Precursor	cd02176: GH16_XET		Glycosylhydrolases

Nearest NCBI GI_ACCESSION (Protein database)*	Gene Name	Conserved Domain Reference	Entrez Gene Reference	Functional Class
37993671	UDP-Glycosyltransferase 89B2	cl10013: glycosyltransferase_GTB_type		Glycosyltransferases
84313479	Xyloglucan Endotransglucosylase	cd02176: GH16_XET		Glycosyltransferases
56606534	Aluminum-Induced Protein	cd01910: Wali7		Miscellaneous
11994234	*Arabidopsis* Thaliana Genomic DNA, Chromosome 3, P1 Clone:MUJ8			Miscellaneous
14334816	Argonaute (AGO1), putative protein	cd04657: Piwi_ago_like	817246	Miscellaneous
6606509	ATP Sulfurylase	cd00517: ATPS		Miscellaneous
1575327	ATP Sulfurylase GI 1575327	cd00517: ATPS		Miscellaneous
4033365	ATP Sulfurylase GI 4033365	cd0517: ATPS		Miscellaneous
20269059	AUX/IAA Protein	pfam02309: AUX_IAA		Miscellaneous
53748473	Dehydrin 1	pfam00257		Miscellaneous
21593191	Desiccation-Related Protein, Putative			Miscellaneous
29120043	DHN1 Protein (Fragment)			Miscellaneous
3201617	DUF231 Hypothetical Protein At2g30900	pfam03005: DUF231	817640	Miscellaneous
20804811	DUF250 - P0028e10.8 Protein	cl1037: DUF6		Miscellaneous
1621268	DUF642 - Hypothetical protein - GI 1621268	pfam04862: DUF642		Miscellaneous
3043426	DUF642 - Hypothetical protein - GI 3043426	pfam04862: DUF642		Miscellaneous
21537379	DUF642 - Hypothetical Protein GI 21537379	pfam04862: DUF642		Miscellaneous
20259856	DUF760 - Hypothetical Protein At3g17800; Meb5.2	pfam05542: DUF760		Miscellaneous
77381733	Flagellar L-Ring Protein Precursor - Lipoprotein	cl00905: FlgH	3717103	Miscellaneous
71556266	Hydrolase, Alpha/Beta Fold Family Protein	cl09107: Esterase_lipase	3558065	Miscellaneous
4039153	Hydrophobic Protein RCI2A - UPF0057	cl00431: UPF	819757	Miscellaneous
24982415	Hypothetical Protein - DOXX	cl00976: DoxX	1044916	Miscellaneous

Nearest NCBI GI_ACCESSION (Protein database)*	Gene Name	Conserved Domain Reference	Entrez Gene Reference	Functional Class
8979711	Nucleoid DNA-binding - CND41-like protein	cd05472: cnd41_like		Miscellaneous
21389647	Photoassimilate-Responsive Protein PAR-1B-Like Protein	pfam06521: PAR1	824571	Miscellaneous
62319609	RSH3	cl01004: Rel-Spo_like		Miscellaneous
21537353	Stress-related - Rubber elongation factor - protein AT3G05500	pfam05755: REF		Miscellaneous
479090	Sulfate Adenylyltransferase	cd00517: ATPS		Miscellaneous
62732987	TUB family, putative protein	pfam01167: Tub		Miscellaneous
48310614	Wound-induced WI12 - AT5G01740/T20L15_10	pfam07107: WI12	831686	Miscellaneous
38194916	Phytochelatin Synthetase-Like Protein	pfam04833: COBRA		Other Cell Wall Enzymes
30841338	Arabinogalactan Protein	cl02663:Fasciclin		Other Cell Wall Proteins
22531042	At4g14500	cl00141: START	827097	Other Cell Wall Proteins
47717925	Fasciclin-like AGP 11	cl02663:Fasciclin	831914	Other Cell Wall Proteins
47717933	Fasciclin-Like AGP 15	cl02663:Fasciclin		Other Cell Wall Proteins
47717907	Fasciclin-like AGP 2	cl02663:Fasciclin		Other Cell Wall Proteins
47717913	Fasciclin-Like AGP 5	cl02663:Fasciclin		Other Cell Wall Proteins
47717915	Fasciclin-like AGP 6	cl02663:Fasciclin		Other Cell Wall Proteins
10178224	Hypothetical protein AT5G64430; T12B11.2	cd06410: PB1_UP2	836564	Other Cell Wall Proteins
28194086	Lipid Transfer Protein Isoform 4	cd01960: nsLTP1	100233106	Other Cell Wall Proteins
7529744	Pectinesterase - AT3G59010/F17J16_60	cl04375: PMEI	825070	Other Cell Wall Proteins
6851366	Putative Cell-Wall P8 Protein		816620	Other Cell Wall Proteins
41053116	Putative Pectin Acetylesterase	pfam03283:PAE	4330443	Other Cell Wall Proteins
3860333	Basic Blue Copper Protein	pfam02298: Cu_bind_like		Oxidoreductases
19070130	Catalase	cd00328: catalase	100232861	Oxidoreductases
33521521	CYP81E8 - Cytochrome P450	COG2124:CypX		Oxidoreductases

Nearest NCBI GI_ACCESSION (Protein database)*	Gene Name	Conserved Domain Reference	Entrez Gene Reference	Functional Class
14009640	Cytosolic 6-Phosphogluconate Dehydrogenase	pfam00393:6PGD		Oxidoreductases
452165	Mono-dehydroascorbate Reductase, Seedling Isozyme	pfam:07992: Pyr_redox_2, PRK09287		Oxidoreductases
17066703	Peroxidase	cd00693: secretory peroxidase		Oxidoreductases
39777536	Peroxidase Precursor	cd00693: secretory peroxidase		Oxidoreductases
927428	Probable Aldehyde Dehydrogenase	cl00545: Aldedh		Oxidoreductases
15808674	Putative Quinone Oxidoreductase	COG0604: Qor		Oxidoreductases
11414898	Uricase	cd00445: Uricase		Oxidoreductases
9965897	Cinnamate-4-Hydroxylase	COG2124:CypX		Phenylpropanoid Biosynthesis
nucleotide: 255558074	Cinnamoyl-CoA Reductase	cl09931: NADB_Rossmann	8280501	Phenylpropanoid Biosynthesis
1044868	P450C4H - Cinnamin Acid 4-Hydroxylase	COG2124:CypX		Phenylpropanoid Biosynthesis
20197554	Hypothetical protein AT2G47860	pfam03000:NPH3	819398	Photosynthesis - Chloroplast
403160	Ribulose 1,5-Bisphosphate Carboxylase Small Subunit Precursor	cd03527: RuBisCO_small		Photosynthesis - Chloroplast
10946375	Ribulose-1,5-Bisphosphate Carboxylase Small Subunit rbcS1 - AF303939_1	cd03527: RuBisCO_small		Photosynthesis - Chloroplast
83584402	EIN3-Binding F-Box Protein	c102535: F-box	817087	Regulation
22530934	GID1-Like Protein 1	cl09107: Esterase_lipase	819674	Regulation
38566726	Leucine Rich Repeat Protein Precursor	pfam08263: LRRNT_2, cl02423: LRR_RI		Regulation
62087121	Transcription Factor LIM1	cl02475: LIM		Regulation
2853219	Glutathione Transferase	cd03185: Gst, GOG0625: Gst		Secondary Metabolism

Nearest NCBI GI_ACCESSION (Protein database)*	Gene Name	Conserved Domain Reference	Entrez Gene Reference	Functional Class
1403044	S-Adenosyl-L-methionine decarboxylase	cl03253: SAM_decarbox		Secondary Metabolism
13540318	S-Adenosyl-L-methionine synthetase	pfam02772: S-AdoMet_synt_M, PTZ00104		Secondary Metabolism
47232488	S-Adenosylmethionine Decarboxylase	cl03253: SAM_decarbox		Secondary Metabolism
1794147	Carrot Root Specific - Aquaporin	cd00333: MIP		Transport
31322147	Ferrous Ion Membrane Transport Protein DMT1	cl00836:Nramp	547711	Transport
1657948	MIPC - Aquaporin	cd00333: MIP		Transport
17939849	Mitochondrial F1 ATP synthase beta subunit	cd01133: F1-ATPase_beta		Transport
49388575	OSTIP2.1 - Aquaporin	cd00333: MIP	4330207	Transport
2384671	Putative Potassium Transporter ATKT2P	cl01227: K_trans	818649	Transport
68347997	Sulfate ABC Transporter, Periplasmic Sulfate-Binding Protein	cl11398: LysR	3480785	Transport
52076090	HYPOTHETICAL PROTEIN B1331F11.3-1	cd04894: ACT_AXR-like_1	4347917	Unaligned
49388541	HYPOTHETICAL PROTEIN P0620H05.6	cd01427: HAD-Like	4328788	Unaligned

Appendix 6– qRT Primer Sets

All primers are written 3’→ 5’.

Primer Set	Primer Name		Primer Sequence
pcs3h1	pcsh1F	Forward Primer	TCCAACTCGGAACTGGCTAAA
	pcsh1R	Reverse Primer	ACTGAACCTCGCCGACGAT
pal1	pal1F	Forward Primer	CCTCCCACAGGAGAACCAAA
	pal1R	Reverse Primer	ACTCTGTTCCGTGCCCAAAT
pal2	pal2F	Forward Primer	GCTCTTCACGGTGGCAACTT
	pal2R	Reverse Primer	CATCAGTTTACCAATTGCTGCAA
pal3	pal3F	Forward Primer	CCACACACTGCCGCACTCT
	pal3R	Reverse Primer	GCCTTGCAACAGGGTATTGATC
pal4	pal4F	Forward Primer	TCATGGATAGCATGGACAAAGG
	pal4R	Reverse Primer	GTCCCCACCCCAAATATTCC
pal5	pal5F	Forward Primer	ATCCCTTGTTGGAATGTCTTAGCA
	pal5R	Reverse Primer	AAAGGGAAAAGGACAATCGAAGT
c4h1	c4h1F	Forward Primer	CAGGTGACCGAGCCAGACAT
	c4h1R	Reverse Primer	AGGAGTGGGATTGCCATACG
c4h2	c4h2F	Forward Primer	AGTCATCGCAAAGCTCAAAGG
	c4h2R	Reverse Primer	TCTCCGACTTGGAGCCAGTT
ccr1	ccr1F	Forward Primer	TGAAGGAGCAAACGAGAGATTG
	ccr1R	Reverse Primer	ACGCCTTGGCAGCCTTTAAT

Primer Set	Primer Name		Primer Sequence
ccr2	ccr2F	Forward Primer	ACGAGGTGGACTGGTCTGTGT
	ccr2R	Reverse Primer	CGAACGGAGTAGCCACGAAT
f5h1	f5h1F	Forward Primer	TGGTACGGCCCAGACAGAGT
	f5h1R	Reverse Primer	CCCATAGTCACCAGGAAATTCAC
f5h2	f5h2F	Forward Primer	GGAGGTCTTGCCCCGGTAT
	f5h2R	Reverse Primer	CAGGTAATTCCCACGTGAAACA
comt1	caom1F	Forward Primer	CACAAGCGTCGCACCTCTT
	caom1R	Reverse Primer	GGGTTTCACCTGACACTTTGTCT
comt2	caom2F	Forward Primer	AGAAATCGCTGCCGCAATC
	caom2R	Reverse Primer	AAAATTTTGCGACGGACGAGTA
comt3	caom3F	Forward Primer	CACTGGTGGAAAAGAAAGAACTGA
	caom3R	Reverse Primer	GCAGGGATTTTGGTGATTTTG
comt4	caom4F	Forward Primer	CGGATATTCACCCAACACCAA
	caom4R	Reverse Primer	TCAGCAGCCATCTGGAAGAGT
comt5	caom5F	Forward Primer	TCGCTCCGTTTGAGAAAGAAC
	caom5R	Reverse Primer	ACGCCATCGCGTTGTTAAG
comt6	caom6F	Forward Primer	CTCCATTGCTCCTCTTTGTCTCA
	caom6R	Reverse Primer	ATTCCCCCATCAAGTACTGCAT
ccoaomt1	ccom1F	Forward Primer	TCCAAAGCGATGCTCTCTATCA
	ccom1R	Reverse Primer	CAGTCACTTCCCTGAGCTCCTT
4cl1	4ccl1F	Forward Primer	ACGAGCTATTTCATTATCCGGTAAA
	4ccl1R	Reverse Primer	CCGGAAGCGAGTAAGAGAGAAG

Primer Set	Primer Name		Primer Sequence
4cl2	4ccl2F	Forward Primer	TCAACAAGCTTGGAGTCAAACAA
	4ccl2R	Reverse Primer	GATATGAAGCTCCTAAGAAGGTGAAGA
4cl3	4ccl3F	Forward Primer	CGTCGTTGCGTTGCCTTACT
	4ccl3R	Reverse Primer	TGAGCCACGCTAGTCACAAGA
4cl4	4ccl4F	Forward Primer	TGAGCGTGGCACCTATTGTG
	4ccl4R	Reverse Primer	GCACTTTCAACGACGACAAGTC
4cl5	4ccl5F	Forward Primer	CACAGCGGTACCTACGATTCAC
	4ccl5R	Reverse Primer	CTCCGAATGAACCGCAACTT
cesa1	cs1F	Forward Primer	TCGATGGGCTTTAGGTTCCA
	cs1R	Reverse Primer	TCGAGCCACTTGAGCTTTCC
cesa2	cs2F	Forward Primer	CTGCCGCGTCTTGTGTATATCT
	cs2R	Reverse Primer	AACTCTCACCAGAGCGTTCATG
cesa3	cs3F	Forward Primer	AAAGCCTGCTTGCGATGTTC
	cs3R	Reverse Primer	TGTTATCGTGGATCGATTTCCA
cesa4	cs4F	Forward Primer	GGACTTGTCGCCGGATCTC
	cs4R	Reverse Primer	TGACCATCCAAGTTCCTCAATG
cesa5	cs5F	Forward Primer	GAATGGTGATTGTGGCTCGTT
	cs5R	Reverse Primer	ACCAGAGCCCAAATGAATCCT
cesa6	cs6F	Forward Primer	GGCCTGGTTTCCAACATCAC
	cs6R	Reverse Primer	TCAACATGAAAGGAGCATTGG
bgal	csgalF	Forward Primer	GCCACGATCCTAATTCCATTTT
	csgalR	Reverse Primer	TTTGCTAACACCAGGACACGTACT

Appendix 7 – Primer Sets and *C. sativa* EST Accessions Examined for qRT-PCR

Gene	Gene Name	Accession Number	qRT-PCR Primer Set	Primer set used for qRT-PCR study	Alignment
PCS3H	p-coumaroyl shikimate 3'-hydroxylase	Genbank gi|110190252|gb|EC855388.1|EC855388 EST00126	PCS3H1	No – poor reaction	PCS3H1
PAL	Phenylalanine ammonia-lyase	Deyholos lab lcl|MD|02_HEMPSE_RP_011_D08_30MAR2006_058	PAL1	No – low expression in snap point	PAL 1
PAL	Phenylalanine ammonia-lyase	Genbank gi|110190233|gb|EC855369.1|EC855369 EST00107	PAL1	No – low expression in snap point	PAL 1
PAL	Phenylalanine ammonia-lyase	Genbank gi|110190232|gb|EC855368.1|EC855368 EST00106	PAL1	No – low expression in snap point	PAL 1
PAL	Phenylalanine ammonia-lyase	Genbank gi|110190256|gb|EC855392.1|EC855392 EST00130	PAL2	No – low expression in snap point	PAL 1
PAL	Phenylalanine ammonia-lyase	Genbank gi|110190237|gb|EC855373.1|EC855373 EST00111	PAL3	No – low expression in snap point	PAL 1
PAL	Phenylalanine ammonia-lyase	Genbank gi|110190238|gb|EC855374.1|EC855374 EST00112	PAL4	No – low expression in snap point	PAL 1
PAL	Phenylalanine ammonia-lyase	Genbank gi|110190236|gb|EC855372.1|EC855372 EST00110	PAL5	No – low expression in snap point	PAL 1
C4H	Cinnamate-4-hydroxylase	Genbank gi|110190212|gb|EC855348.1|EC855348 EST00086	C4H1	Yes	C4H 1

Gene	Gene Name	Accession Number	qRT-PCR Primer Set	Primer set used for qRT-PCR study	Alignment
C4H	Cinnamate-4-hydroxylase	Genbank gi|110190211|gb|EC855347.1|EC855347 EST00085	C4H2	No – Same expression pattern as c4h1	C4H 1
C4H	Cinnamate-4-hydroxylase	Genbank gi|110190210|gb|EC855346.1|EC855346 EST00084	C4H2	No – Same expression pattern as c4h1	C4H 1
C4H	Cinnamate-4-hydroxylase	Deyholos lab lcl|MD|02_HEMPSE_RP_015_B07_30MAR2006_061	C4H2	No – Same expression pattern as c4h1	C4H 1
C4H	Cinnamate-4-hydroxylase	Deyholos lab lcl|MD|02_HEMPSE_RP_019_D11_01APR2006_089	C4H2	No – Same expression pattern as c4h1	C4H 1
CCR	Cinnamoyl CoA reductase	Genbank gi|110190257|gb|EC855393.1|EC855393 EST00131	CCR1	Yes	CCR 1
CCR	Cinnamoyl CoA reductase	Genbank gi|110190225|gb|EC855361.1|EC855361 EST00099	CCR2	Yes	CCR 2
CCR	Cinnamoyl CoA reductase	Deyholos lab lcl|MD|02_HEMPSE_RP_017_B01_30MAR2006_013	CCR1	Yes	CCR 1
CCR	Cinnamoyl CoA reductase	Deyholos lab lcl|MD|02_HEMPSE_RP_011_A12_30MAR2006_096	None made	No – No Primers	
F5H	Ferulate 5-hydroxylase	Deyholos lab lcl|MD|HEMPSEQ_RP_005_G04_16DEC2005_020	None made	No – No Primers	F5H 1

Gene	Gene Name	Accession Number	qRT-PCR Primer Set	Primer set used for qRT-PCR study	Alignment
F5H	Ferulate 5-hydroxylase	Deyholos lab lcl|MD|HEMPSEQ_RP_006_B08_16DEC2005_062	F5H2	Yes	F5H 1
COMT	Caffeic acid O-methyltransferase	Deyholos lab lcl|MD|02_HEMPSE_RP_013_H02_01APR2006_002	COMT 1	No – low expression in snap point	COMT 1
COMT	Caffeic acid O-methyltransferase	Deyholos lab lcl|MD|02_HEMPSE_RP_015_B11_30MAR2006_093	COMT 1	No – low expression in snap point	COMT 1
COMT	Caffeic acid O-methyltransferase	Deyholos lab lcl|MD|HEMPSEQ_RP_009_E06_06MAR2006_040	COMT 1	No – low expression in snap point	COMT 1
COMT	Caffeic acid O-methyltransferase	Deyholos lab lcl|MD|02_HEMPSE_RP_016_G07_30MAR2006_051	COMT 2	Yes	COMT 1
COMT	Caffeic acid O-methyltransferase	Deyholos lab lcl|MD|02_HEMPSE_RP_017_G04_30MAR2006_020	COMT 3	No – low expression in snap point	COMT 2
COMT	Caffeic acid O-methyltransferase	Deyholos lab lcl|MD|02_HEMPSE_RP_019_H02_01APR2006_002	COMT 3	No – low expression in snap point	COMT 2
COMT	Caffeic acid O-methyltransferase	Deyholos lab lcl|MD|HEMPSEQ_RP_002_A06_19AUG2005_048	COMT 3	No – low expression in snap point	COMT 2
COMT	Caffeic acid O-methyltransferase	Deyholos lab lcl|MD|HEMPSEQ1_RP_001_H01_10JUN2005_001	COMT 3	No – low expression in snap point	COMT 2

Gene	Gene Name	Accession Number	qRT-PCR Primer Set	Primer set used for qRT-PCR study	Alignment
COMT	Caffeic acid O-methyltransferase	Deyholos lab lcl|MD|HEMPSEQ_RP_009_D03_06MAR2006_025	COMT 4	No – same pattern as caom2	COMT 3
COMT	Caffeic acid O-methyltransferase	Deyholos lab lcl|MD|02_HEMPSE_RP_015_A10_30MAR2006_080	COMT 4	No – low expression in snap point	COMT 3
COMT	Caffeic acid O-methyltransferase	Deyholos lab lcl|MD|COO34_UP_001_C08_20MAY2005_060	COMT 3	No – low expression in snap point	COMT 2
COMT	Caffeic acid O-methyltransferase	Genbank gi|110190184|gb|EC855320.1|EC855320 EST00058	COMT 6	No – low expression in snap point	COMT 4
COMT	Caffeic acid O-methyltransferase	Deyholos lab lcl|MD|02_HEMPSE_RP_015_A09_30MAR2006_079	COMT 6	No – low expression in snap point	COMT 4
COMT	Caffeic acid O-methyltransferase	Genbank gi|110190186|gb|EC855322.1|EC855322 EST00060	COMT 6	No – low expression in snap point	COMT 4
COMT	Caffeic acid O-methyltransferase	Deyholos lab lcl|MD|HEMPSEQ_RP_006_A11_16DEC2005_095	COMT 6	No – low expression in snap point	COMT 4
COMT	Caffeic acid O-methyltransferase	Genbank gi|110190185|gb|EC855321.1|EC855321 EST00059	COMT 6	No – low expression in snap point	COMT 4
COMT	Caffeic acid O-methyltransferase	Deyholos lab lcl|MD|HEMPSEQ_RP_005_H08_16DEC2005_050	COMT 6	No – low expression in snap point	COMT 4

Gene	Gene Name	Accession Number	qRT-PCR Primer Set	Primer set used for qRT-PCR study	Alignment
COMT	Caffeic acid O-methyltransferase	Deyholos lab lcl|MD|HEMPSEQ_RP_006_C02_16DEC2005_012	COMT 6	No – low expression in snap point	COMT 4
CCoAOMT	Caffeoyl-CoA O-methyltransferase	Genbank gi|110190209|gb|EC855345.1|EC855345 EST00083	CCoAOMT1	No – poor primer reactions	CCoAOMT 1
CCoAOMT	Caffeoyl-CoA O-methyltransferase	Genbank gi|110190208|gb|EC855344.1|EC855344 EST00082	CCoAOMT1	No – poor primer reactions	CCoAOMT 1
4CL	4-coumarate-CoA ligase	Deyholos lab lcl|MD|02_HEMPSE_RP_019_G05_01APR2006_035	4CL1	Yes	4CL 1
4CL	4-coumarate-CoA ligase	Genbank gi|110190204|gb|EC855340.1|EC855340 EST00078	4CL4	No – low expression in snap point	4CL 2
4CL	4-coumarate-CoA ligase	Genbank gi|110190205|gb|EC855341.1|EC855341 EST00079	4CL3	No – low expression in snap point	4CL 2
4CL	4-coumarate-CoA ligase	Genbank gi|110190206|gb|EC855342.1|EC855342 EST00080	4CL2	No – low expression in snap point	4CL 2
4CL	4-coumarate-CoA ligase	Genbank gi|110190207|gb|EC855343.1|EC855343 EST00081	4CL2	No – low expression in snap point	4CL 2
4CL	4-coumarate-CoA ligase	Genbank gi|110190230|gb|EC855366.1|EC855366 EST00104	4CL5	Yes	4CL 1
4CL	4-coumarate-CoA ligase	Genbank gi|110190231|gb|EC855367.1|EC855367 EST00105	4CL1	Yes	4CL 1

Gene	Gene Name	Accession Number	qRT-PCR Primer Set	Primer set used for qRT-PCR study	Alignment
CESA	Cellulose synthase	Genbank gi|110190255|gb|EC855391.1|EC855391 EST00129	CESA1	Yes	CESA 1
CESA	Cellulose synthase	Genbank gi|110190253|gb|EC855389.1|EC855389 EST00127	CESA2	Yes	CESA 2
CESA	Cellulose synthase	Genbank gi|110190246|gb|EC855382.1|EC855382 EST00120	CESA3	No – Similar expression to cs2	CESA 2
CESA	Cellulose synthase	Genbank gi|110190245|gb|EC855381.1|EC855381 EST00119	CESA4	Yes	CESA 3
CESA	Cellulose synthase	Genbank gi|110190235|gb|EC855371.1|EC855371 EST00109	CESA5	No – low expression in snap point	CESA 1
CESA	Cellulose synthase	Genbank gi|110190234|gb|EC855370.1|EC855370 EST00108	CESA6	No– low expression in snap point	CESA 1
CESA	Cellulose synthase	Deyholos lab 02_HEMPSE_RP_011_H03_30MAR2006_017	CESA1	Yes	CESA 1
CESA	Cellulose synthase	Deyholos lab HEMPSEQ_RP_006_C12_16DEC2005_092	CESA5	No – low expression in snap point	CESA 1
CESA	Cellulose synthase	Deyholos lab HEMPSEQ_RP_006_D08_16DEC2005_058	None made	No – No Primers	CESA 3
BGAL	beta-D-galactosidase	Deyholos lab HEMPSEQ_RP_007_F02_06FEB2006_006	BGAL1	Yes	XGAL 1
BGAL	beta-D-galactosidase	Deyholos lab 02_HEMPSE_RP_021_E05_01APR2006_039	None made	No– No Primers	

Appendix 8 – Determination of Bast Fiber Development in Hemp Hypocotyls

Tissue from *Cannabis sativa* genotype Carmen was collected to determine a time frame for secondary wall formation in bast fiber of hemp hypocotyls. Collections were done of emergent hemp (3 to 30 days old) in two independent samplings (Sept and October). Daily measurements of the height and width of 30 hypocotyls were made using a ruler and calipers. Sections of hypocotyls from 5 to 10 plants were taken daily and stained with fast green. Then measurements of height and the width of the cell wall and the width of the "ring" of developing cell wall cells were made to determine the time after germination that fiber development starts. Tissue from the cotyledons and various portions of hypocotyl from 10 plants was collected and frozen for later RNA extractions (Figure 0-1).

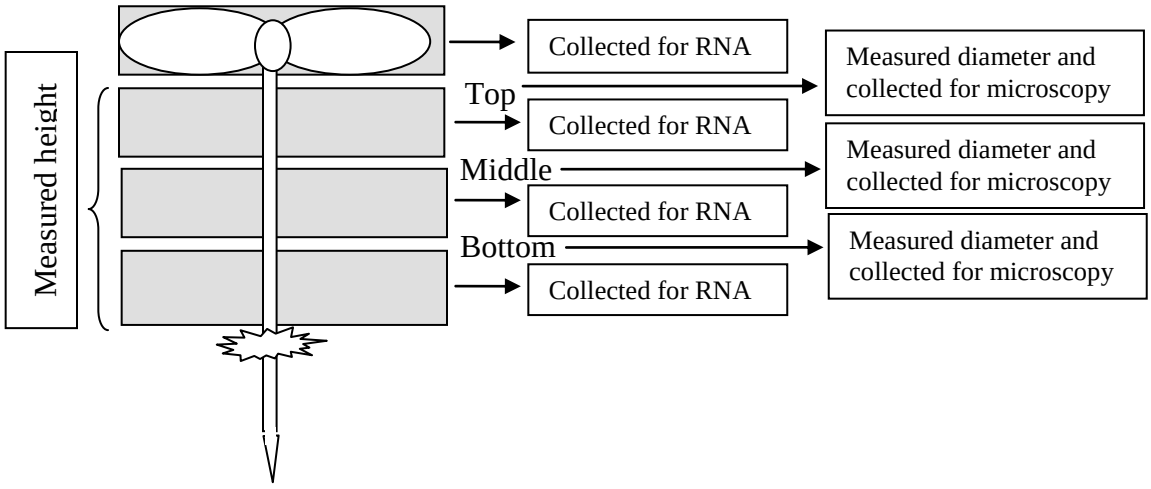


Figure 0-1: Diagram of dissected tissue for cotyledon and hypocotyl study.

Microscopy sections showed that 14 to 16 days after planting, bast fibers started to develop, and lignification could be seen in the phloem. The diameter of the ring of cell wall increases in two distinct segments as the fibers cells below grow past each other. The photos of the top segment show that 6 days after planting the xylem starts to develop. 14 days after planting distinctive xylem structures are seen but there is no visible lignification in the phloem. By day 16 there is visible lignification and by day 19 a second row of intrusive fibers is seen (Figure 0-2). The middle segment (pictures not shown) mimics the same timeline as the top segment. The bottom segment develops fiber one day later.

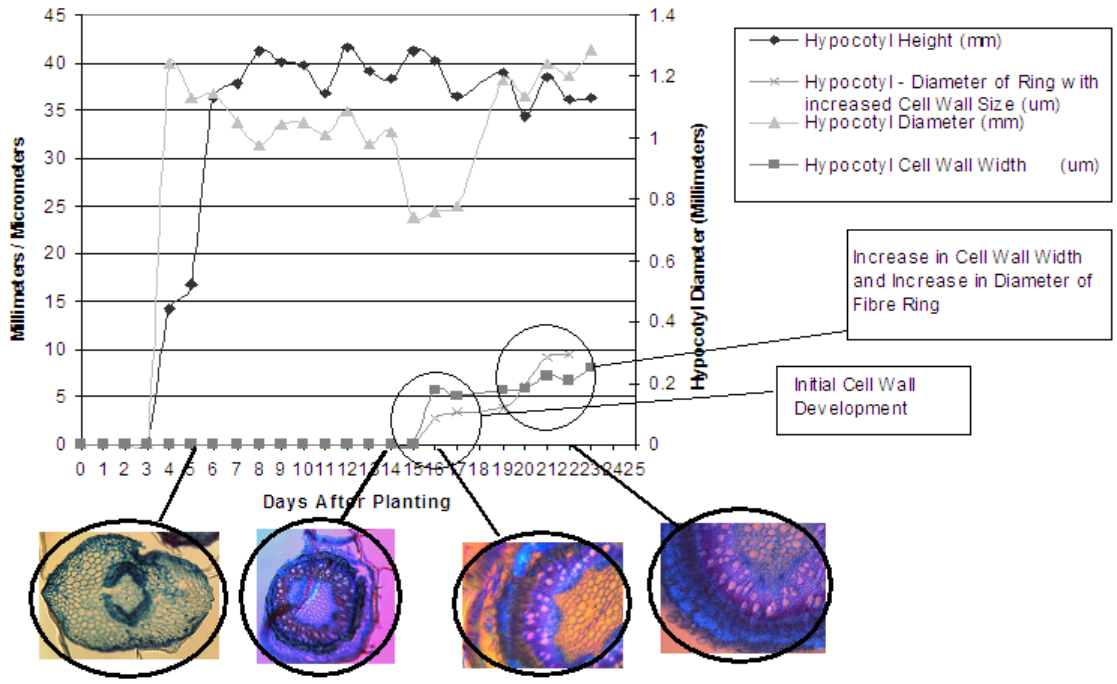


Figure 0-2: Graph of height and width of hypocotyl(mm) and hypocotyl cell wall (µm) in *Cannabis sativa* genotype Carmen.
Pictures of the top segment of hypocotyl are (from left to right) are from five, fourteen, sixteen, and twenty two days after planting.

The height and the diameter of the stem did not correlate directly with the onset of fiber development in the hypocotyl. Only the date after planting was used to determine the samples used in later RNA extractions.