

Abstract

Cameron R. Morris. Applications of Metabolic Profiling in the Study of Wood Formation. (Under the direction of Dr. Hou-min Chang and John F. Kadla)

In this work, a method of metabolic profiling of developing xylem tissue was developed and utilized in various applications. The first study examines the metabolic differences between four families of trees, involved in a tree breeding program, that exhibit different levels of α -cellulose content. Profiling results detect an increase in the concentration levels of glucose in the high α -cellulose samples. In the low α -cellulose samples, there was an accumulation of fructose and sucrose. The accumulation of sucrose in the low α -cellulose samples suggests an inhibition of P-Susy, which converts sucrose to UDP-glucose.

The second α -cellulose study utilizes clones that were planted on two different sites. One clone was classified as a high α -cellulose clone and the other as a low α -cellulose clone. By having the clones planted on two different sites, metabolic variations between the clones were observed on the same site and between sites, allowing a comparison of genetic and environmental differences associated with cellulose content. Based on the statistical analysis of the profiling data, the genetic differences associated with the high α -cellulose clone is an increase in the concentration levels of alanine, phosphoric acid, and free sugars, fructose, glucose, and mannose. The accumulation of fructose suggests that the greater amounts of cellulose produced releases more fructose in the cell that is not be recycled or slow to be recycled back into the cellulose biosynthesis.

Tree growth was also studied to examine the metabolic variation and transcript levels of genes associated with growth. Two families grown on a genetic site were chosen from a group of families based on their differences in average tree height. Statistical analysis of the profiling data identifies glutamine as a metabolite that has a positive correlation with tree height. Microarray results identify a NAM (no apical meristem)-like protein as being overly expressed in the fast growing family. NAM is a protein that regulates shoot apical meristem formation, which is responsible for vertical growth.

The last study, examines the metabolic effects of BWA infestation on the Fraser Fir species. In response to the feeding process of the BWA various changes occur in the tree metabolism, creating an increase in the concentration levels of shikimic acid, polyols, and some amino acids. The accumulations of polyols and amino acids suggest infestation causes the trees to enter a state of water stress.

**Applications of Metabolic Profiling in the Study of
Wood Formation**

by

Cameron R. Morris

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Wood and Paper Science

Raleigh
2005

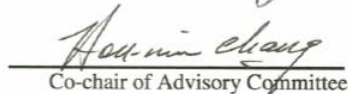
Approved by





Co-chair of Advisory Committee





Co-chair of Advisory Committee

I would like to dedicate this dissertation to my parents, Wilbert and Sharon Morris, and sister, Melanie, for their patience, love and guidance. You all have had a great influence on my life and made me the man I am today!

Biography

Cameron R. Morris was born in Raleigh, N.C. on August 15, 1978. He received his primary and secondary education in Raleigh. In the fall 1996, he entered North Carolina State University where he graduated Summa Cum Laude with a B.S. in Chemical Engineering and Wood and Paper Science, while a four-year letterman in track and field and a member of Phi Beta Sigma, Inc.

In the fall of 2000, the author began the Ph. D. curriculum in the Department of Wood and Paper Science under the direction of Dr. Hou-min Chang and Dr. John F. Kadla.

Acknowledgements

The author would like to express his indebtedness to Dr. Hou-min Chang and Dr. John F. Kadla for their thoughtful and constructive comments throughout the process of developing this dissertation. Your valuable assistance will always be remembered.

Sincere thanks are also extended to Dr. Barry Goldfarb and Dr. Ronald Sederoff for their priceless assistance and guidance through the years. The author would also like to thank Dr. Fikret Isik, Dr. Jason Osborne, Dr. John Frampton, Deborah Craig, and Chris Smith for their assistance and valuable suggestions. Also, he would like to thanks Dr. Satoshi Kubo, Dr. Tomoya Yokoyama, Dr. Tatsuhiko Yamada, and Ting-feng Yeh for their in-depth discussions and friendship.

The author would like to thanks his friends far and wide for all of their support and many words of encouragement and laughs along the way. Special appreciation is also extended to immediate and extended family for their sacrifices, patience, love, and other contributions that kept me going throughout this experience. Lastly, the author would like to thank GOD for all the blessings and love bestowed throughout life.

Financial Support from US DOE (DE-PS36-00GO 10480) and USDA-IFAFS (2001-52104-11224) is greatly appreciated.

Table of Contents

	Page
List of Figures	vii
List of Tables	x
1. Introduction	1
1.1 Background	1
1.2 Cell Wall Components	3
1.3 Photosynthesis	4
1.4 Cellulose	6
1.4.1 Cellulose Biosynthesis	7
1.5 Hemicellulose	15
1.6 Lignin	19
1.6.1 Lignin Precursors	19
1.6.2 Monolignol Precursor Biosynthesis	22
1.6.3 Lignin Polymerization	24
1.7 Extractives	28
1.8 Functional Genomics	31
1.9 Metabolomics	31
1.10 Metabolic Profiling	32
1.10.1 Instruments for Metabolic Profiling	33
1.10.2 Sample Preparation	41
1.10.3 Quantitation of Metabolites	42
1.10.4 Statistical Analysis	43
1.11 Microarrays	43
1.12 References	48
2. Metabolic Profiling – a New Tool in the Study of Wood Formation	54
2.1 Abstract	55
2.2 Introduction	56
2.3 Materials and Methods	58
2.4 Results and Discussion	64
2.5 References	77
3. Variation of α-Cellulose Content and Related Metabolites During Wood Formation in Loblolly Pine	79
3.1 Abstract	80
3.2 Introduction	81
3.3 Materials and Methods	83
3.4 Results and Discussion	87

3.5 Conclusion	99
3.6 References	100
4. Metabolic Profiling and Microarray Analysis of a Loblolly Pine Growth Study	101
4.1 Abstract	102
4.2 Introduction	103
4.3 Materials and Methods	104
4.4 Results and Discussion	110
4.5 Conclusion	126
4.6 References	126
5. The Utilization of Metabolic Profiling in the Study of BWA Infestation of Fraser Fir	128
5.1 Abstract	129
5.2 Introduction	130
5.3 Materials and Methods	133
5.4 Results and Discussion	137
5.5 Conclusion	150
5.6 References	150
6. Metabolite Analysis of Clones in a Cellulose Study	153
6.1 Abstract	154
6.2 Introduction	155
6.3 Materials and Methods	156
6.4 Results and Discussion	160
6.5 Conclusion	169
6.6 References	169
7. Future Work	170
8. Appendix	173
8.1 Metabolic Profiling Protocol	174
8.2 Microarray Protocol	177
8.3 Structure of Significant Metabolites	181
8.4 Lipid Phase of Metabolite Profiles	182

List of Figures

	Page
1. Introduction	1
Reaction 1. Overall Chemical Reaction of Photosynthesis	5
Figure 1. Structure of Glucan Chain	7
Reaction 2. Conversion of Sucrose to UDP-glucose	9
Figure 2. Cellulose Organization in Cell Wall	10
Figure 3. Structure of Wood Cell	11
Figure 4. Model Pathway of Cellulose Production from Sucrose in Plants	12
Figure 5. Cellulose Rosette Assembly	15
Figure 6. Hemicelluloses in Wood	17
Figure 7. Hemicellulose Biosynthetic Pathway	18
Figure 8. Lignin Precursors	20
Figure 9. Major Lignin Linkages	21
Figure 10. Shikimic Acid Pathway	26
Figure 11. Monolignol Biosynthesis Pathway	27
Figure 12. Simplistic Extractives Biosynthetic Scheme	30
Figure 13. Metabolite Profile of Polar Phase of Leaf Extract from Arabidopsis	38
Figure 14. Experimental Scheme of Microarrays	45
2. Metabolic Profiling – a New Tool in the Study of Wood Formation	54
Scheme 1. Metabolite Isolation Protocol	69
Figure 1. Chromatograms of Polar Phase Extraction Times	68
Figure 2. Chromatograms of Lipid Phase Extraction Times	68
Figure 3. Lipid Phase Metabolite Profile	73
Figure 4. Polar Phase Metabolite Profile	74
Figure 5. Metabolic phenotype clustering. Representation of clusters created after PCA of polar metabolite data taken from thirteen samples, which were spilt into two groups representing separate families. In order to provide, the best clusters between families, principal components 1, 2 and 4 were chosen. These principal components contain 74% of the ‘total information content derived from metabolite variances’	75
Figure 6. Principal Component 1 vs. Principal Component 2. This provides a graphical representation of the impact that individual polar metabolites have on the clustering results. The metabolites that are further away from zero have a greater impact on the linear combination that is used to calculate the principal component vector. Likewise, the metabolites closer to zero have less of an impact on the linear combination. The labeled metabolites, which are further from zero are responsible from the phenotypic differences seen between the two families of trees: 4 = shikimic acid; 5 = citric acid; 7 = unknown; 12 and 13 = D-Fructose; 18 = Ononitol;	76

3. Variation of α-Cellulose Content and Related Metabolites During Wood Formation in Loblolly Pine	79
Figure 1. Cellulose Production of the 14 Full-sib Families	88
Figure 2. Metabolic phenotypic clustering results of extreme high and low α -cellulose for the 13 th year. Principal components 1, 2, and 3 contain 57% of the total information substance resulting from the metabolite variances.	91
Figure 3. Stepwise Regression Model for α -cellulose Content Prediction	94
Figure 4. Graph of the differences between normalized mean relative concentrations for the 13 th year. Metabolites that have positive integers are high in cellulose content. Metabolites that have negative integers are part of the low cellulose classification.	96
Figure 5. Cellulose Biosynthetic Pathway	97
4. Metabolic Profiling and Microarray Analysis of a Loblolly Pine Growth Study	101
Figure 1. Graph of Tree Height versus α -Cellulose Content	113
Figure 2. Graph of Tree Height versus Lignin Content	113
Figure 3. Phenotypic Clustering Analysis (PCA) Fast-Slow Growing Trees. Principal components 1, 2, and 3 contain 65.36% of the total information resulting from the metabolite variances. The filled squares represent trees from the slow growing family. The circles represent trees from the fast growing family.	114
Figure 4. Hierarchical Cluster Analysis (HCA) of Fast-Slow Growing Trees. Results of this clustering analysis indicate that there is an overlap between trees in the fast and slow growing families.	115
Figure 5. Ratio of Metabolite Levels in Fast and Slow Growing Families. Metabolites on the right side of the graph are high in concentrations for the fast growing family. Metabolites on the left side of the graph represent metabolites higher in concentrations for slow growing family.	116
Figure 6. Volcano plot comparing fast growing family versus slow growing family. Each point of the graph represents a gene. The y-axis is $-\log_{10}$ (P-value) of the comparison. The x-axis is the estimated fold change on a \log_2 scale. Horizontal line in the graph is obtained from Bonferroni correction across the multiple hypothesis tests. Genes above the line are statistically significant under strict control of false positive rate.	120
5. The Utilization of Metabolic Profiling in the Study of BWA Infestation of Fraser Fir	128
Figure 1. Metabolic Phenotypic clustering results for xylem tissue circle = infested Fraser Fir square = non-infested Fraser Fir x = Veitch Fir z = infested Fraser Fir from Veitch Fir	140

Figure 2. Metabolic Phenotypic clustering results for phloem tissue circle = infested Fraser Fir square = non-infested Fraser Fir x = Veitch Fir z = infested Fraser Fir from Veitch Fir	141
Figure 3. Plot of PC 1 versus PC 2 separating infested and non-infested Fraser Fir xylem tissue samples. Metabolites further away from zero play a greater role in the clustering results.	142
Figure 4. Graph of xylem tissue metabolite results: Infested/Non-infested Fraser Fir. This graph represents the ratio of the average relative concentration of individual metabolites. The right side of the graph represents metabolites that are higher in concentrations for the non-infested samples. Metabolites on the right side of the graph denote pooling of metabolites in the infested Fraser Fir sample group which may accumulate in response to BWA infestations. Metabolites with colored bars were determined to be significantly different using the student t-test.	146
Figure 5. Graph of the ratio of the average relative concentration of individual metabolites: unknowns 45, 30 and tartaric acid. Unknown 45 accumulates in non-infested Fraser Fir and unknown 30 and tartaric acid accumulate in infested Fraser Fir.	147
Figure 6. Graph of phloem tissue metabolite results: Non-infested Fraser Fir/Veitch Fir. The right side of the graph corresponds to the accumulation of metabolites for the Veitch Fir species. The left side of the graph represents the amassing of metabolites for non-infested Fraser Fir samples. Metabolites with colored bars were determined to be significantly different using the student t-test.	149
6. Metabolite Analysis of Clones in a Cellulose Study	153
Figure 1. Site by Clone Interaction for D-Fructose. The solid bars represent clone 75 and the diagonal shaded bars represent clone 24. Results of the graph indicates that D-fructose contains a slope change interaction; thus, suggesting that the changes observed in metabolic concentration levels are due mostly to genetics.	167
Figure 2. Cellulose Metabolic Biosynthetic Pathway	168

List of Tables

	Page
1. Introduction	1
2. Metabolic Profiling – a New Tool in the Study of Wood Formation	54
Table 1. Extraction time variation. The initial polar phase extractions were conducted holding the lipid phase extraction constant. Once an optimum polar phase extraction time of thirty minutes was achieved, the lipid phase extraction was varied holding the polar phase extraction constant.	67
Table 2. Relative concentrations of random metabolites for various Polar Phase Extraction Times	67
3. Variation of α-Cellulose Content and Related Metabolites During Wood Formation in Loblolly Pine	79
Table 1. ANOVA Table of Stepwise Regression Model	92
Table 2. Stepwise Regression Model Information	93
4. Metabolic Profiling and Microarray Analysis of a Loblolly Pine Growth Study	101
Table 1. ANOVA Table of Stepwise Regression Model	117
Table 2. Stepwise Regression Model Information	118
Table 3. Significant Genes Up-regulated in the Fast Growing Family (Model 1)	121
Table 4. Significant Genes Up-regulated in the Slow Growing Family (Model 1)	122
Table 5. Differentially Expressed Genes for Fast/Slow Growing Families (Model 2). Fold change greater than 1 represents genes up-regulated in the fast growing trees. Fold change less than 1 represent genes down-regulated in slow growing trees. Levels of significance are based on John Storey's False Discovery Rate q-value.	124
5. The Utilization of Metabolic Profiling in the Study of BWA Infestation of Fraser Fir	128
6. Metabolite Analysis of Clones in a Cellulose Study	153
Table 1. Four Clone Analysis of Variance	162
Table 2. Least Means Squares Analysis for Clonal Effect	162
Table 3. Significant Metabolites Results from the 2 clone Analysis of Variance	164

1. INTRODUCTION

1.1 Background

The overall objective of this study is the utilization of metabolic profiling to study the genetic and/or environmental differences of differentiating xylem tissue. In this work, two of the studies reflect on or report the metabolic differences in trees that contain difference in α -cellulose content. The first study makes metabolic observations on seedlings that were grown in South Carolina. The second study involves clonal samples to distinguish the various genetic and environmental effects associated with the observed phenotypes. In the next study, two families of trees were found to vary significantly in tree height. Metabolic profiling was utilized to observe the metabolic differences between the families. Microarrays were also conducted to determine if any genes were differentially expressed between the fast and slow growing families. The last study looked at the impact of infestation of the Balsam Woolly Adelgid (BWA) on the metabolite levels of the Fraser Fir species.

The significance of this project was the potential to better understand how wood is formed. Metabolic profiling provides a means to evaluate changes in metabolic pathways of the function/identity of changes in metabolic systems arising from genetic engineering or natural mutation. Properties altered by variations in cellulose production could be a step towards engineering trees with properties needed by the pulp and paper industry. If the principles of cellulose production were better understood, there could be a significant improvement in a \$300 billion global industry by increasing the efficiency of the pulp and paper making process.

The techniques needed for these series of profiling experiments have been developed and are discussed in detail in later sections of the document. Metabolic profiling is a technique that will enable the analysis of wood structure through both genetic and biochemical approaches. The extraction and derivatization protocols required for profiling have only been applied to a woody plant, Arabidopsis, but a protocol for loblolly pine (*Pinus taeda* L.) has been developed for profiling of xylem tissue. Whetten et al. have done extensive research using microarrays on pine cDNA clones (Whetten, Zhang et al. 2001). There is some difficulty associated in working with microarrays due to the large amounts of data generated from an array as well as biological and technical variation. To date, clustering analysis has been used as a tool to group genes that act in similar ways in different conditions. Another possible solution to handling microarray data was to determine the genes that are up-regulated or down-regulated based on phenotype utilizing various models to produce p values or John Storey q values.

Metabolic profiling is a powerful technique that will be utilized due to its ability to analyze the effect of changes, amongst those gene expression and its effects on metabolites and phenotypes during development or as a result of environmental or chemical stresses (Glassbrook, Beecher et al. 2000). Fiehn et al. have developed an extraction protocol for Arabidopsis tissue samples (Fiehn, Kopka et al. 2000). The protocol involves isolation followed by derivatization, and the metabolites subsequently analyzed by gas chromatography/mass spectrometry (GC/MS). In addition to GC/MS analysis, metabolites are also analyzed by of high-performance liquid chromatography (HPLC). In this case, samples are set aside prior to derivatization due to the fact that

derivatization is not required for the use of HPLC. With the use of this protocol as a reference, another protocol has been developed for the use of loblolly pine tissue samples to conduct metabolic profiling on such samples.

When coupled to genomics, metabolic profiling offers a way to elucidate the function of novel genes. Profiling is a valuable tool in decoding the biochemical functions of plant genes and the regulatory networks present in plant metabolism. A unique aspect of this project was the selection of extreme phenotypes that have potential to provide improved wood quality to both paper and sawmills. To date, there has been little genetic analysis applied to extreme phenotypes, such as those that produce higher cellulose content or have higher growth rates. Both traits could provide higher wood volume and wood with superior properties, thereby increasing the value of a wood as a plantations crop.

Expression microarrays are another technique that will be used in this project. The technique uses expressed sequence tag (EST) cDNA clones to investigate variation in transcript level for a large number of genes. The advantage of using microarrays is the ability to compare transcript variation from two different tissue samples representing differences in development or environmental treatment. For example, gene transcript variation can be compared between CAD mutated tissue and wild-type tissue to determine its potential role for cell wall biosynthesis.

1.2 Cell Wall Components

The plant cell wall consists of three layers: middle lamella, primary wall and secondary wall. The middle lamella comprised of pectic compounds, protein, and lignin, is the first layer formed during cell division, which is made up of the outer wall of the

cell and is shared by adjacent cells. The primary wall formed after the middle lamella consists of cellulose, lignin, hemicellulose and pectins (Cosgrove 1997). The secondary wall comprised of three layers: S1 (outer layer), S2 (middle layer), and S3 (inner layer), is deposited at the conclusion of cell growth and is made up of cellulose, hemicellulose, and lignin (Reiter 2002).

Cell wall formation occurs in three steps. The first step is the formation of a cell plate (Terashima, Fukushima et al. 1993). The cell plate is comprised mostly of a hard pectin gel, formed by the cross-linking of the partially methylated α -1,4-polygalacturonic acid, that is secreted by the Golgi apparatus (Terashima, Fukushima et al. 1993; Gibeaut and Carpita 1994). The second step in cell wall formation involves the formation of a primary wall, comprised of cellulose microfibrils, hemicellulose, and pectins (Terashima, Fukushima et al. 1993; Gibeaut and Carpita 1994; Bonetta, Facette et al. 2002; Reiter 2002).

Prior to formation of the S₁ layer, lignification begins in the cell corner followed by the middle lamella (Terashima, Fukushima et al. 1993). At this point, cellulose microfibrils produced in the plasma membrane are deposited onto the secondary wall. Usually a step behind, hemicellulose is deposited onto the secondary wall at the conclusion of cellulose deposition. Always lagging behind polysaccharide deposition, lignification advances through the primary wall into the secondary wall. Upon the conclusion of the lignification process, the cell dies.

1.3 Photosynthesis

Photosynthesis is the conversion of light energy to chemical energy in the chloroplast. This conversion takes place in two different stages: light and dark reactions.

The light reaction consists of two photosystems that generate NADPH and ATP. Photosystems are an arrangement of pigment molecules such as chlorophyll and various proteins. The dark reaction converts CO₂ into glucose by utilizing the energy (NADPH and ATP) produced from the light reaction.

In the light reaction, light is absorbed at 700nm, where it excites photosystem I and releases an electron that ultimately produces NADPH via two electrons that combine NADP⁺ and H⁺. NADPH is a strong reducing agent that donates its electrons to the process of glucose synthesis. In order to replace the electron loss from photosystem I, photosystem II absorbs light at 680 nm and releases one electron. On its downward path, ADP is converted to ATP prior to entering photosystem I. In photosystem II, it takes two H₂O molecules to convert to O₂; therefore 4 ATP are created within a cycle. Lastly, if too much NADPH is stored, the Shunt pathway takes place to prevent a toxic build up of NADPH. Therefore, the light reaction would shut down with the initiation of the Shunt pathway.

The dark reaction, also known as the Calvin Cycle, occurs in the chloroplast stroma. It utilizes NADPH and ATP created in the light reaction to generate glucose by reducing CO₂. The overall reaction for photosynthesis is provided in reaction 1. The ultimate result of photosynthesis is the production of a sugar pool that is utilized by other metabolic pathways such as cellulose, hemicellulose, lignin, and extractives biosynthesis.



Reaction 1.1. Overall Chemical Reaction of Photosynthesis

1.4 Cellulose

In 1838, Anselm Payen first recognized the existence of cellulose. Cellulose has been determined to be the most abundant plant polymer (Gibeaut and Carpita 1994; Cosgrove 1997). It serves as the major structural component of the walls that surround the cells of plants (Delmer and Haigler 2002). Cellulose is the major source of tensile strength for the plant cell wall (Cano-Delgado, Penfield et al. 2003).

The main component of wood is cellulose, it makes up as much as 40-45% of a tree's dry weight (Delmer and Haigler 2002). Due to the high amounts of cellulose produced that provide the tree with its strength, wood from trees have various end uses such as wood chips, fiberboard, lumber, beams, posts, telephone poles, and furniture. Also, cellulose fibers are isolated from wood to produce pulp and used to produce various forms of paper and paperboard.

Cellulose has been described as a partial crystal of glucan chains, made of D-glucose monomers, which is a simple sugar. On average, cellulose in wood has a degree of polymerization of about 10,000. The D-glucose units, anhydroglucopyranose, of cellulose are joined by a β -(1-4) acetal linkage (glycoside linkage) (Delmer and Haigler 2002; Doblin, Kurek et al. 2002).

This glycoside linkage is created by the loss of water when an alcohol and hemiacetal react. Due to the nature of the linkage between glucose units, adjacent anhydroglucopyranose molecules exist in opposite orientation in space. Therefore, the repeat unit of cellulose is generally accepted to be cellobiose (figure 1) (Gibeaut and Carpita 1994).

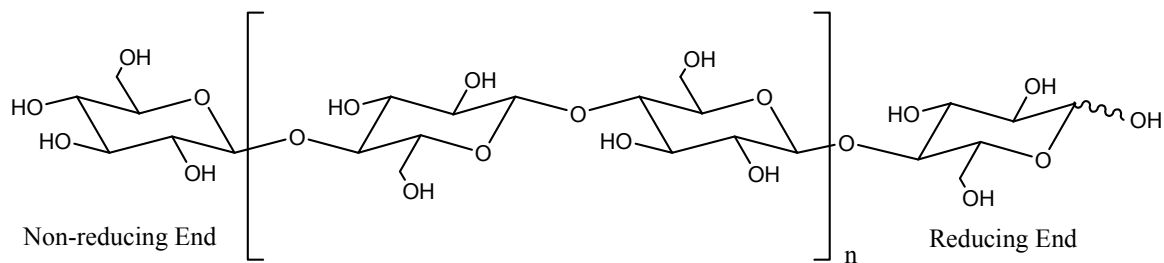


Figure 1.1. Structure of Glucan Chain

The hydroxyl groups of the glucan chain are in the equatorial position and therefore to the side as apposed to above or below the plane of the anhydroglucopyranose ring. This outward orientation of the hydroxyl groups allow multiple chains to interact with each other. Due to the crystalline structure of cellulose, there is the formation of hydrogen bonds.

1.4.1 Cellulose Biosynthesis

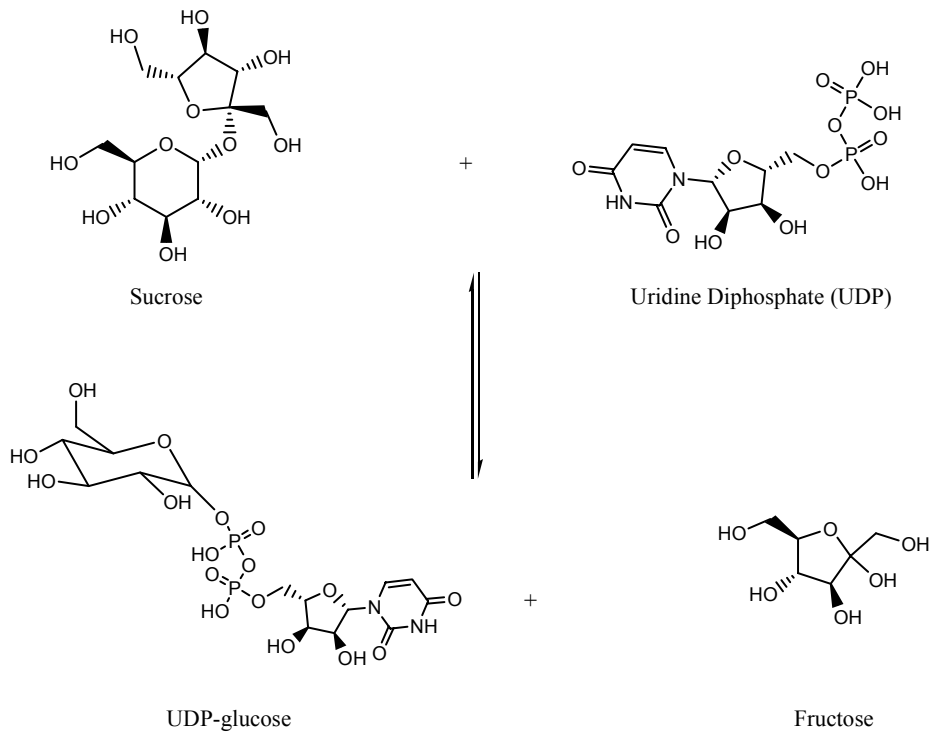
Cellulose has three levels of organization. The first level is the linkage of anhydroglucopyranose units to form a single polymer. Crystallization of numerous cellulose polymers to create a microfibril is the second level. The third and last level is the orientation of microfibrils within the cell wall (Figure2). The orientation of cellulose is altered during cell wall formation. The direction at which cellulose is laid down in the cell wall is controlled by cellulose synthesizing rosettes (Taylor, Howells et al. 2003). In the outer segment of the primary wall, cellulose microfibrils do not contain a set orientation. However, in the inner portion of the primary wall the microfibrils are perpendicular to the cell axis. In the S_1 layer of the secondary wall, microfibrils are

oriented at an angle between 60-80°. The microfibril angle of the middle S₂ layer varies from 5-30°. Lastly, the inner S₃ layer has a microfibril angle that fluctuates between 60-90° (Figure 3) (Plomion, Leprovost et al. 2001).

UDP-glucose is thought to be the immediate substrate for cellulose polymerization (Naki, Tonouchi et al. 1999; Salnikov, Grisom et al. 2001; Delmer and Haigler 2002). Models for the biosynthetic pathway of cellulose suggest the sources of UDP-glucose are complex, which involve numerous possibilities for the recycling of the intermediates (Delmer 1999; Haigler, Ivanova-Datcheva et al. 2001; Delmer and Haigler 2002). As shown in figure 4, the first source of UDP-glucose is suggested to be produced in the cytoplasm via enzymes 6 and 4 that convert glucose-6-phosphate to glucose-1-phosphate followed by the conversion to UDP-glucose by pyrophosphorylase (figure 2) (Amor, Haigler et al. 1995; Babb and Haigler 2001; Delmer and Haigler 2002). This formation of this UDP-glucose in the cytoplasm is noted as being a “free pool” of UDP-glucose which would sustain universal metabolism (Babb and Haigler 2001). The second source of UDP-glucose is converted from sucrose via sucrose synthase (SuSy) by enzymes 2 or 3 (figure 4 and reaction 2) (Amor, Haigler et al. 1995; Babb and Haigler 2001; Delmer and Haigler 2002). Formation of this form of UDP-glucose is suggested to be a “no free pool” of UDP-glucose, suggesting that it is directly fed into the cellulose synthesis machinery (Naki, Tonouchi et al. 1999; Babb and Haigler 2001).

However, there are at least of two forms of SuSy in two locations that may lead to two independent pools of UDP-glucose (Figure 4) (Delmer and Haigler 2002). SuSy may interact with the inner bilayer of the plasma membrane and/or the cytoplasmic portion of the secondary wall cellulose synthase (Salnikov, Grisom et al. 2001). The two forms of

SuSy are known as S-SuSy (the soluble form of SuSy) and P-SuSy (the membrane-associated form of SuSy) (Delmer and Haigler 2002). P-SuSy is present in addition to S-SuSy in non-photosynthetic tissue (Armor, Haigler et al. 1995; Carlson and Chourey 1996; Winter, Huber et al. 1998; Delmer 1999; Winter and Huber 2000; Delmer and Haigler 2002). The membrane-associated form of SuSy is thought to be responsible for the flow of carbon from sucrose to cellulose synthase via the following reaction (Naki, Tonouchi et al. 1999; Delmer and Haigler 2002):



Reaction 1.2. Conversion of Sucrose to UDP-glucose

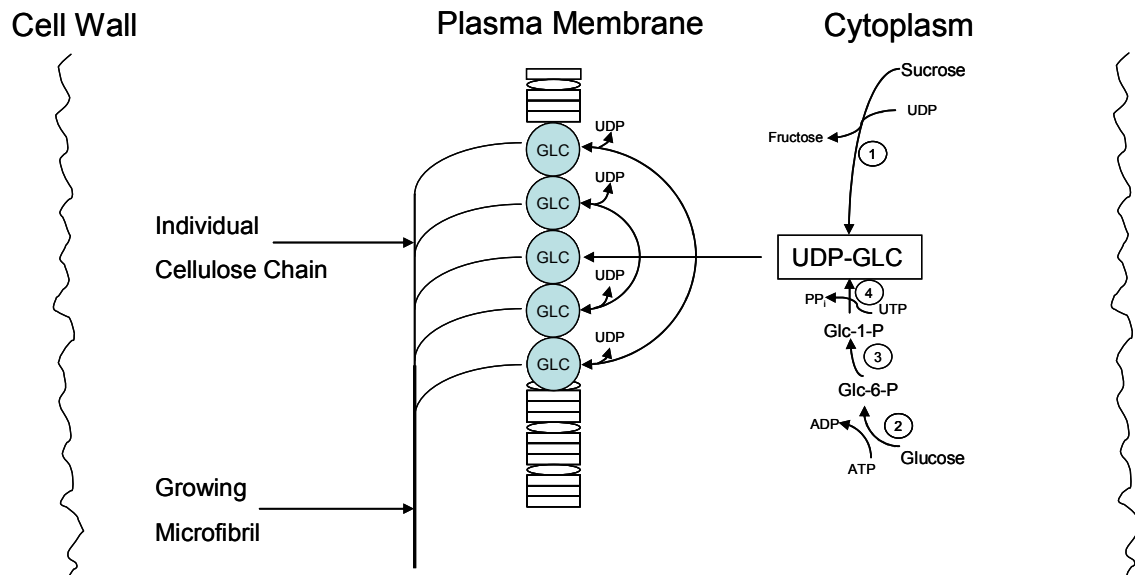


Figure 1.2. Cellulose Organization in Cell Wall

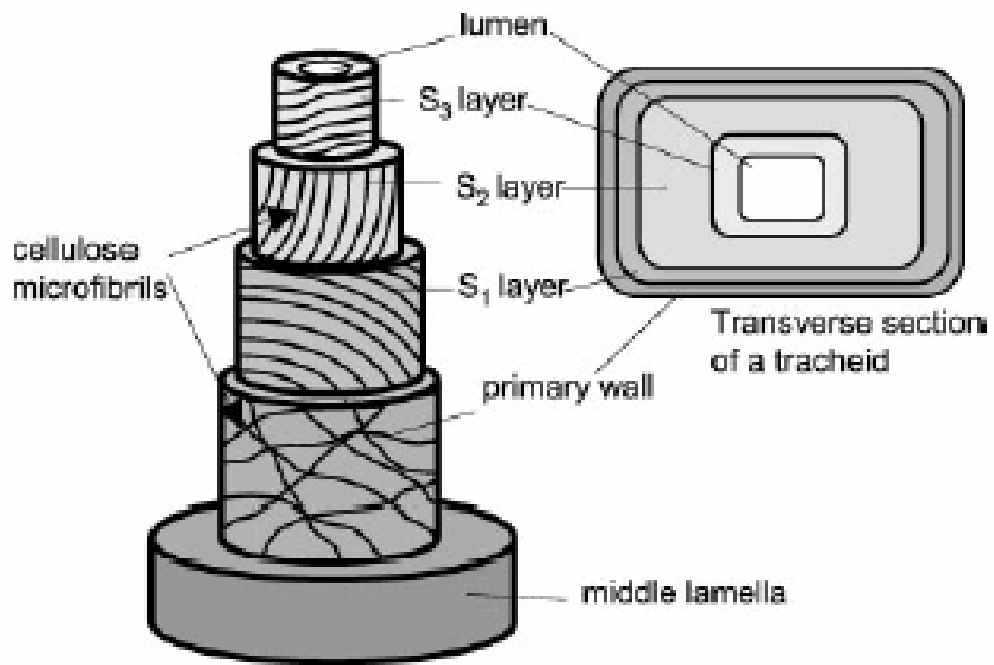


Figure 1.3. Structure of Wood Cell(Plomion, Leprovost et al. 2001)

The following enzymes catalyze reactions that directly or indirectly affect cellulose biosynthesis:

1	CesA: Catalytic subunit of cellulose synthase
2	S-SuSy: Soluble form of SuSy
3	P-SuSy: membrane-associated form of SuSy
4	UDP-glucose pyrophosphorylase
5	Sucrose-P phosphatase
6	Phosphoglucomutase
7	Sucrose-P synthase
8	Fructokinase
9	Glucose-6-P isomerase
10	Phosphofructomutase

Fructose, which is the second product formed from the conversion of sucrose to UDP-glucose, can act as an inhibitor of SuSy (Figure 4) (Naki, Tonouchi et al. 1999; Weckwerth, Willmitzer et al. 2000; Haigler, Ivanova-Datcheva et al. 2001; Delmer and Haigler 2002). For the proposed pathway to operate efficiently it is argued there must be some mechanism to prevent the accumulation of fructose near the site of cellulose synthesis (Delmer and Haigler 2002). Some fructose produced by cellulose synthesis can be involved in other biosynthetic pathways, used for respiration or stored in the vacuole, which is used to store nutrients, metabolites and waste products (Delmer and Haigler 2002). In cotton, the fructose created by the conversion of sucrose to UDP-glucose is rapidly phosphorylated and recycled back to sucrose to be converted by P-SuSy again (Martin 1999).

Thus, plant cellulose biosynthesis can be considered to be a process comprised of three major steps. The first step is the channeling of the UDP-glucose substrate to a cellulose synthase complex, called rosettes, by P-SuSy (Amor, Haigler et al. 1995; Doblin, Kurek et al. 2002). The second step involves the polymerization of UDP-glucose into glucan chains by the rosettes (Figure 5). Upon the completion of the second step, UDP is recycled by to P-SuSy. The third and final major step in cellulose biosynthesis is the monitor of the conversion of the glucan chains into a cellulose microfibril by a membrane associated cellulase, called KORRIGAN (KOR) (Doblin, Kurek et al. 2002).

A rosette is comprised of six subunits containing six cellulose synthase catalytic subunits (CesA) (Doblin, Kurek et al. 2002). Each subunit of the rosette is suggested to synthesize one glucan chain that is eventually added to the other glucan chains that are combined to form a cellulose microfibril (Doblin, Kurek et al. 2002). In the last step of

plant cellulose biosynthesis, KOR acts to remove flawed glucan chains prior to its addition to the cellulose microfibril (Doblin, Kurek et al. 2002).

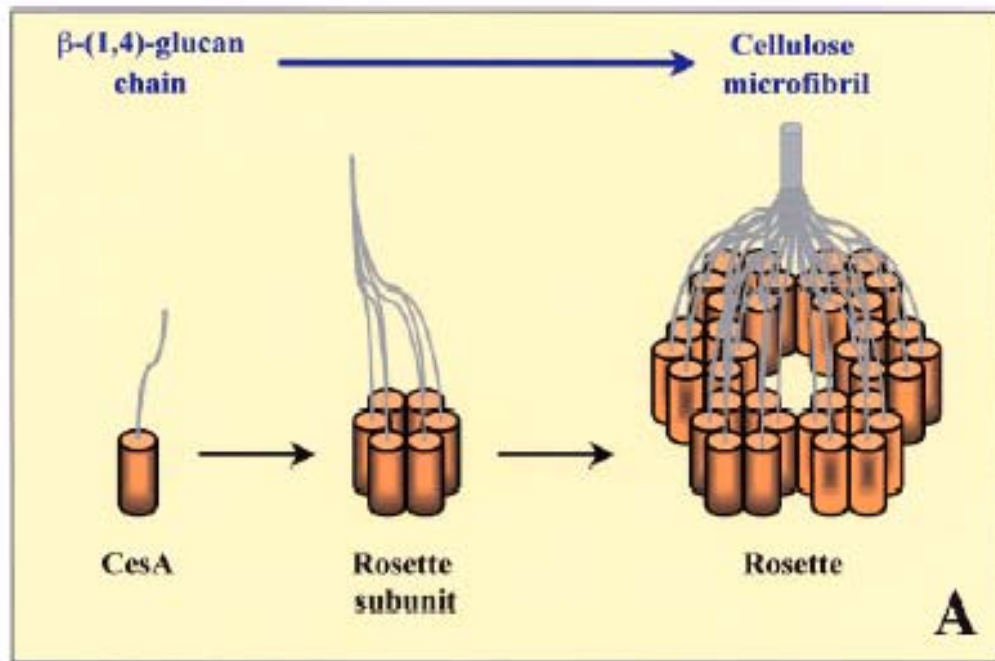


Figure 1.5. Cellulose Rosette Assembly (Doblin, Kurek et al. 2002)

1.5 Hemicellulose

Hemicellulose usually makes up 25-35% of the dry weight of wood. In woody plants, there are two basic types of hemicelluloses: D-glucomannans, and D-glucuronoxylans. Structurally, hemicellulose consists of co-polymers of two or more sugars and sugar acids. The main sugar components of hemicelluloses are xylose, mannose, galactose, glucose, arabinose and 4-O-methylglucuronic acid. In softwood, the major hemicellulose is galactoglucomannans and the minor hemicellulose is

arabinoglucuronoxylans (Figure 6). In hardwood, the principle hemicellulose is glucuronoxylans and the minor hemicellulose is glucomannans (Figure 6).

The main substrate for hemicellulose precursors is NDP-glucose (Johansson, Sterky et al. 2002). NDP-glucuronic acid, a product of dehydrogenation of NDP-glucose by NDP-glucose dehydrogenase, is proposed to be the essential intermediate of hemicellulose precursors (Tenhaken and Thulke 1996; Seitz, Klos et al. 2001). As an essential intermediate NDP-glucuronic acid can be readily converted into NDP-arabinose, NDP-galacturonic acid, and NDP-xylose as seen in figure 7. In Arabidopsis, there is the suggestion of an alternative pathway in the production of NDP-glucuronic acid via the oxidation of inositol to glucuronic acid followed by conversion to its nucleotide sugar (Seitz, Klos et al. 2001).

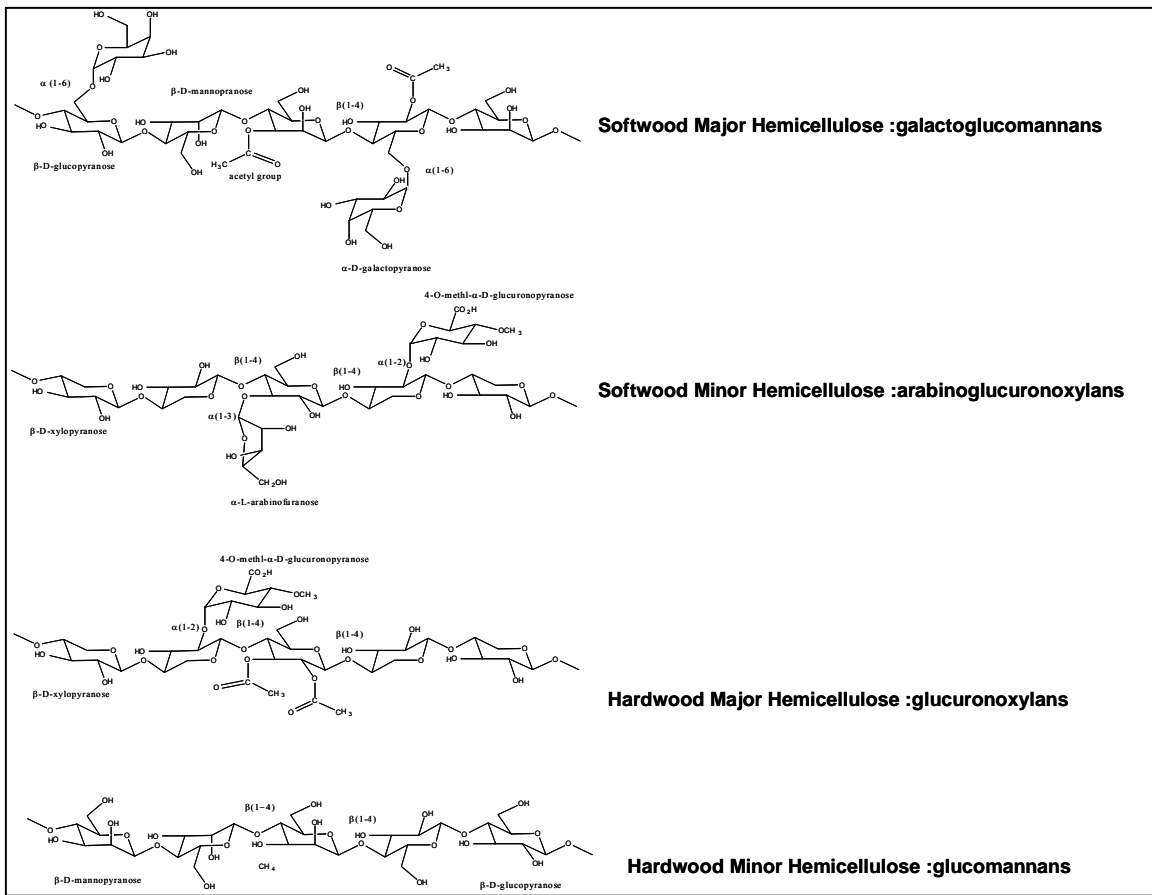


Figure 1.6. Hemicelluloses in Wood

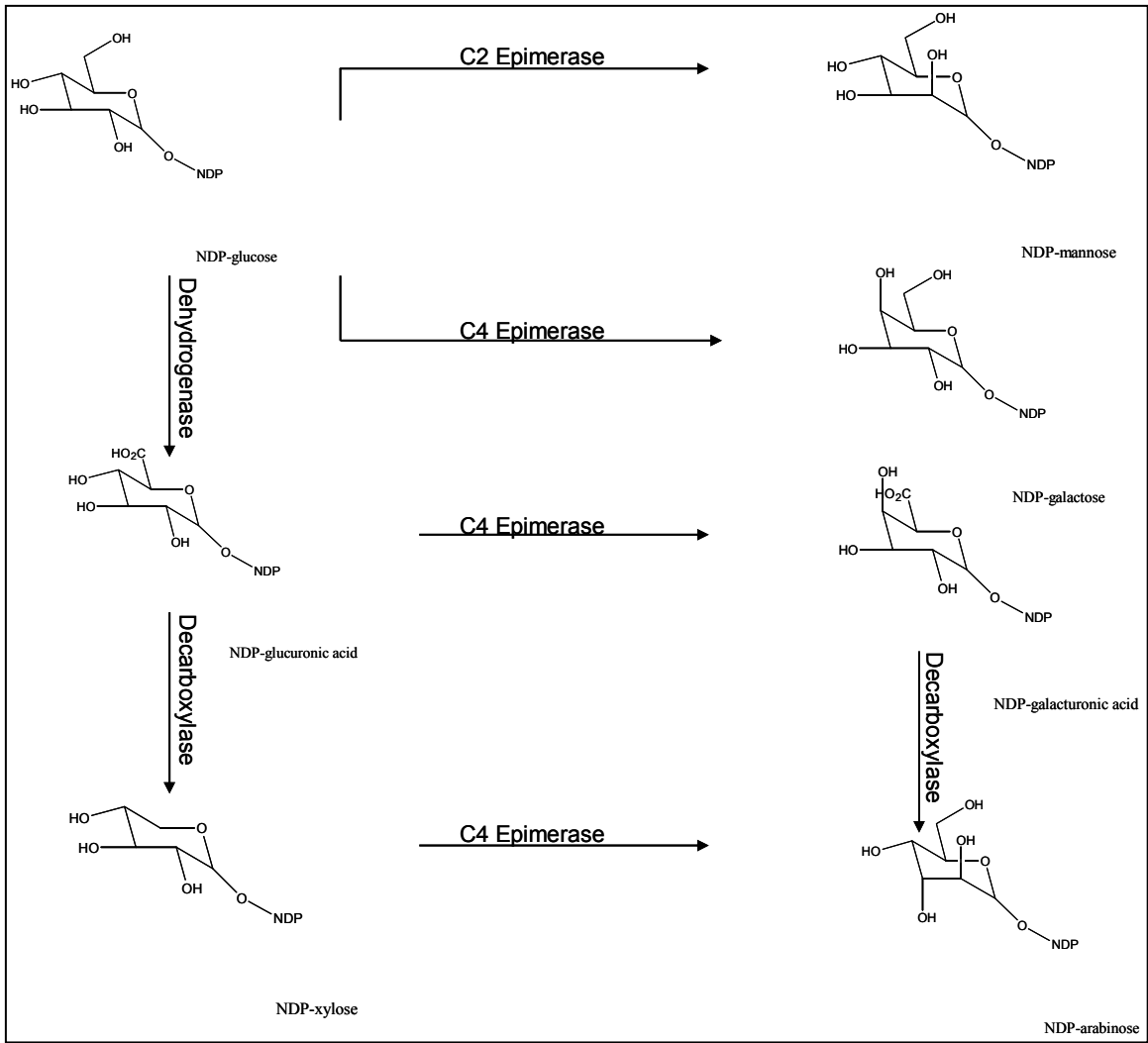


Figure 1.7. Hemicellulose Biosynthesis

1.6 Lignin

A major component of plant secondary walls is lignin. In wood, lignin makes up 20 to 35% of the dry weight of wood on average. It provides plants with its structural support and also protects plants against attack by pest and pathogens. Lignin also assists in transporting water and solute through xylem tissue (Boerjan, Ralph et al. 2003).

1.6.1 Lignin Precursors

Lignin is primarily made of three hydroxycinnamyl alcohol precursors through the coupling of radicals produced by a single-electron oxidation (Boudet 2000). The three lignin precursors are: p-coumaryl alcohol, coniferyl alcohol, and sinapyl alcohol (Figure 8) (Higuchi 2003). Hardwood lignin is composed of various proportions of coniferyl alcohol derived subunits and sinapyl alcohol derived subunits. On the other hand, softwood lignin is derived mostly from coniferyl alcohol monomers. Lignin structure not only consists of subunits derived from these three precursors, in foliage, it also contains derivatives of hydroxycinnamyl alcohol like acylated hydroxycinnamyl alcohols and hydroxycinnamate esters (Ralph, MacKay et al. 1997; Kim, Ralph et al. 2000). Prior to lignification, the precursors may exist as glucosides that accumulate in the cambial tissue of gymnospermous trees (Terashima, Ralph et al. 1995; Matsui, Fukushima et al. 1996; Steeves, Forster et al. 2001). Glucosides may be considered to be the transported form of monolignols (Steeves, Forster et al. 2001). This transported form is required to allow the monolignols to travel through the aqueous phase of the cell to the zone of lignification since the monolignol itself is hydrophobic and will not diffuse

through the cell. Monolignols are also toxic and it may be necessary to have glucosides so that larger amounts can be stored.

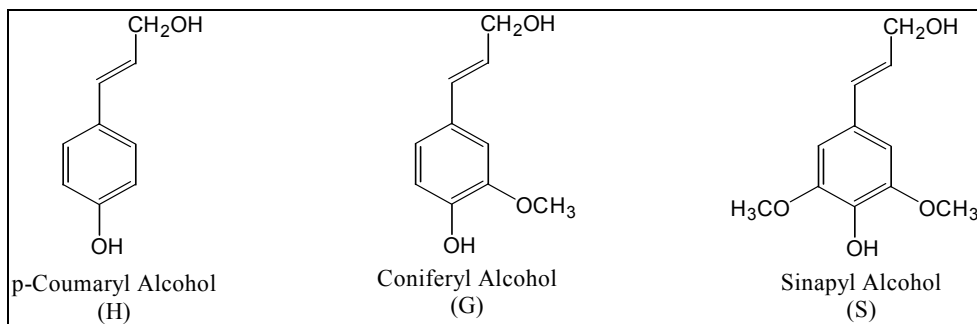


Figure 1.8. Lignin Precursors

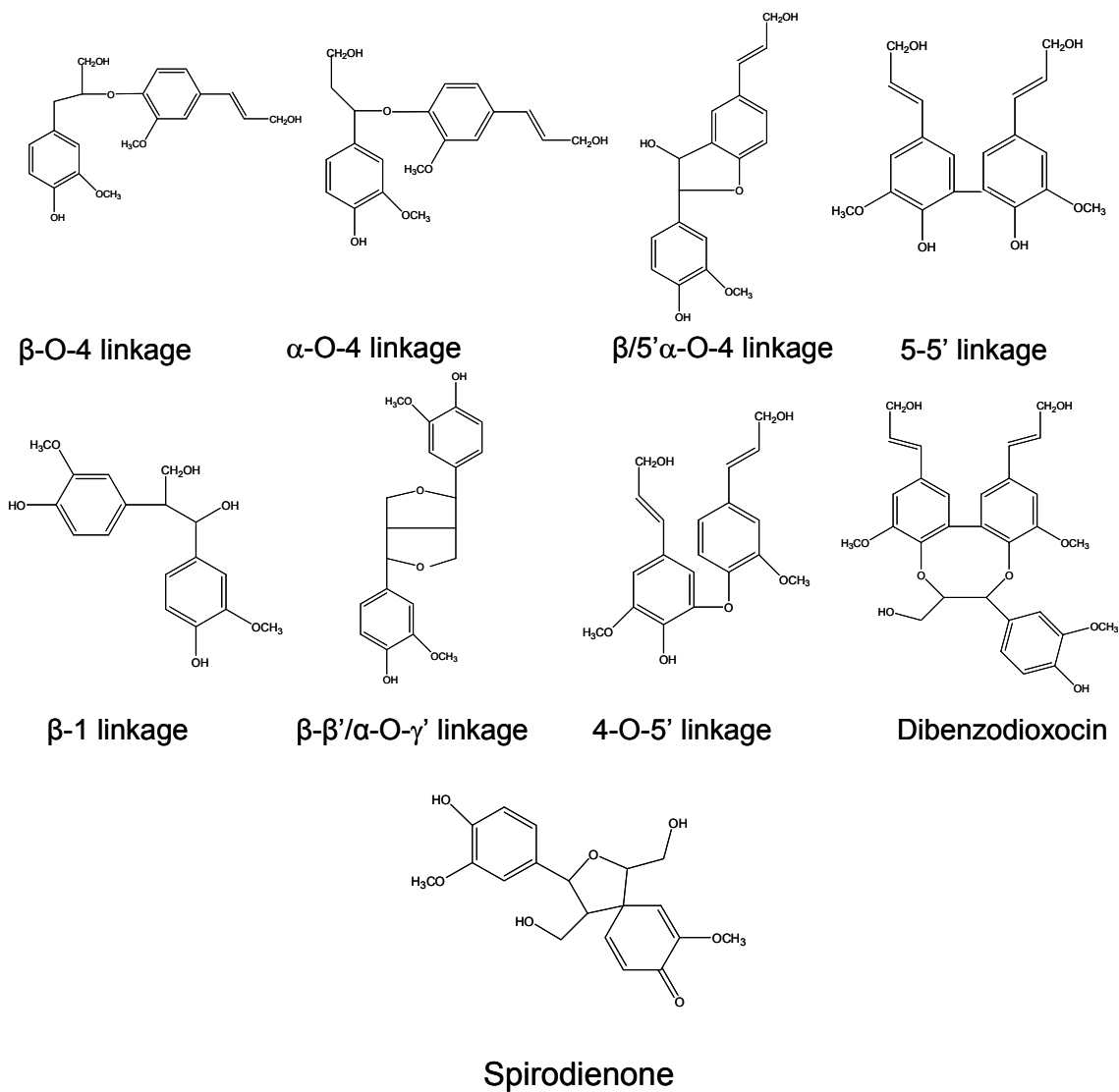


Figure 1.9. Major Lignin Linkages

1.6.2 Monolignol Precursor Biosynthesis

The starting point of monolignol biosynthesis begins with the production of glucose from photosynthesis. Glucose is subsequently converted to shikimic acid and phenylalanine via the shikimic acid pathway (Figure 10) (Higuchi 2003). The first step of the monolignol biosynthetic pathway is the conversion of phenylalanine to cinnamic acid using PAL, phenylalanine ligase, shown in the top left portion of figure 11 (Li, Popko et al. 2000; Boerjan, Ralph et al. 2003; Higuchi 2003). The second step is the conversion of trans-cinnamic acid to p-coumaric acid through cinnamate 4-hydroxylase (C4H), which provides a hydroxyl group at the C-4 position (Figure 11) (Whetten, MacKay et al. 1998). This reaction is driven in the presence of O₂ and NADPH (Higuchi 1998). PAL is suggested as being responsible for controlling the formation of the phenylpropanoids from aromatic amino acids (Higuchi 1998).

Recent developments suggest that the preferred substrates of p-coumarate 3-hydroxylase (C3H) are the shikimic and quinic esters of p-coumaric acid and not p-coumaric acid as previously believed (Schoch, Goepfert et al. 2001; Franke, Humphreys et al. 2002; Nair, Xia et al. 2002). In *Arabidopsis*, thioesterification of p-coumaroyl-CoA is proposed to be produced from p-coumaric acid by 4CL, 4-coumarate-CoA ligase. It is suggested that the next step is esterification with shikimic acid and quinic acid, in separate molecules, to produce the substrates for C3H, p-coumaroyl-shikimate and p-coumaroyl-quininate (Schoch, Goepfert et al. 2001). From this point it is noted that these substrates, caffeoyl-shikimate or caffeoyl-quininate are subsequently converted to caffeoyl-CoA by p-hydroxycinnamoyl-CoA:D-quininate (CQT) or p-hydroxycinnamoyl-CoA:shikimate p-hydroxycinnamoyltransferase (CST) (Ulbrich and Zenk 1980; Schoch,

Goepfert et al. 2001). Once caffeoyl-CoA is produced, caffeoyl-CoA O-methyltransferase (CCoAOMT) readily converts it to feruloyl-CoA, as seen in the second row of figure 11 (Ye, Kneusel et al. 1994; Zobel and Sprague 1998; Li, Popko et al. 2000). Feruloyl-CoA is noted as being the best substrate for cinnamoyl-CoA reductase (CCR), to produce the corresponding coniferaldehyde (Figure 11) (Higuchi 1998).

It was once thought that the 3-hydroxyl group of caffeate was methylated by caffeic acid O-methyltransferase (COMT) to produce ferulic acid (Figure 11). The methyl donor of this reaction was believed to be S-adenosyl methionine (Higuchi 1998). Once ferulic acid was produced, it was thought to be hydroxylated by ferulate 5-hydroxylase (F5H), which requires molecular oxygen and NADPH, to synthesize 5-hydroxyferulic acid (Higuchi 1998). However, feruloyl-CoA is the best substrate for CCR in the production of coniferaldehyde. Coniferylaldehyde is not only a good substrate for F5H, it also has been suggested to inhibit the hydroxylation of ferulic acid to 5-hydroxyferulate (Humphreys, Hemm et al. 1999; Osakabe, Tsao et al. 1999). These recent findings have created a more direct route to monolignol synthesis of coniferyl alcohol as seen by the following pathway in figure 11.

Since coniferaldehyde is utilized to synthesize syringyl lignin monomers, the sinapoyl-CoA ligase activity is not required. It has been determined that coniferylaldehyde 5-hydroxylase (CAld5H) is responsible for producing 5-hydroxyconiferaldehyde, as seen in the third row of the pathway (Figure 11) (Osakabe, Tsao et al. 1999). The next step converts 5-hydroxyconiferaldehyde to sinapaldehyde by 5-hydroxyconiferyl aldehyde O-methyltransferase (AldOMT). AldOMT was found to

catalyze this reaction rather than COMT, through the use of enzyme kinetics coupled with mass spectrometry (Osakabe, Tsao et al. 1999; Li, Popko et al. 2000).

The last step of monolignol biosynthesis, seen in the bottom of figure 11, involves the reduction of the hydroxycinnamaldehydes to its analogous alcohol, in the presence of NADPH (Savidge and Forster 2001). The products of this reaction produce the three monolignols that operate as lignin precursors. The enzyme responsible for this reduction of p-coumaryl alcohol and coniferyl alcohol is cinnamyl alcohol dehydrogenase (CAD) (Higuchi 2003). Li et al. discovered a novel enzyme, sinapyl alcohol dehydrogenase (SAD), that is responsible for the reduction of sinapaldehyde to sinapyl alcohol in angiosperms (Li, Cheng et al. 2001).

Although CAD plays a very important role in monolignol biosynthesis, it is not considered the rate-limiting step. Also, it is noted that CAD plays a minor role in regulating metabolic flux through the lignin biosynthetic pathway (MacKay, Omalley et al. 1997). However, a little CAD activity is required to create the lignin structure found in wild-type loblolly pine.

1.6.3 Lignin Polymerization

The exact mechanism of monolignol polymerization is still under debate. The common belief is that monolignols are converted into free phenoxy radicals that then polymerize spontaneously (Freudenberg and Neish 1968). The enzymes thought to be responsible for the dehydrogenation of monolignols are peroxidases and laccases (Higuchi 1997). The hydrogen peroxide required for the peroxidase reaction could be from the reduction of oxygen by glucose oxidase and glucose, which is cleaved from the

monolignol glucoside (Freudentberg and Neish 1968). The first step of lignin polymerization involves the coupling of radicals to form a dilignol, in a nonrandom process. Despite have a number of possibilities for linkage, only 3 linkages are prevalent as coupling products: β -O-4, β -5, β - β . As a result of the coupling process, each of the dilignols formed have at least one phenolic hydroxyl group. This phenolic hydroxyl group can be dehydrogenated by a peroxidase to produce dilignol radicals. Subsequent dilignol radicals can couple with a monolignol or another dilignol to produce a trilignol or a tetralignol. In either product, there is at least one phenolic hydroxyl group for further dehydrogenation. This repetition of dehydrogenation and coupling of lignol radicals form the basis of the polymerization process of lignin.

Numerous lignin linkages have been described in literature. Figure 9 displays some of the major linkages in lignin. The major linkage in lignin is β -O-4. In monolignol coupling, the ratio of β -5 linkages is greater than β -O-4 linkages, which is approximately equal to β - β linkages. However, the predominant gradual addition of monolignols to the growing lignin polymer in the cell wall produces a high ratio of β -O-4 linkages that is greater than β -1 linkages which is greater than β -5 linkages. In the case that monolignols are released at once and not gradually, there would be a higher number of β -5, 5-5', and 5-O-4 linkages. Couplings between oligolignols have greater ratios of 5-5', 4-O-1, 4-O-5, 1-5 linkages. The β -O-4 coupling products produce a quinone methide intermediate, which plays a major role in the lignification process. Many things can add to quinone methide such as carbohydrates to form lignin carbohydrate complexes to bond lignin to cell wall polymers, cellulose, and hemicellulose (Whetten, MacKay et al. 1998).

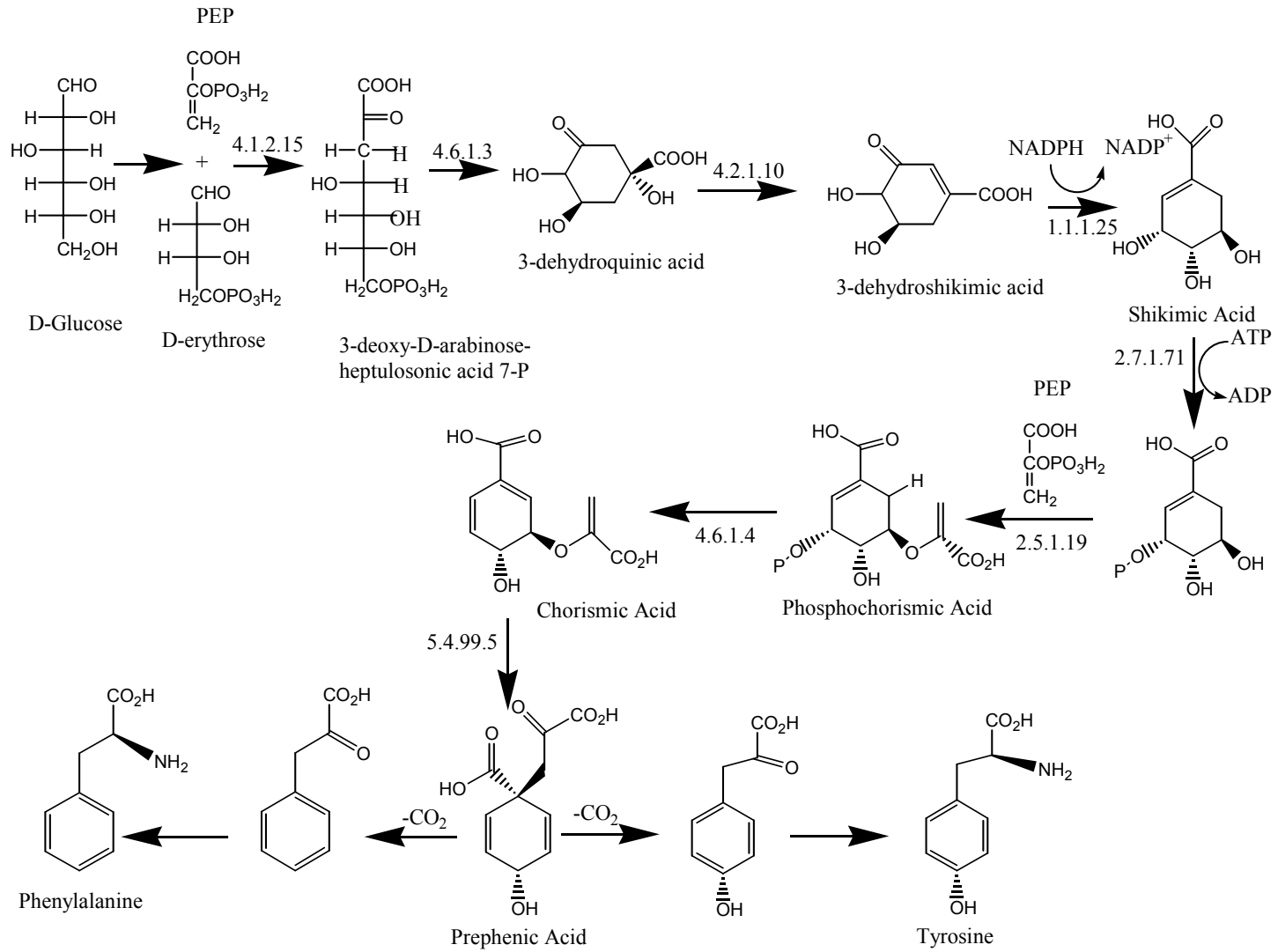


Figure 1.10. Shikimic Acid Pathway

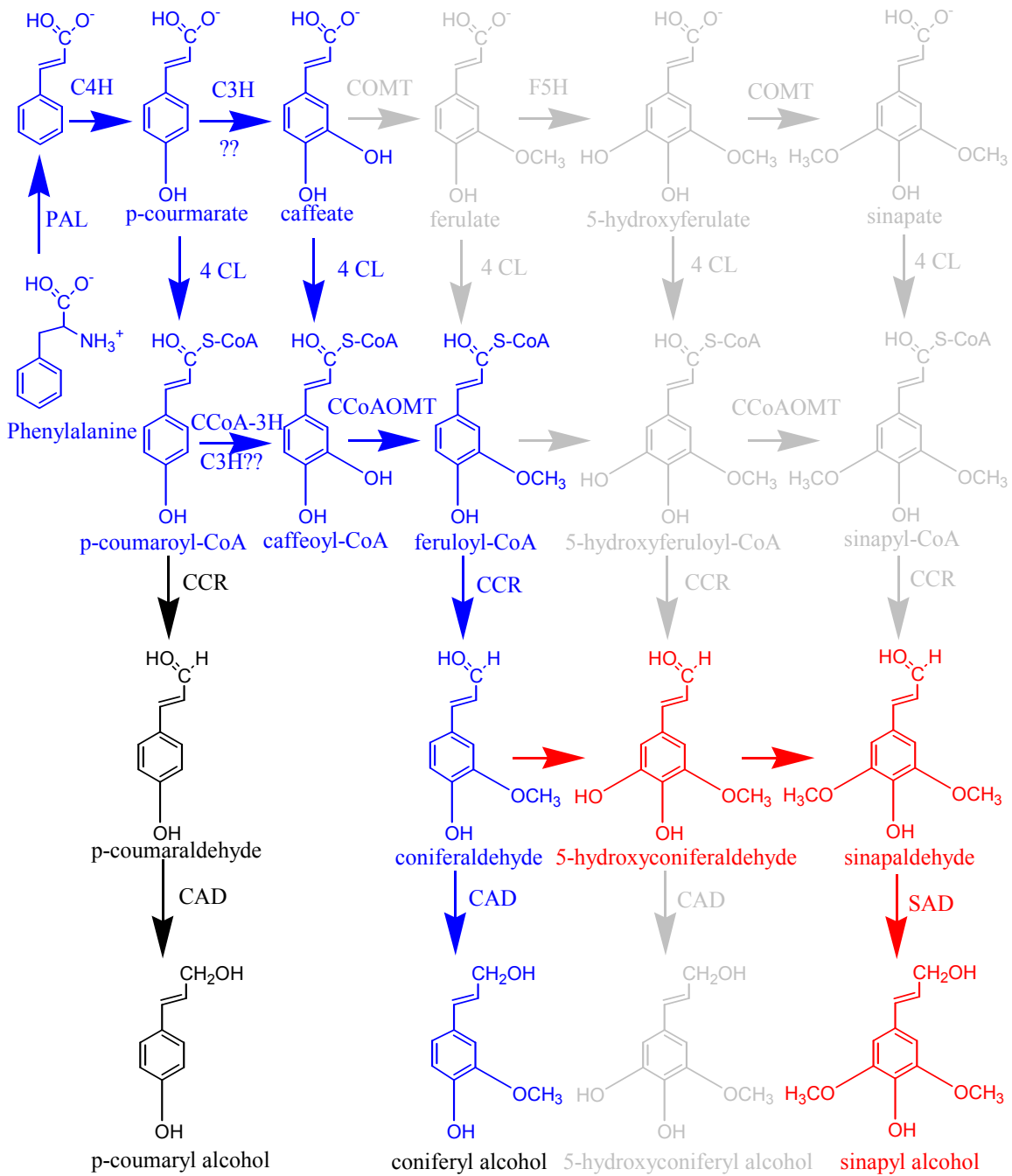


Figure 1.11. Monolignol Biosynthesis Pathway. Overall scheme of pathway: a deamination, hydroxylation of aromatic ring, methylation, thioesterification, and reduction of the thioester to an alcohol

1.7 Extractives

Extractives make up a small percentage (2-3%) of the dry weight of wood. Much like cellulose, hemicellulose, and lignin, extractives metabolism starts with the sugar pool created by photosynthesis. Figure 12 provides a simple schematic of the production of these extractives. Among the extractives in trees are terpenoids, fatty acids, phenolics, and tannins. Terpenoids and fatty acids are constructed via pathways through phosphoenol pyruvate and Acetyl CoA. Phenolics are produced from the shikimic acid pathway and also converted from Acetyl CoA. Lastly, tannins are produced solely through the shikimic acid pathway.

The main building block for fatty acid synthesis is acetyl-SCoA. The intramitochondrial oxidation of pyruvate serves as the sole cellulose source of acetyl-SCoA. Due to the fact that the mitochondrial membrane is impermeable to acetyl-SCoA, it is suggested that the additional mitochondrial production of acetyl-SCoA occurs through three possible paths. In the first path, it diffuses as citrate and decomposes into acetyl-SCoA and oxaloacetate via the citric acid cycle. The second path diffuses as acetate and regenerates acetyl-SCoA with acetate thiokinase. In the third pathway, carnitine serves as acetyl carrier.

Terpenoids are derived from isoprene units that are coupled 'head to tail' or 'tail to tail' (Obst 1998). Each isoprene represents a five carbon building block for the terpenoid structure. For example, a monoterpene is made up of two isoprene units totaling 10 carbons in the structure (Obst 1998). Likewise, diterpenes, resin acids, have a total of 20 carbons in the structure. Terpenoids can also exist in various cyclic forms

such as monocyclic and bicyclic monoterpenes. Limonene, a cyclic monoterpene, is major component of turpentine in the *Pinus species* (Obst 1998).

Phenolics are produced via two major pathways. These pathways are the shikimic acid pathway and the acetate pathway. These compounds are distinguished by the presence of a hydroxyl group, attached to a benzene ring. Some phenols are produced solely by the acetate pathway, such as gallic acid, and phenols such as lignans and coumarins are produced by the shikimic acid pathway. In addition, stilbenes and flavonoids are derived from a combination of both shikimic and acetate pathway. Stilbenes are a class of microbial agents. Flavonoids provide trees with resistance to fungal and insect attack.

Tannins are a class of plant polyphenols that convert animal skins into leather. There are two main classes of tannins: hydrolyzable tannins and condensed tannins. Condensed tannins are polymers of flavonoids; however, it is unclear whether these tannins are polymerized via oxidation by enzymes or acid catalyzed. Hydrolyzable tannins are sugar esters of gallic acid and/or ellagic acid.

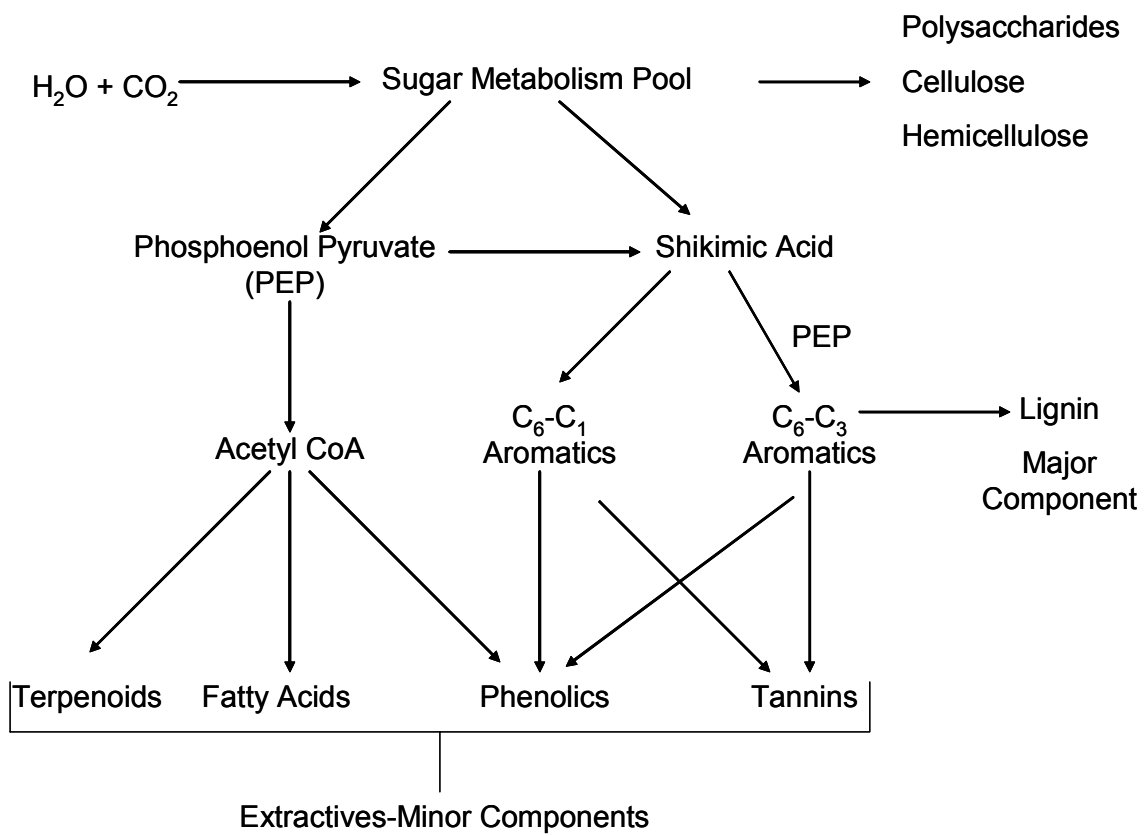


Figure 1.12. Simplistic Extractives Biosynthetic Scheme

1.8 Functional Genomics

Functional genomics is coined as being “the study of all the genes expressed by a specific cell or group of cells and the changes in their expression pattern during development, disease, or environmental exposure” (Freeman, Roberston et al. 2000). The purpose of functional genomics is to define the function of all gene sequences by many methods including examining the activity of the encoded protein and the role of that particular protein within a cell, tissue, or organism. Recently, two new tools have been developed to aid in understanding gene function; metabolic profiling and microarrays. Metabolic profiling is a powerful tool that can provide information towards the understanding of gene function in the context of metabolic networks (Trethway 1999), while microarrays also provide information about gene function (Cheung, Morely et al. 1999). Microarrays have the capability to profile the patterns of transcript abundance for thousands of genes in a solitary experiment (Duggan, Bittner et al. 1999).

1.9 Metabolomics

Metabolic analysis can be divided into four areas: metabolomics, metabolic profiling, metabolic fingerprinting, and target compound analysis. Metabolomics is often defined as the measurements of metabolites in an organism’s cells at a specific point in time under specific environmental conditions (Fiehn 2002; Weckwerth 2003). Metabolic fingerprinting is sample classification by a rapid and global analysis with high throughput (Fiehn 2002). Target or screening analysis is the extraction and quantitation of specific metabolites within the samples. Metabolic profiling is the quantitative measurement of a

group of related metabolites to determine the function of an entire pathway or intersecting pathways (Fiehn 2002).

1.10 Metabolic Profiling

Metabolic profiling is a multi-component chromatographic analysis of extractible metabolites or compounds from plants and animals. The method is designed to correlate metabolic states at different points in time or state using characteristic chromatographic profiles. The most important step in metabolic profiling experiments is the extraction step. The vitality of the profiling technique is the ability of an extraction protocol to consistently and reliably recover and quantify as many metabolites as possible in a single measurement. Although not all the intermediates of the metabolic pathways are detectable the pooled metabolites serve as projectors and regulators of the various complex metabolic networks (Trethway 1999).

In order for metabolic profiling to be useful, some general requirements must be met. The instrumental and chemical methods chosen must be appropriate for the analysis of very large number of metabolites of different classes present in a biological sample. Due to the large number of metabolites, these components must be analyzed with a high degree of selectivity. The analytical methods must separate and identify the various components (Holland, Leary et al. 1986).

The next requirement is that the various metabolites must be measured with a detection system that does not bias to the quantitation of the different metabolites. There is an extremely large dynamic range with which the measurements of metabolites must be made. Metabolites found in biological mixtures vary widely in concentration, thereby instruments must be able to identify and precisely measure a metabolite in the mixture

that may be substantially more dilute than another metabolite in the mixture (Holland, Leary et al. 1986).

1.10.1 Instruments for Metabolic Profiling

Paper Chromatography

Paper chromatography is a method for the separation of complex mixtures. It is a relatively quick and can be carried out with small quantities of material. The separation process in paper chromatography involves the same principles as thin layer chromatography in that the substances are distributed between a stationary phase and a mobile phase. The stationary phase is usually a piece of high quality filter paper, while the mobile phase is a developing solution that travels up the stationary phase carrying the samples with it. The proportions of mixture separation are based on the compounds ability to absorb to the stationary phase versus how much they dissolve in the mobile phase.

Paper chromatography was originally utilized to follow chemical reactions and screen compound purity. However, it also found use in the analysis of metabolites. R.J. Williams first applied the theory of metabolite profiling using paper chromatography to create metabolic patterns of individual men (Williams 1951; Holland, Leary et al. 1986). In 1956, Armstrong et al. utilized paper chromatography as a method to detect phenolic acids in urine. The method identified patients who were drinking coffee during the study (Armstrong, Shaw et al. 1956).

Thin-Layer Chromatography (TLC)

A few years later researchers turned to TLC as the choice of analysis for metabolite analysis. As with paper chromatography, TLC was a technique utilized to separate organic compounds. Due to the speed and simplicity of the technique, it is frequently used to examine the progress of organic reactions and to check the purity of precursors and reaction products. TLC consists of a stationary phase immobilized on a glass or plastic plate and a solvent. The sample is introduced as a spot on the bottom portion of the stationary phase. By running standards with the unknown products, the components of the sample can often be identified. The bottom portion of the plate is placed in a solvent, which travels up the plate by capillary action. Once the solvent reaches the top of the plate, the plate is removed from the solvent and the separated spots are visualized with ultraviolet light. Much like paper chromatography, the components of the sample are separated based on the rates at which they diffuse up the plate with the solvent.

TLC provided several advantages over paper chromatography. The first advantage was that minimum amounts of metabolites could be processed and evaluated with less extensive preparation without loss of sample. Another advantage of TLC is that during the development of the chromatogram, the spots or bands diffuse less. Also, once the TLC plate is developed, it is amenable to direct staining and densitometry, quantitative scanning method by photoelectric means, which is used to create the metabolic profiles. The profiles were created by two different methods. This first is created by photographing the stained plate with the print serving as the template to be compared to other profiles. The second method was created by the densitometer

recordings from linear scans of the plate, which provides a more quantitative estimator of components (Holland, Leary et al. 1986).

Gas Chromatography/Mass Spectrometry (GC/MS)

Over the past couple of decades, gas chromatography-mass spectrometry (GC/MS) has become the premiere technique for metabolic profiling. Horning and Horning first developed metabolic profiling through the utilization of GC/MS in 1971 (Williams 1951; Holland, Leary et al. 1986; Niwa 1986). This technique was applied to the analysis of steroids, organic acids, and drug metabolites in human urine. The ultimate goal of this work was to use the profiles to study pathological conditions and drug metabolism, as well as investigate the development of humans (Holland, Leary et al. 1986; Niwa 1986).

GC/MS is an instrument that separates and detects chemical mixtures. The GC component contains a column, and a sensitive detector, which is the MS component that sends a signal to a computer component that, collects the data. For this instrument, the sample is injected into the GC inlet, where it is vaporized immediately due to the high temperatures of the inlet. The sample then travels to the detector by the carrier gas, which is generally helium. In the column, the sample undergoes the normal separation process based on the specifics of the column packing material.

Once the sample components travel through the column, they enter into the capillary column mass spectroscopy detector interface, which connect the GC and MS. The interface aids in concentrating the sample by removing the carrier gas prior to entering the ionization chamber. When using GC/MS typically two forms of ionization

are applied: electron impact (EI) or chemical ionization (CI). CI is a softer ionization technique and creates less fragmentation than EI, which allows this ionization method to provide information on the molecular ion. However, EI produces a stronger fragmentation pattern, which provides information about the structure of the compound through the analysis of fragmentation losses.

After ionization occurs, a small electrical potential is used to repel the ions out of the ionization chamber to the mass analyzer. There are two kinds of mass analyzers or filters available for the GC/MS system. They are a quadrupole and ion trap. There are various limitations and benefits for both the quadrupole and ion trap.

The benefits of quadrupole mass spectrometer are that it is a classical mass analyzer that provides good reproducibility on small low costs systems. The low-energy collision-induced dissociation (CID) MS/MS spectra in triple quadrupole provide efficient conversion of precursor to product. The limitations of the quadrupole mass spectrometer are that it has limited resolution, and the peak heights vary as a function of mass and must be consistently tuned. Also, quadrupole mass analyzers are not well suited for pulse ionization methods, and the CID MS/MS spectra depend strongly on energy, collision gas and pressure.

The benefits of the ion trap mass spectrometer are that it provides the best mass resolution of all mass spectrometers, and it contains powerful capabilities for ion chemistry and MS/MS experiments. The ion trap is also well-suited for pulse ionization techniques and provides non-destructive ion detection. The limitation of the ion trap is that it has limited dynamic range and requires an external source to maintain low-pressure during most analytical applications. Artifacts such as sidebands are present in

mass spectra produced by the ion trap. Based on the construction of the ion trap, parameters such as excitation, trapping, and detection conditions can alter the quality of the mass spectra.

Once the mass analyzer separates the ions, the analyte enters the detector and then goes on to the amplifier, where the signal is subsequently boosted. The computer records all the data created by the detector, and converts the electrical pulses into visual display. At the completion of a run, two methods can be used in the identification of a compound. This identification is made through the availability of a library of known mass spectra by comparing the spectra created by the run for the compound to all compounds in the library. Retention time for a particular column can also be used in conjunction with the library to aid in the correct identification of a compound by running standards on the column.

GC/MS has been the instrument used to conduct metabolic profiling as early as 1971. An adoption of the technique of metabolic profiling of plant tissue using GC/MS was first introduced by Sauter, in 1991 (Sauter, Lauer et al. 1991). The profiling technique was used to study the effect of herbicide treatments on barley leaves (Sauter, Lauer et al. 1991). Recently, Fiehn *et. al.* developed a protocol to study metabolites in *Arabidopsis thaliana* leaf tissue (Fiehn 2000). The profiling procedure involves a methanol and chloroform extraction of the plant tissue, followed by phase separation into polar and non-polar phases. The corresponding phases are methoxylated and trimethylsilylated for the analysis of total fatty acids, fatty alcohols, sterols, and aliphatics, as well as hydroxy-amino acids, amino acids, sugars, sugar alcohols, organic monophosphates, (poly)amines, and aromatic acids (Fiehn, Kopka et al. 2000). By using

GC/MS analysis of the extracts, over 300 metabolites were quantified, with 15 of them being uncommon to plants (Fiehn, Kopka et al. 2000). Figure 13 shows an example of a polar phase profile of the leaf extract.

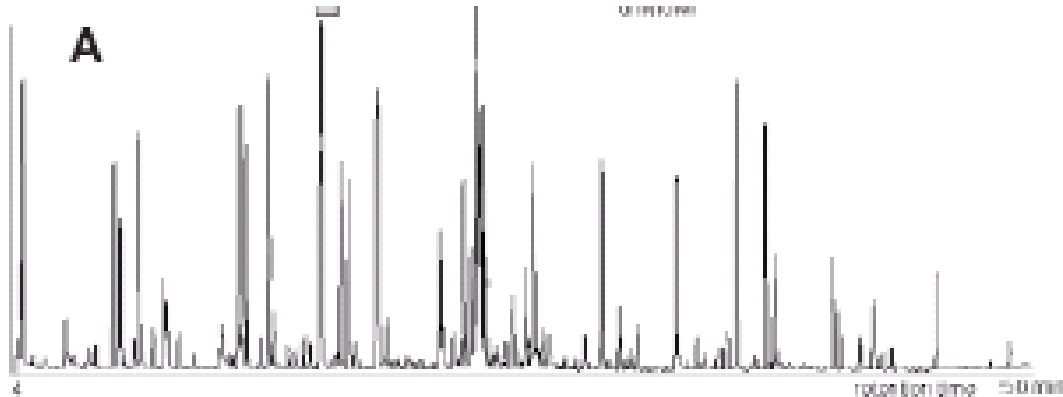


Figure 1.13. Metabolite Profile of Polar Phase of Leaf Extract from Arabidopsis (Fiehn, Kopka et al. 2000)

High-Performance Liquid Chromatography/Mass Spectrometry (HPLC/MS)

High performance liquid chromatography is an important analytical tool for separating and quantifying components in complex liquid mixtures. The efficiency of the separation process is greatly increased by using high-pressure pumps. By selecting the appropriate column and detector, the HPLC system has the capability to separate out

samples with a wide range of components varying from small inorganic and organic molecules to polymers and proteins (Skoog, Holler et al. 1998).

The HPLC instrument is comprised of several components. The first component of the system is the reservoirs that hold solvents, which make up the mobile phase. The next component in some systems could consist of a solvent degasser, which is used to prevent bubbles in the mobile phase. Programmable pumps provide a steady high pressure with no pulsating and also allow the mobile phase to be run as gradient by varying the composition of the solvents during the experiment. The next component is the column housing, which is equipped with a thermostat that aids in controlling the temperature of the column. Current HPLC instruments also come equipped with an auto-sampler that draws a determined volume from the sample vial and injects onto the column (Skoog, Holler et al. 1998). Finally, most HPLC systems are equipped with some sort of detection system, typically a refractive index and/or UV-Vis detector such as a photo-diode array detector to monitor the entire UV-Vis spectrum of the column effluent at frequent intervals.

HPLC coupled with electrospray ionization (ESI) MS is rapidly replacing GC/MS techniques for analyzing metabolites of interest in metabolic profiling. ESI MS uses a needle that has an applied potential at the end to force the sample solution through to produce charged droplets. The potential applied is high enough to diffuse the solution into a spray of like charged droplets. The charge concentration at the surface of the droplet gets increased as the solvent evaporates and the size of the droplet decreases. Eventually, a Coulombic explosion takes place, when charge repulsion overcomes the surface tension of the droplet and the droplet explodes. This explosion creates smaller,

lower charged droplets. Droplets continue to shrink and explode until distinct charged analyte ions are formed. These charges are dispersed throughout the available charge sites of the analyte, which create multiply charged ions under the correct conditions. By introducing a counter current flow of drying gas, the extent of multiple charged ions can be increased. This technique has been found to be a capable tool for the determination of molecular weight (Evtuguin, Domingues et al. 1999).

As mentioned in the GC/MS section, mass spectrometry is used to identify the molecules separated out by chromatography. The mechanism of mass spectrometry involves molecules being ionized and isolated based on each individual molecules' mass-to-charge ratio (m/z) (Glassbrook, Beecher et al. 2000). Not only can the masses of a molecule be measured, the masses of a molecule's fragments can also be determined. This can be helpful in determining the identity of a molecule by analyzing the fragmentation pattern of a molecule at a particular m/z value.

Pros and Cons of HPLC/MS versus GC/MS

Metabolic profiling coupled with mass spectrometry has only recently been used for plant research. In order to increase the usage of metabolic profiling, it must be shown to be fast, reliable, sensitive and have the ability to cover a large number of metabolites (Fiehn, Kopka et al. 2000). As a result, a HPLC/MS system could be used as the mode for metabolite separation in hopes to expand the number of monitored metabolites due to the limited mass range of the GC/MS. This mode of separation has been applied to biomedical research as well as for the detection of plant isoprenoids (Stanley, Newport et al. 1978; Gelpi 1992; Fraser, Elisabete et al. 2000; Chace 2001; Cremin, Kasim-Karakas

et al. 2001). There are other classes of compounds that could be better separated by a HPLC system rather than a GC system such as lipids (Fiehn 2001).

Despite liquid chromatography's use in clinical research, GC/MS has been found to be the most established technique for metabolic profiling of plants (Roessner, Leudemann et al. 2001a). There are several differences that apply between liquid and gas chromatography use for metabolic profiling. The main advantage of the GC/MS is that compounds can be identified more easily by mass spectrometry due to the samples complete interfacing with the mass spectrometry. The other advantage of the GC/MS is that open-tubular capillary columns usually provide better separation than HPLC (Holland, Leary et al. 1986). The main advantage of HPLC/MS is that samples do not require derivatization for the detection of the metabolites, whereas GC/MS requires derivatization. Also, HPLC/MS can identify higher molecular weight compounds that cannot be detected by GC/MS (Holland, Leary et al. 1986). Therefore, the utilization of HPLC will support the findings of the GC/MS as well as provide more information on the higher molecular weight compounds.

1.10.2 Sample Preparation

In metabolic profiling of plant tissue, one of the most important steps is tissue collection. During tissue collection, it is imperative to stop enzymatic activity immediately; one method involves immediate immersion in liquid nitrogen. It is critical that the tissue not partially thaw before the extraction of the metabolites, because enzymes may cause changes in metabolite levels or create new metabolites. One way to work around this problem is to add an organic solvent (Fiehn 2001). However, this is

difficult to accomplish when collecting tissue out in the field do to the lack of resources available.

Prior to the extraction of metabolites, it is important to grind the tissue into a fine powder with liquid nitrogen. There are several methods of grinding reported to achieve a fine powder such as a ball-mill or pestle and mortar (Fiehn, Kopka et al. 2000; Fiehn 2001). After the tissue is ground, the polar metabolites are extracted using a organic solvent like methanol ((Fiehn, Kopka et al. 2000; Fiehn 2001). Once the polar metabolites are extracted, a non-polar solvent such as chloroform is added to extract out lipophilic metabolites (Fiehn 2001). At the completion of the extractions and fractionation the metabolites in each phase are derivatized and run on the GC/MS.

1.10.3 Quantitation of Compounds

In metabolic profiling, the quantitation of a compound is determined by the ratio of the peak height or peak area of the compound and internal standard. This ratio estimates the relative concentration of the metabolite. The choice of internal standard is critical and is typically a chemical analogue of the analytes being measured with similar extraction and derivatization behavior, e.g. isomers, or compounds labeled with stable isotopes (Niwa 1986; Fiehn, Kopka et al. 2000). The relative concentration is calculated from a total or reconstructed ion-monitoring chromatogram, mass chromatogram or selective ion monitoring (SIM) chromatogram (Sweeley, Elliot et al. 1966; Niwa 1986).

1.10.4 Statistical Analysis

Large amounts of data are generated by metabolic profiles; therefore the chromatographic data is stored and analyzed utilizing statistical analysis. Principle component analysis, PCA, is a technique of multivariate analysis that can categorize large amounts of data in metabolic profiling experiments (Fiehn, Kopka et al. 2000; Roessner, Leudemann et al. 2001a; Jeong, Jiang et al. 2004). Statistical analysis defines the technique of PCA as transforming large data sets into a smaller number of meaningful underlying variables called principal components. Metabolic profiling in combination with PCA in this study is used to 1) separate the samples into metabolic phenotypic clusters and 2) indicate the relationships of metabolites.

A great deal of work has been done with this metabolic profiling work over the past five or six years. Recent work has been conducted extensively on the leaves of *Arabidopsis* and potato tuber (Fiehn, Kopka et al. 2000; Roessner 2001b). Metabolic profiling also has an impact in agriculture by observing fruit development of transgenic tomato plants (Roessner-Tunali, Hegemann et al. 2003). Work has also been conducted on Aspen leaves to observe the comparison of sink-to-source transition in developing leaves (Jeong, Jiang et al. 2004). Most recently, metabolic profiling has been conducted on the xylem tissue of loblolly pine (Morris, Scott et al. 2004)

1.11 Microarrays

Another research tool utilized by the field of functional genomics is known microarrays. One of the first microarray experiments were conducted on yeast in 1997 (DeRisi, Iyer et al. 1997). Almost two years later, the first microarray experiments were conducted on the whole genome of *E. coli* and *M. tuberculosis* (Behr, Wilson et al. 1999;

Richmond, Glasner et al. 1999). As the microarray techniques improved over the next couple of years the experiments were applied to other systems such as soy beans, *Arabidopsis* and loblolly pine.

Microarrays are used to study transcript level variation for thousands of genes. This analysis of gene expression measures the relative abundance of gene transcripts, not only the rate of transcription, which requires an estimate of stability of the transcript. For a number of genes, this technique provides the first information on the function of a gene (Zweiger 1999). In order to study transcript levels, microarrays makes use of four different elements (Freeman, Roberston et al. 2000):

- 1) Target labeling
- 2) Target-probe hybridization
- 3) Detection
- 4) Data analysis

As seen in figure 14, the target sequence is labeled so that the amount of probe-target hybridization can be detected for measurement. Generally, fluorescent dyes are used to label the target RNA sequence; however, the RNA can also be converted to radioactively labeled cDNA (Freeman, Roberston et al. 2000). The conversion of RNA to cDNA occurs through reverse transcription. In order for this conversion to take place, the reverse transcriptase requires a primer to synthesize the cDNA and d(X)TP's. Since a mature mRNA sequence contains a poly A tail, the primer on oligo T is made to complement the poly A. As a result, the reverse transcriptase utilizes the mRNA as a template to produce the cDNA (Baldwin, Crane et al. 1999).

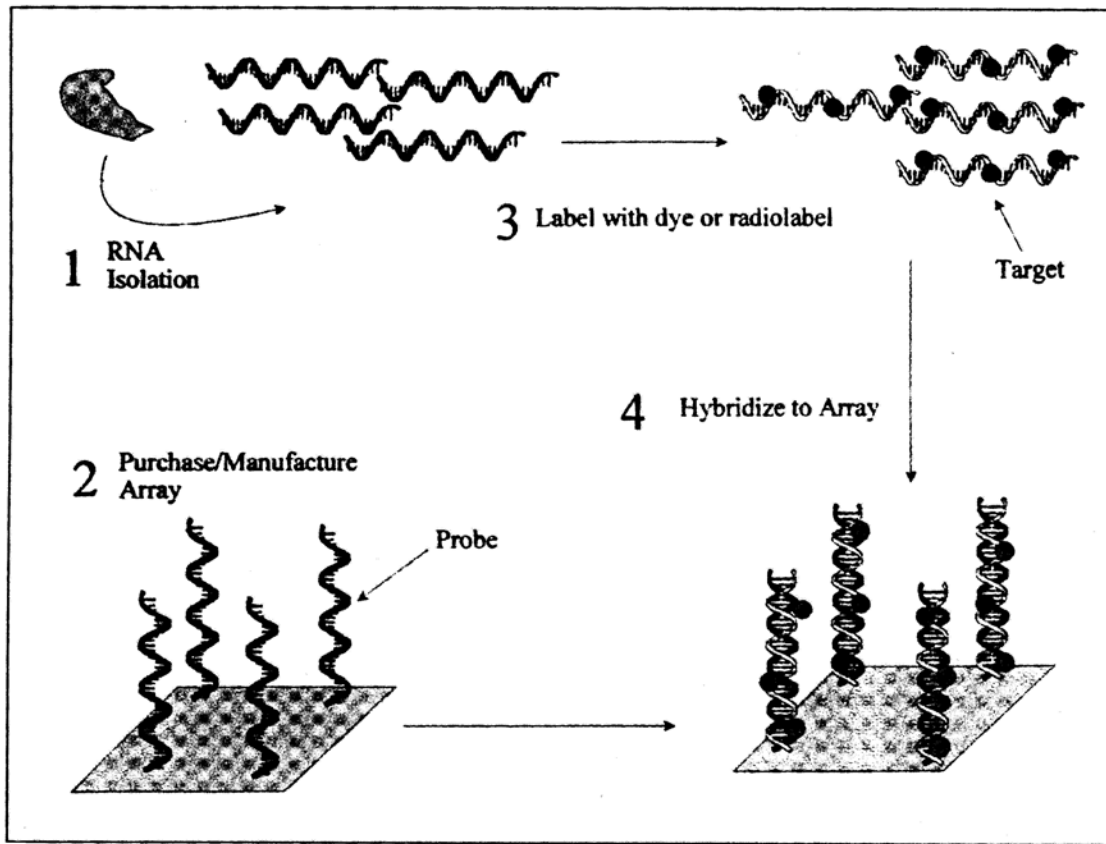


Figure 1.14. Experimental Scheme of Microarrays (Freeman, Roberston et al. 2000)

Transcript abundance data can be obtained through hybridization of the cDNA to the arrays followed by washing. In order to optimize the interactions between the target and probe sequences, there must be a uniform dispersion of the hybridization solution and

the wash solution. Since the intention of the hybridization solution is to allow the binding of complementary target and probe sequences, the hybridization solution needs to be evenly blended so that the labeled sequences can be spread equally over the probes. Consequently, the wash solution must be uniformly diffused over the array. The uniform distribution will minimize background and cut down on the nonspecific probe-target hybridization (Freeman, Roberston et al. 2000). However, this is a minor issue compared to others such as RNA quality, bias in cDNA synthesis, hybridization scanning, bleaching and stringency, and gene family effects.

In some microarray systems is the glass microscope slide is enclosed in a flow cell, for example Affymetrix. The slide is bound to the bottom of a sealed chamber, where it comes in contact with solutions that pass through the chamber. Under the slide, there is a metal plate that is used to control temperature. Consistent hybridization is obtained by optimizing the volume and temperature of the system (Freeman, Roberston et al. 2000). There are advantages of using a glass slide rather than a nylon array. The first is that glass is a resilient material that washes of high ionic strength and can withstand high temperatures. Also, the glass surface allows the DNA samples to be covalently attached to the treated surface. Next, by glass being non-porous, the kinetics of the annealing of the probes to the targets are enhanced by keeping the hybridization volume down to a minimum. Lastly, it does not contribute to background noise due to its low fluorescence (Cheung, Morely et al. 1999).

In order to analyze changes in transcript abundance, it is important to have sensitive and accurate signal output detection. The two methods of detection are radioisotopic and fluorescence (Freeman, Roberston et al. 2000). The downfall of a

radiolabelled probe is that it is unable to mix control and experimental probes for hybridization on the same probe (Duggan, Bittner et al. 1999). Fluorescence is the detection method that will be utilized for this research project. This method of detection allows the use of a double labeled hybridization scheme. Samples are to be compared labeled with different emission wavelength dyes, Cy3 and Cy5, mixed together and hybridized to the same array. The ratios of the Cy3 and Cy5 signals indicate the difference in relative abundance for the target-probe hybridization (Baldwin, Crane et al. 1999; Cheung, Morely et al. 1999; Freeman, Roberston et al. 2000). A scanning confocal microscope with a laser excitation source is used to detect the hybridization signal (Baldwin, Crane et al. 1999; Freeman, Roberston et al. 2000). The laser excites the dyes and the light generated and fluorescence is recorded as an optical image. The image is scanned integrated and converted to raw fluorescence intensity score. (Duggan, Bittner et al. 1999).

A major problem associated with microarrays is data analysis (Zweiger 1999). Microarray experiments generate large amounts of data. Background subtraction and normalization are required. For each spot, the background noise can be subtracted from the signal detected to provide a more accurate estimate of specific signals. Normalization is used to correct for the variation between individual arrays and other effects due to dye, and the bias due to interactions. These differences are created from unequal starting amounts of RNA or cDNA.

For those experiments that have two or more experimental conditions, clustering analysis is applied. This application is used to see correlated effect of different cDNAs over a time course or under different conditions. Clustering analysis groups genes

together that behave the similarly over the different experimental conditions. These genes are then studied to find a common function or regulatory control to explain the related behavior. Clustering analysis does not only apply to those genes of known sequence or known function (Zweiger 1999). Cluster analysis has also been utilized for metabolic profiling (Roessner, Leudemann et al. 2001a).

Extensive microarray analysis has been done on *Arabidopsis*. For example, analysis has been conducted on the development of *Arabidopsis* seeds (Girke, Todd et al. 2000). Microarray experiments have also applied to observe the gene expression levels of soybean (Maguire, Grimmond et al. 2002). Numerous micrarray experiments observe the expression of genes on loblolly pine (Sun, Whetten et al. 2000; Yang, No et al. 2002). Microarray analysis and metabolic profiling has also been conducted on loblolly pine cultures cells to study the lignification process (Stasolla, Scott et al. 2002).

1.12 References

- Amor, Y., C. Haigler, et al. (1995). "A membrane-associated form of sucrose synthase and its potential role in synthesis of cellulose and callose in plants." Proc. Natl. Acad. Sci. **92**: 9353-9357.
- Armor, Y., C. H. Haigler, et al. (1995). "A membrane-associated form of sucrose synthase and its potential pole in synthesis of cellulose and callose in plants." Proc. Natl. Acad. Sci. **92**: 9353-9357.
- Armstrong, M. D., K. N. F. Shaw, et al. (1956). J. Biol. Chem. **218**: 293.
- Babb, V. M. and C. Haigler (2001). "Sucrose Phosphate Synthase Activity Rises in Correlation with High-Rate Cellulose Synthesis in Three Heterotrophic Systems." Plant Physiology **127**: 1234-1242.
- Baldwin, D., V. Crane, et al. (1999). "A Comparison of Gel-based, Nylon Filter and Microarray Techniques to Detect Differential Expression in Plants." Current Opinion in Plant Biology **2**: 96-103.
- Behr, M. A., M. A. Wilson, et al. (1999). "Comparative genomics of BCG vaccines by whole-genome DNA microarray." Science **284**: 1520-1523.
- Boerjan, W., J. Ralph, et al. (2003). "Lignin Biosynthesis." Annu. Rev. Plant Biol **54**: 519-546.

- Bonetta, D. T., M. Facette, et al. (2002). "Genetic dissection of plant cell-wall biosynthesis." Biochemical Society Transactions **30**: 298-301.
- Boudet, A. M. (2000). "Lignins and lignification: Selected issues." Plant Physiology and Biochemistry **38**(1-2): 81-96.
- Cano-Delgado, A., S. Penfield, et al. (2003). "Reduced cellulose synthesis invokes lignification and defense responses in *Arabidopsis thaliana*." The Plant Journal **34**: 351-362.
- Carlson, S. J. and P. Chourey (1996). "Evidence for plasma membrane-associated forms of sucrose synthase in maize." Mol. Gen. Genet. **252**: 303-310.
- Chace, D. (2001). "Mass Spectrometry in Clinical Research." Chem. Rev. **101**: 445-477.
- Cheung, V., M. Morely, et al. (1999). "Making and Reading Microarrays." Nature Genetics Supplement **21**: 15-19.
- Cosgrove, D. J. (1997). "Assembly And Enlargement Of The Primary Cell Wall In Plants." Annu. Rev. Cell. Dev. Biol **13**: 171-201.
- Cremin, P., S. Kasim-Karakas, et al. (2001). "LC/ES-MS Detection of Hydroxy Cinnamates in Human Plasma and Urine." J. Agric. Food Chem. **49**: 1747-1750.
- Delmer, D. P. (1999). Cellulose biosynthesis in developing cotton fibers. In "Cotton Fibers". A. S. Basra. New York, Hayworth/Food Sci. Press: 85-102.
- Delmer, D. P. and C. H. Haigler (2002). "The Regulation of Metabolic Flux to Cellulose, a Major Sink for Carbon Plants." Metabolic Engineering **4**: 22-28.
- DeRisi, J. L., V. R. Iyer, et al. (1997). "Exploring the metabolic profiling and genetic control of gene expression on a genomic scale." Science **1997**: 680.
- Doblin, M. S., I. Kurek, et al. (2002). "Cellulose Biosynthesis in Plants: from Genes to Rosettes." Plant Cell Physiology **43**(12): 1407-1420.
- Duggan, D., M. Bittner, et al. (1999). "Expression Profiling using cDNA Microarrays." Nature Genetics Supplement **21**: 10-14.
- Evtuguin, D., P. Domingues, et al. (1999). "Electrospray Ionization Mass Spectrometry as a Tool for Lignins Molecular Weight and Structural Characterization." Holzforschung **53**: 525-528.
- Fiehn, O. (2000). Protocol for Plant Leaf Metabolite Profiling.
- Fiehn, O. (2001). "Combining genomics, metabolome analysis, and biochemical modeling to understand metabolic networks." Comp Funct Genom **2**: 155-168.
- Fiehn, O. (2002). "Metabolomics - the link between genotypes and phenotypes." Plant Mol Biol **48**: 155-171.
- Fiehn, O., J. Kopka, et al. (2000). "Metabolite profiling for plant functional genomics." Nature Biotechnology **18**(11): 1157-1161.
- Fiehn, O., J. Kopka, et al. (2000). "Identification of Uncommon Plant Metabolites Based on the Calculation of Elemental Compositions Using Gas Chromatography and Quadrupole Mass Spectrometry." Anal. Chemistry **72**: 3573-3580.
- Franke, R., J. M. Humphreys, et al. (2002). "The Arabidopsis REF8 gene encodes the 3-hydroxylase of phenylpropanoid metabolism." Plant J. **30**: 33-45.
- Fraser, P. D., M. Elisabete, et al. (2000). "Application of High-Performance Liquid Chromatography with Photodiode Array Detection to the Metabolic Profiling of Plant Isoprenoids." The Plant Journal **24**(4): 551-558.
- Freeman, W., D. Roberston, et al. (2000). "Fundamentals of DNA Hybridization Arrays for Gene Expression Analysis." Biotechniques **29**: 1042-1055.

- Freudenberg, K. and A. C. Neish (1968). Constitution and Biosynthesis of Lignin (Molecular Biology, Biochemistry, and Biophysics). New York, Springer-Verlag.
- Gelpi, E. (1992). "Trends in Biochemical and Biomedical Applications of Mass-Spectrometry." International Journal of Mass Spectrometry and Ion Processes **118**: 683-721.
- Gibeaut, D. M. and N. C. Carpita (1994). "Biosynthesis of plant wall polysaccharides." The FASEB Journal **8**: 904-915.
- Girke, T., J. Todd, et al. (2000). "Microarray Analysis of Developing Arabidopsis Seeds." Plant Physiology **124**: 1570-1581.
- Glassbrook, N., C. Beecher, et al. (2000). "Metabolic profiling on the right path." Nature Biotechnology **18**(11): 1142-1143.
- Haigler, C. H., M. Ivanova-Datcheva, et al. (2001). "Carbon partitioning to cellulose synthesis." Plant Mol. Biol. **47**: 29-51.
- Higuchi, T. (1997). Biochemistry and molecular biology of wood. Berlin ; New York, Springer.
- Higuchi, T. (1998). The Discovery of Lignin. Discoveries in Plant Biology. S.-D. Kung and S.-F. Yang.
- Higuchi, T. (2003). "Pathways for monolignol biosynthesis via metabolic grids: coniferyl aldehyde 5-hydroxylase, a possible key enzyme in angiosperm syringyl lignin biosynthesis." Proc. Japan Acad. **79**: 227-236.
- Holland, J. F., J. J. Leary, et al. (1986). "Advanced Instrumentation and Strategies For Metabolic Profiling." Journal of Chromatography **379**: 3-26.
- Humphreys, J., M. Hemm, et al. (1999). "New Routes for Lignin Biosynthesis Defined by Biochemical Characterization of Recombination Ferulate 5-hydroxylase a Multifunctional Cytochrome P450-dependent Monooxygenase." Proc. Natl. Acad. Sci. of United States of America **96**(1999): 10045-10050.
- Jeong, M. L., H. Jiang, et al. (2004). "Metabolic Profiling of the Sink-to-Source Transition in Developing Leaves of Quaking Aspen." Plant Physiology **136**: 3364-3375.
- Johansson, H., F. Sterky, et al. (2002). "Molecular cloning and characterization of a cDNA encoding poplar UDP-glucose dehydrogenase, a key gene of hemicellulose/pectin formation." Biochimica et Biophysica **1576**: 53-58.
- Kim, H., J. Ralph, et al. (2000). "Cross-Coupling of Hydroxycinnamyl Aldehydes in Lignins." Organic Letter **2**(15): 2197-2200.
- Li, L., X. Cheng, et al. (2001). "The Last Step of Syringyl Monolignol Biosynthesis in Angiosperms is Regulated by a Novel Gene Encoding Sinapyl Dehydrogenase." the Plant Cell **13**: 1567-1585.
- Li, L. G., J. L. Popko, et al. (2000). "5-Hydroxyconiferyl aldehyde modulates enzymatic methylation for syringyl monolignol formation, a new view of monolignol biosynthesis in angiosperms." Journal of Biological Chemistry **275**(9): 6537-6545.
- MacKay, J. J., D. M. Omalley, et al. (1997). "Inheritance, gene expression, and lignin characterization in a mutant pine deficient in cinnamyl alcohol dehydrogenase." Proceedings of the National Academy of Sciences of the United States of America **94**(15): 8255-8260.

- Maguire, T. L., S. Grimmond, et al. (2002). "Tissue-specific gene expression in soybean (*Glycine max*) detected by cDNA microarray analysis." Plant Physiology **159**(12): 1361-1374.
- Martin, L. K. (1999). *Cool-Temperature-Induced Changes in Metabolism Related to Cellulose Synthesis in Cotton Fibers*. Lubbock, Texas Tech.
- Matsui, N., K. Fukushima, et al. (1996). "On the behavior of monolignol glucosides in lignin biosynthesis .3. Synthesis of variously labeled coniferin and incorporation of the label into syringin in the shoot of *Magnolia kobus*." Holzforschung **50**(5): 408-412.
- Morris, C. R., J. T. Scott, et al. (2004). "Metabolic Profiling: a new tool in the study of wood formation." J. Agric. Food Chem. **52**(6): 1427-1434.
- Nair, R. B., Q. Xia, et al. (2002). "Arabidopsis CYP98A3 mediating aromatic 3-hydroxylation. Developmental regulation of the gene, and expression in yeast." Plant Physiology **130**: 210-220.
- Naki, T., N. Tonouchi, et al. (1999). "Enhancement of cellulose production by expression of sucrose synthase in *Acetobacter xylinum*." Proc. Natl. Acad. Sci. **96**: 14-18.
- Niwa, T. (1986). "Metabolic Profiling with Gas Chromatography-Mass Spectrometry and its Applications to Clinical Medicine." J Chromatography **379**: 313-345.
- Obst, J. R. (1998). *Special (Secondary) Metabolites from Wood*. Forest Products Biotechnology. A. Bruce and J. W. Palfreyman, Taylor and Francis.
- Osakabe, K., C. C. Tsao, et al. (1999). "Coniferyl aldehyde 5-hydroxylation and methylation direct syringyl lignin biosynthesis in angiosperms." Proceedings of the National Academy of Sciences of the United States of America **96**(16): 8955-8960.
- Plomion, C., G. Leprovost, et al. (2001). "Wood Formation in Trees." Plant Physiology **127**: 1513-1523.
- Ralph, J., J. MacKay, et al. (1997). "Abnormal Lignin in a Loblolly Pine Mutant." Science **277**: 235-239.
- Reiter, W.-D. (2002). "Biosynthesis and properties of the plant cell wall." Current Opinion in Plant Biology **5**: 536-542.
- Richmond, C. S., J. D. Glasner, et al. (1999). "Genome-wide expression profiling in *Escherichia coli* K-12." Nucleic Acid Res. **27**: 3821-3835.
- Roessner, U., A. Leudemann, et al. (2001a). "Metabolic Profiling allows comprehensive phenotyping of genetically or environmentally modified plant systems." Plant Cell **13**: 11-29.
- Roessner, U., Willmitzer, L., Fernie, A.R. (2001b). "High-resolution metabolic phenotyping of genetically and environmentally diverse potato tuber systems. Identification of phenocopies." Plant Physiology **127**: 1-15.
- Roessner-Tunali, U., B. Hegemann, et al. (2003). "Metabolic Profiling of Transgenic Tomato Plants Overexpressing Hexokinase Reveals That the Influence of Hexose Phosphorylation Diminishes during Fruit Development." Plant Physiology **133**(1): 84-99.
- Salnikov, V. V., M. J. Grisom, et al. (2001). "Sucrose synthase localizes to cellulose synthesis sites in tracheary elements." Phytochemistry **57**: 823-833.
- Sauter, H., M. Lauer, et al. (1991). "Metabolic Profiling of Plants - a New Diagnostic-Technique." Acs Symposium Series **443**: 288-299.

- Savidge, R. and H. Forster (2001). "Coniferyl Alcohol Metabolism in Conifers - II. Coniferyl Alcohol and Dihydroconiferyl Alcohol Biosynthesis." Phytochemistry **57**: 1095-1103.
- Schoch, G., S. Goepfert, et al. (2001). "CYP98A3 from *Arabidopsis thaliana* is a 3'-hydroxylase of phenolic esters, a missing link in the phenylpropanoid pathway." J. Biol. Chem **276**: 36566-36574.
- Seitz, B., C. Klos, et al. (2001). "Matrix polysaccharide precursors in Arabidopsis cell walls are synthesized by alternative pathways with organ-specific expression patterns." The Plant Journal **21**(6): 537-546.
- Skoog, D., J. Holler, et al. (1998). Principles of Instrumental Analysis. Philadelphia, Saunders College Publishing.
- Stanley, J., G. Newport, et al. (1978). "An Analytical Approach to Metabolic Profiling of Aromatic Compounds using Liquid Chromatography." Journal of Liquid Chromatography **1**: 305-325.
- Stasolla, C., J. Scott, et al. (2002). Large scale gene expression analysis during lignification of CAD (cinnamyl alcohol dehydrogenase) - mutant and wild type cultured cells of loblolly pine. International Plant & Animal Genome Conference, San Diego, Ca.
- Steeves, V., H. Forster, et al. (2001). "Coniferyl Alcohol Metabolism in conifers - I. Glucosidic Turnover of Cinnamyl Aldehydes by UDPG: Coniferyl Alcohol Glucosyltransferase from Pine Cambium." Phytochemistry **57**: 1085-1093.
- Sun, Y.-H., R. Whetten, et al. (2000). Stresses Make Wood Younger? A cDNA Microarray Analysis of Juvenile and Mature Wood Development in Loblolly Pine. Plant & Animal Genome VIII Conference, San Diego.
- Sweeley, C. C., W. H. Elliot, et al. (1966). Anal. Chem. **38**: 1549.
- Taylor, N. G., R. M. Howells, et al. (2003). "Interactions among three distinct Cesa proteins essential for cellulose synthesis." PNAS **100**(3): 1450-1455.
- Tenhaken, R. and O. Thulke (1996). "Cloning of an enzyme that synthesizes a key nucleotide-sugar precursor of hemicellulose biosynthesis from soybean: UDP-glucose dehydrogenase." Plant Physiology **112**(3): 1127-1134.
- Terashima, N., K. Fukushima, et al. (1993). Comprehensive Model of the Lignified Plant Cell Wall. Forage cell wall structure and digestibility. H. G. Jung. Madison, ASA, CSSA, SSSA: 247-270.
- Terashima, N., S. Ralph, et al. (1995). "New Facile Synthesis of Monolignol Glucosides:p-Glucocoumaryl Alcohol, Coniferin, and Syringin." Holzforschung **50**: 151-155.
- Trethway, R., Krotzy, A., Willmitzer, L. (1999). "Metabolic Profiling: a Rosetta Stone for Genomics?" Current Opinion in Plant Biology **2**: 83-85.
- Ulbrich, B. and M. H. Zenk (1980). "Partial purification and properties of para-hydroxycinnamoyl-CoA-shikimate-para-hydroxycinnamoyl transferase from higher plants." Phytochemistry **19**: 1625-1629.
- Weckwerth, W. (2003). "Metabolomics in Systems Biology." Annu. Rev. Plant Biol **54**: 669-689.
- Weckwerth, W., L. Willmitzer, et al. (2000). "Comparative quantification and identification of phosphoproteins using stable isotope labeling and liquid

- chromatography/mass spectrometry." Rapid Communications in Mass Spectrometry **14**(18): 1677-1681.
- Whetten, R., J. MacKay, et al. (1998). "Recent Advances in Understanding Lignin Biosynthesis." Annual Reviews Plant Physiology Plant Molecular Biology **49**: 585-609.
- Whetten, R. S., Y. Zhang, et al. (2001). "Functional Genomics and Cell Wall Biosynthesis in Loblolly Pine." Plant Mol. Biol. **0**: 1-17.
- Williams, R. J. (1951). Biochemical Institute Studies IV. Individual Metabolic Patterns and Human Disease: An Exploratory Study Utilizing Predominantly Paper Chromatographic Methods. Austin, University of Texas.
- Winter, H., J. L. Huber, et al. (1998). "Membrane association of sucrose synthase: Changes during graviresponse and possible control by protein phosphorylation." FEBS Lett. **420**: 151-155.
- Winter, H. and S. C. Huber (2000). "Regulation of sucrose metabolism in higher plants: Localization and regulation of activity of key enzymes." Crit. Rev. Plant Sci. **19**: 31-67.
- Yang, S.-H., E.-G. No, et al. (2002). Microarray analysis of Gene Expression in Different Loblolly Pine Tissue. Plant, Animal & Microbe Genomes X Conference, San Diego, CA.
- Ye, Z. H., R. E. Kneusel, et al. (1994). "An Alternative Methylation Pathway in Lignin Biosynthesis in Zinnia." Plant Cell **6**(10): 1427-1439.
- Zobel, B. and J. R. Sprague (1998). Juvenile Wood in Forest Trees. Berlin ; New York, Springer.
- Zweiger, G. (1999). "Knowledge Discovery in Gene-Expression-Microarray Data: Mining the Information." TIBTECH **17**: 429-436.

Metabolic Profiling – a New Tool in the Study of Wood Formation

published in J. Agric. Food Chem. 52, 1427-1434 (2004)

Cameron R. Morris¹, Jay T. Scott¹, Hou-min Chang¹, Ronald R. Sederoff²,
David O'Malley² and John F. Kadla^{1*}

¹ Wood Chemistry, ² Forest Biotechnology,
North Carolina State University, Raleigh, NC 27695

* Telephone number (919) 513-2455, Fax number (919) 515-6302

email address jfkadla@ncsu.edu

2.1 Abstract

In the realm of plant genomics, metabolic profiling has become a valuable tool with which to assess the effect of genetic and/or environmental factors on plant development. In this paper we report the first application of metabolic profiling on differentiating xylem tissue of Loblolly pine. We present a protocol for the analysis of loblolly pine xylem tissue. The effect of sample preparation, extraction and derivatization on the corresponding metabolite profiles and yields have been investigated and are reported. Using gas chromatography/mass spectroscopy (GC/MS), we have quantified more than 60 polar and lipophilic metabolites from wood forming tissue. It was possible to assign chemical structures to approximately half of these compounds. Comparison of six loblolly pine genotypes, three high cellulose (50-52%) and three medium (45-48%) cellulose, showed distinct metabolic profiles. Principal component analysis (PCA) enabled the assignment of metabolic phenotypes using these large data sets. Metabolic phenotype clustering occurred in which the three high cellulose genotypes were segregated from the medium cellulose genotypes. These results demonstrate the use of metabolic profiling for the study of wood forming tissue and as a tool in functional genomics.

Keywords: Metabolic profiling, loblolly pine, xylem tissue, principal component analysis (PCA)

2.2 INTRODUCTION

Metabolic profiling is a relatively new, but valuable technique being applied in the plant genomic community (Trethewey 1999). It has its roots in toxicology and blood screening (Holland, Leary et al. 1986; Toshimitsu 1986; Tayek, Bergner et al. 1990; Matsumoto and Kuhara 1996; Ning, Kuhara et al. 1996; Hong and Baldwin 1997; Rashed, Bucknall et al. 1997; Lim 1999; Garner and Leong 2000), where chemists have long used state-of-the-art chromatography to separate components, followed by mass spectrometry (MS) to identify and quantify them (Adams 1974; Holland, Leary et al. 1986; Toshimitsu 1986). The main goal of metabolic profiling at that time was to utilize the characteristic features of chromatographic profiles to study the differences between normal and disease states (Niwa 1986). To date, when coupled to genomics, the direct use of metabolic profiling offers a quick way to elucidate the function of novel genes. This circumvents the uncertainties due to regulation at the levels of transcription, translation and protein modification.

Metabolic profiles have been defined as multi-component analyses that produce patterns for a group of metabolically or analytically related metabolites. They are a two-dimensional cross section of a complex multidimensional physiological state delineated by the sample source, chemical processing, and method of analysis. Traditionally defined to be gas chromatography (GC), a profile may be generated by any one of a number of instrumental techniques such as gas chromatography/mass spectrometry (GC/MS), liquid chromatography/mass spectrometry (LC/MS), liquid chromatography/nuclear magnetic

resonance spectroscopy (LC/NMR), or NMR (Fraser, Elisabete et al. 2000; Glassbrook, Beecher et al. 2000; Le Gall, Colquhoun et al. 2003).

A distinction must be made between screening and profiling analyses. In screening, or multiple metabolite analysis, several compounds are measured, sometimes simultaneously, but in all situations the amount of these metabolites are compared to the reference ranges in which they are normally found. The concentration of each metabolite is to be considered as a single variable and used in the diagnosis of disease or the evaluation of the state-of-health based on the clinically determined relevance of that particular metabolite. In metabolic profiling, however, the absolute values of particular metabolites are not a major concern; but rather, the relative relationship of these metabolites to one another is an equally important consideration (Fiehn, Kopka et al. 2000). Another major difference between typical multi-component analysis and metabolic profiling is that in the profile, the various metabolites are usually metabolically or chemically related. The metabolic profile usually arises from a single analytical procedure, which imposes limitations on the chemical types of metabolites that can be determined in a given analysis. That is profiles may be obtained for steroids, organic acids, peptides etc., but not usually mixtures of these different kinds of metabolites. In addition, the profile, which utilizes information from all substances in the sample, those being known as well as unknown, makes it more able to serve in distinguishing metabolic differences as the results of analyses are studied.

An adoption of the technique of metabolic profiling of plant tissue using GC/MS was first introduced by Sauter, in which the profiling technique was used to study the effect of herbicide treatments on barely leaves (Sauter, Lauer et al. 1991). Recently,

Fiehn *et al.* developed a protocol to study *Arabidopsis thaliana* leaf tissue (Fiehn 2000). The profiling procedure involves a methanol and chloroform extraction of the plant tissue, followed by phase separation into polar and non-polar phases. The corresponding phases was methoxylated and trimethylsilylated for the analysis of total fatty acids, fatty alcohols, sterols, and aliphatics, as well as hydroxy-amino acids, amino acids, sugars, sugar alcohols, organic monophosphates, (poly)amines, and aromatic acids (Fiehn, Kopka et al. 2000). Using GC/MS analysis of the extracts, over 300 metabolites were quantified, with 15 of them being uncommon to plants (Fiehn, Kopka et al. 2000).

In this paper, we present a protocol for the analysis of loblolly pine xylem tissue. We discuss the effect of sample preparation, extraction and derivatization on the corresponding metabolite profiles and yields. Also, a data-mining tool, known as principal component analysis (PCA), is applied to the data to illustrate its effectiveness in interpreting the large amounts of data created by metabolic profiling. The overall goal of PCA is to reduce the dimensionality of the data set and ascertain new ‘meaningful underlying variables’ known as principal components.

2.3 MATERIALS AND METHODS

Materials. Materials and reagents were of analytical reagent grade (Aldrich Chemicals Co.) and used as received.

Methods. Samples were collected during the height of wood formation (between May and June). Xylem tissue was obtained from loblolly pine (*Pinus taeda L.*) by cutting away a small window of bark and phloem, typically 4” x 8”, followed by rapid scraping of the xylem tissue into a pre-chilled sterile centrifuge tube. The tissue was then immediately immersed in liquid nitrogen and transferred to a dry ice chest for transport

back to the lab. NOTE: once the tree bark has been scored, the trees defense mechanisms will be triggered and the metabolites present will begin to change, therefore, it is imperative that the tissue be collected as fast as possible, moreover, that consistent collection times be maintained during the course of repetitive sampling. Generally, the entire process takes about 30 seconds to collect enough tissue and place the samples in liquid nitrogen once the bark is removed. At the lab, the samples were placed in a -80°C freezer for storage.

300 mg \pm 30 mg of frozen xylem tissue was ground into a fine powder prior to extraction. Three grinding methods were used: a coffee bean grinder (Gloria Jean's Gourmet Coffee's/Model:202), amalgamator (Zenith/Model:Z14), and pestle and mortar. All three methods exhibited nearly identical results provided that the tissue samples remained frozen during the grinding procedure.

The ground tissue (300 mg \pm 30 mg) was mixed with 1.4-mL of 100% methanol and vortexed for 10 seconds to stop any enzymatic activity. Two internal standards were added, 100- μL ribitol solution (10 mg/mL H_2O) and 100- μL nonadecanoic acid methyl ester (10mg/mL CHCl_3). A 100- μL aliquot of water was added, the mixture was vortexed, and the tissue was extracted by shaking the mixture for 30 minutes at 70°C . The sample was then centrifuged for 3 minutes at 14000g. The supernatant was removed and diluted with 1.4-mL of H_2O in a second centrifuge tube, vortexed and set aside. The remaining centrifugate was further extracted for 10 minutes at 37°C with 2-mL of CHCl_3 . The mixture was centrifuged at 14000g for 3 minutes, and the supernatant transferred to the methanol-water extract. The combined extracts were vortexed and centrifuged at 4000 rpm for 15 minutes.

The polar (upper) phase and the lipid (lower) phase of the combined extract were separated into two 7-mL glass vials. Care was taken not to transfer any of the lipid phase over with the polar phase. Both the polar and lipid phase solutions were filtered into 7-mL vials using a 3-mL syringe and 0.2 μm syringe filter. The polar phase was dried in a speed vac concentrator overnight at 25°C.

The dried polar phase was dissolved in 100- μL of methoxyamine hydrochloride (20mg/mL pyridine) and shaken for 90 minutes at 30°C. 200- μL of N-Methyl-N-(trimethylsilyl) trifluoroacetamide (MSTFA) was then added, the solution was shaken for 30 minutes at 37°C, and then left standing for 2 hours at 25°C. The mixture was then diluted with 200- μL of pyridine, vortexed and directly injected (2- μL) onto the GC/MS.

The lipid phase was transmethylated prior to GC/MS analysis. Accordingly, a 700- μL aliquot of the lipid phase was placed into a 10-mL round bottom flask. To the flask was added 900- μL of CHCl_3 and 1-mL MeOH containing 3%v/v H_2SO_4 . The reaction was refluxed for 4 hours. The reaction was then extracted 2 times with water (4 mL), centrifuged at 4000 rpm to promote phase separation, and the water phase was removed. The remaining CHCl_3 phase was dried over anhydrous Na_2SO_4 and concentrated under reduced pressure to ~ 80 - μL . The acidic protons were derivatized with 10- μL of MSTFA in 10- μL pyridine for 30 minutes at 37°C. After the 30 minutes of derivatization the sample was injected (2- μL) onto the GC/MS.

The GC/MS system consisted of a ThermoFinnigan TraceGC and PolarisQ ion trap mass spectrometer. The system was controlled by Xcalibur software (version 1.3). GC/MS analyses were conducted using a 25 m x 0.2mm J&W Scientific INC. DB-1 column. The injection temperature was set to 230°C, the interface temperature was set to

200°C, and the ion source was set to 250°C. Helium flow was 1 ml min⁻¹. After a 5 minute solvent delay at 70°C, the oven temperature was increased at 5°C min⁻¹ to 310°C. The oven was then held isocratic for 1 minute and cooled down to 70°C. Mass spectra were recorded from m/z = 50 to 650 at 0.58 s scan⁻¹ with an electron ionization of 70 eV.

Lipid Metabolites

Propanoic acid methyl ester (**L1**) MW=234; MS m/z (rel. int.) 219 (M-15, 8), 191(36), 147(100), 117(22), 73(64)

Hexadecanoic acid methyl ester (**L7**) MW=270; MS m/z (rel. int.) 270 (M+, 30), 227(51), 199(41.85), 185(40), 171(50), 157(36), 143(79), 129(35), 101(46), 87(69), 74(50), 55(100)

Dibutyl Phthalate (**L8**) MW=278; MS m/z (rel. int.) 278 (M+, 1), 223(6), 205(7), 149(100), 121(4), 76(3)

Hexadecanoic acid trimethylsilyl (TMS) ether (**L11**) MW=328; MS m/z (rel. int.) 328 (M+, 9), 313(77), 201(13), 185(14), 143(10), 131(36), 129(50), 117(100), 75(80), 73(63)

5,8,11,14 Eicosatetraenoic acid, ethyl ester (**L12**) MW=332; MS m/z (rel. int.) 332 (M+, 1), 252(14), 184(21), 150(6), 119(11), 105(15), 91(23), 79(100), 67(44)

Trans 9,12-Octadecadienoic acid methyl ester (**L13**) MW=294; MS m/z (rel. int.) 294 (M+, 4), 262(14), 150(11), 135(20), 109(116), 95(29), 81(77), 67(100)

9-Octadecadienoic acid methyl ester (**L14**) MW=296; MS m/z (rel. int.) 296 (M+, 8), 264(29), 222(17), 180(17), 138(21), 123(20), 111(27), 96(48), 81(60), 67(68), 55(100)

Hexadecanoic acid, 2-[(trimethylsilyl)oxy]-methyl ester (**L15**) MW=358; MS m/z (rel. int.) 343 (M-15, 56), 299(100), 207(9), 159(18), 111(32), 97(61), 89(29), 73(81), 69(32), 57(31)

1-trimethylsilyloxooctadecane (**L16**) MW=370; MS m/z (rel. int.) 327 (M-43, 100), 227(3), 207(4), 139(5), 125(13), 111(31), 97(43), 83(24), 75(46), 57(27)

1-Octadecanoic acid TMS ester (**L19**) MW=356; MS m/z (rel. int.) 356 (M+, 15), 341(92), 313(16), 257(19), 201(26), 145(17), 129(68), 117(90), 75(100)

Methyl (Z)-5,11,14,17-eicosatetraenoate (**L20**) MW=318; MS m/z (rel. int.) 318 (M+, 1), 281(2), 249(4), 235(7), 203(10), 189(14), 149(15), 135(34), 121(36), 108(22), 91(42), 79(100), 67(53), 55(18)

Phenanthrene carboxylic acid methyl ester (**L22**) MW=314; MS m/z (rel. int.) 314(M+,5), 299(12), 239(100), 207(12), 197(10), 141(9)

7-Isopropyl-1,4a-dimethyl-1,2,3,4,4a,9,10,10a-octahydro-phenanthrene-1-carboxylic acid-trimethylsilyl ester(**L23**) MW=372; MS m/z (rel. int.) 372(M+, 5), 357(7), 250(9), 239(100), 173(6), 141(5), 73(5)

Hexanedioic acid, bis(2-ethylhexyl) ester (**L24**) MW=258; MS m/z (rel. int.) 259 (M+1, 6), 241(7), 207(14), 147(40), 129(100), 111(54), 83(24), 70(12), 55(37)

Phthalic acid bis-(6-methyl-heptyl) ester (**L28**) MW=390; MS m/z (rel. int.) 279 (M-111, 11), 207(10), 167(47), 149(100), 71(5), 55(7)

Polar Metabolites

2-trimethylsilanyloxy-succinic acid bis(trimethylsilyl) ester, (**P1**) MW=350; MS m/z (rel. int.) 351(M+1, 14), 263(19), 233(24), 189(17), 147(63), 133(12), 73(100)

N,N bis(trimethyl silanyl) amino butyric acid trimethylsilyl ester (**P2**) MW=319;
MS m/z (rel. int.) 304(M-15, 40), 246(5), 216(25), 174(100), 147(42), 131(10), 100(13),
73(62)

3,4,5-Tris-trimethylsilyloxy-cyclohex-1-enecarboxylic acid trimethylsilyl ester
(**P4**) (MW=462; MS m/z (rel. int.) 462(M+, 3), 372(7), 357(8), 282(17), 254(19),
204(100), 189(17), 147(25), 73(85)

Citronensaeure-tris-trimethylsilylester-trimethylsilylether (**P5**) MW=480; MS m/z
(rel. int.) 465(M-15, 4), 375(14), 363(13), 347(25), 305(7), 273(100), 257(34), 217(42),
147(82), 73(80)

1,3,4,5-Tetrakis-trimethylsilyloxy-cyclohexanecarboxylic acid
trimethylsilylester (**P11**) MW=552; MS m/z (rel. int.) 553(M+1, 13), 537(11), 419(8),
345(100), 334(6), 255(66), 204(9), 191(11), 147(25), 73(66)

1,3,4,5,6-Pentakis-trimethylsilyloxy-D-fructose-O-methyl-oxime (**P12**)
MW=570; MS m/z (rel. int.) 570 (M+1, 14), 480(14), 364(10), 307(20), 277(10),
217(100), 205(2), 147(10), 103(14), 73(50)

2,3,4,5,6-Pentakis-trimethylsilyloxy-D-glucose O-methyl-oxime (**P14**)
MW=569; MS m/z (rel. int.) 570 (M+1, 32), 480(27), 319(100), 229(12), 217(32),
205(11), 157(43), 147(21), 129(53), 73(80)

1-Methoxy-2,3,4,5,6-pentakis-trimethylsilyloxy-cyclohexane (**P16**) MW=554;
MS m/z (rel. int.) 555 (M+1, 2), 433(10), 318(27), 305(47), 260(72), 247(11), 217(43),
191(25), 159(11), 147(34), 133(17), 73(100)

2,3,4,5,6-Pentakis-trimethylsilyloxy-D-Glucopyranose (**P17**) MW=540; MS m/z (rel. int.) 539 (M-1, 2), 435(14), 361(18), 243(5), 217(26), 204(100), 191(48), 147(12), 129(8), 73(69)

1,2,3,4,5,6-Hexakis-trimethylsilyloxy-cyclohexane (**P18**) MW=612; MS m/z (rel. int.) 613 (M+1, 4), 507(8), 432(12), 318(56), 305(100), 265(33), 217(94), 191(45), 147(36), 129(20), 73(100)

(Tetrakis-O-trimethylsilyl- β -D-fructofuranosyl)-(tetrakis-O-trimethylsilyl- α -D-glucopyranoside) (**P19**) MW=918; MS m/z (rel. int.) 452 (M-466, 1), 437(19), 361(100), 271(13), 217(13), 169(24), 147(6), 129(6), 73(39)

Software. Principal Component Analysis (PCA) was carried out in SAS JMP 4.04.

2.4 RESULTS AND DISCUSSION

Variation in Xylem Tissue Grinding Methods and Grinding Times

The grinding of xylem tissue is an important step in the extraction of metabolites. Three methods of tissue grinding were used; pestle and mortar, amalgamator, and a coffee bean grinder with dry ice. The effect of grinding time (1, 2, 5 and 10 minutes) on metabolite yield was evaluated using the standard extraction and derivatization protocol.

For each of the three grinding methods, it was noted that one minute of grinding did not sufficiently pulverize the xylem tissue, with grinding times of two or more minutes being required to produce finely ground powders. However, the 1 and 2 minute grinding times gave very similar polar phase metabolite results. This was in stark contrast to the level of metabolites obtained from the lipid phase. Two minutes of grinding time produced twice as many metabolites as that obtained from grinding the xylem tissue for

only 1 minute. Grinding the xylem tissue for longer than 2 minutes showed no increase in the number of metabolite for both the polar and lipid phases. This is likely due to the fact that 2 minutes of grinding is sufficient to completely pulverize the tissue and enable the liberation of both the polar and lipid metabolites. Any increase in the duration of mechanical action may only serve to destroy sensitive and unstable metabolites, and lead to a decrease in metabolite yields. In fact, the tissue ground for 5 and 10 minutes showed a decrease in the number of metabolites seen in the lipid phase, despite refreezing the tissue samples every twenty seconds to ensure they remained frozen.

These results indicate (1) tissue grinding for 2 minutes provides the optimal yield in metabolites from xylem tissue, and (2) the method of tissue grinding does not greatly affect metabolite yield. It must be stressed that although it is important to accurately weigh the tissue prior to grinding, if not careful, pre-weighing the tissue for the amalgamator or pestle and mortar methods can lead to the tissue thawing out, allowing the continuous function of the enzymes within the xylem tissue. Therefore, the method employing the coffee grinder, which utilizes dry ice during the grinding process, is preferred. It not only enables the grinding of all of the tissue at one time, but also enables the temperature to be easily maintained at -80°C . This method permits the finely ground xylem to never increase in temperature and undergo any adverse changes. As needed, the xylem powder can be easily and quickly transferred into a frozen tarred vial, weighed for the extraction process and the remaining tissue placed back into the -80°C freezer.

Variation in Extraction Time

The second variable examined in the profiling protocol was the extraction time. As seen in Table 1, several extraction times for both the polar and lipid phases were

examined, and the effect of extraction time on the appearance/abundance of metabolites monitored. The polar phase extraction was analyzed by first holding the lipid phase extraction time constant at 5 minutes. Results from the polar phase extractions for the 5, 15 and 30 minute time periods, found in Figure 1, presented very similar chromatograms, which indicates that the same number of metabolites were released for these extraction time periods. Comparing the relative concentrations between metabolite levels in each chromatogram, Table 2 shows a difference between the three extraction times; an increase in extraction time has a direct relationship with the concentration of metabolites. However, a change in the profile of metabolites was observed when the extraction time was increased to 60 minutes. The concentration of metabolites decreased and new metabolites appeared, thus indicating metabolite decomposition or instability under these conditions. Thus, from these results a polar phase extraction time of 30 minutes appears to produce a maximum amount of polar metabolites in the shortest period of time.

The effect of extraction time for the lipid phase was analyzed by maintaining a constant polar phase extraction time of 30 minutes. The lipid phase extraction time was varied over the range of 5 to 60 minutes. Analogous to the polar phase, Figure 2 shows that the extraction time had a dramatic effect on the levels of metabolites detected by GC. A 5-minute lipid extraction was enough time to release metabolites; however, compared to the 10-minute lipid phase extraction there were fewer metabolites released at lower concentrations. As extraction time was increased to twenty minutes, little change in lipid metabolite levels occurred. Again at 60 minutes, a substantial decrease in metabolite levels accompanied by an increase in the total number of metabolites observed occurred,

analogous to that observed for the polar phase. Scheme 1 summarizes the optimal conditions for the isolation of polar and lipid metabolites.

Table 2.4.1. Extraction time variation. The initial polar phase extractions were conducted holding the lipid phase extraction constant. Once an optimum polar phase extraction time of thirty minutes was achieved, the lipid phase extraction was varied holding the polar phase extraction constant.

Extraction Phase	Extraction Time (min)							
	Polar	5	15	30	60	30	30	30
Lipid	5	5	5	5	10	15	30	60

Table 2.4.2. Relative concentrations of random metabolites for various Polar Phase Extraction Times

Polar Extraction Time (min)	$A_{27.95}/A_{1.S}$	$A_{29.04}/A_{1.S}$	$A_{31.07}/A_{1.S}$	$A_{42.73}/A_{1.S}$
5	1.96	4.57	2.49	4.52
15	1.97	4.70	2.76	4.15
30	3.33	8.02	4.03	7.46
60	-	-	-	-

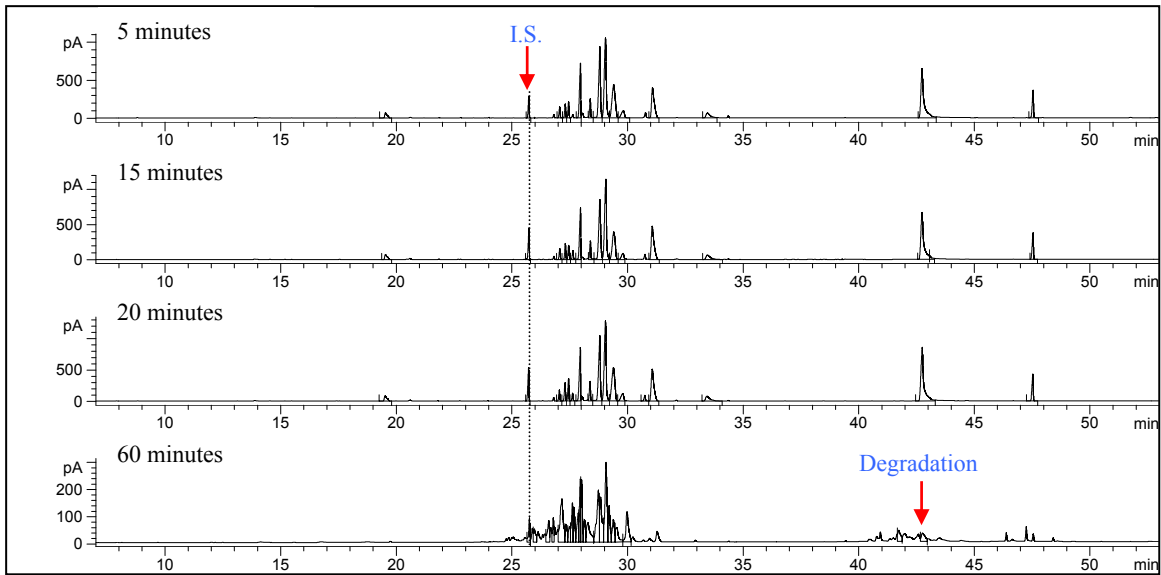


Figure 2.4.1. Chromatograms of Polar Phase Extraction Times

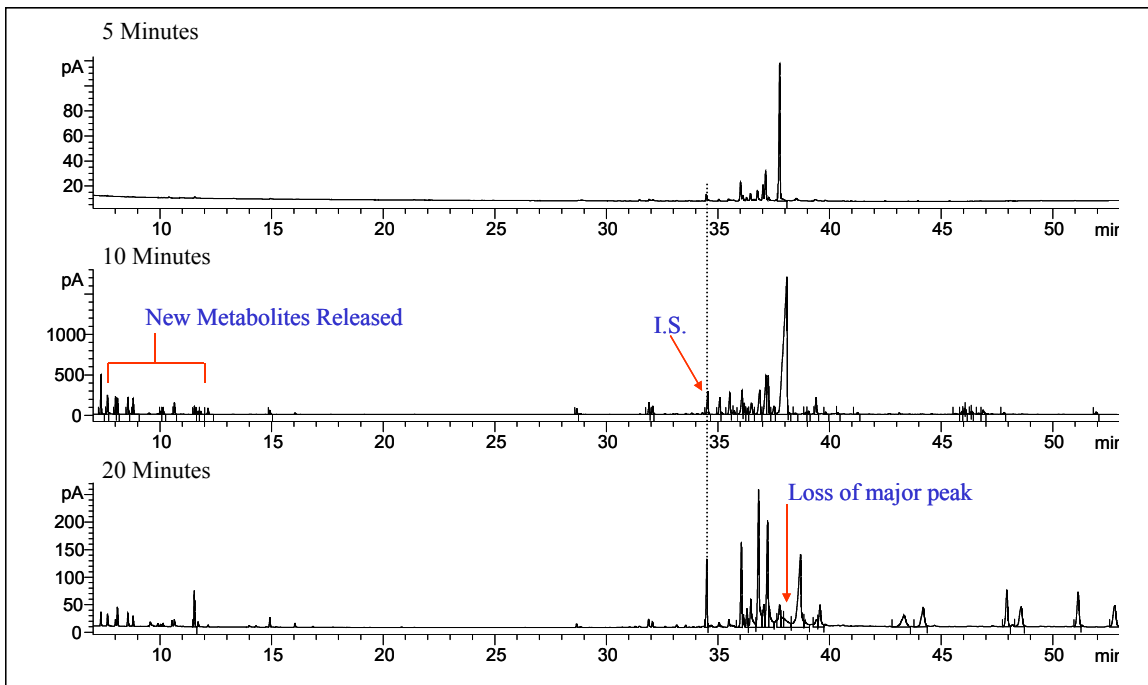
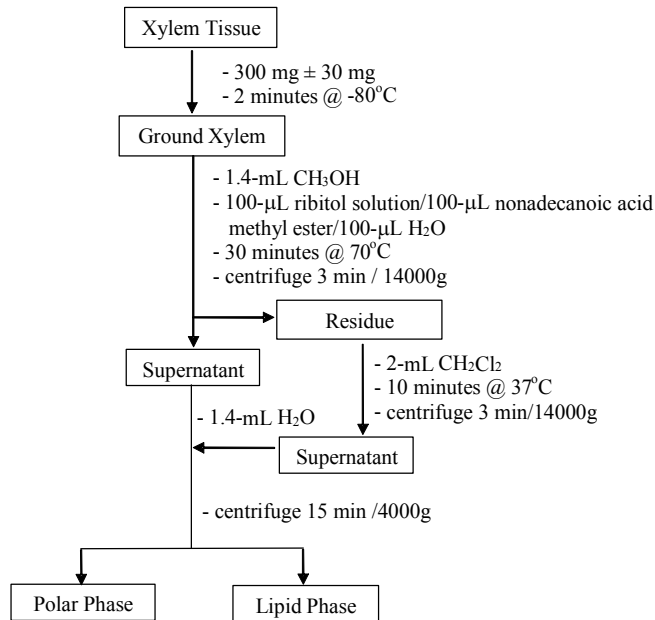


Figure 2.4.2. Chromatograms of Lipid Phase Extraction Times Figure Legend



Scheme 2.4.1. Metabolite isolation protocol

Variation in Derivatization Time

Once the appropriate extraction times were determined, it was decided to look at the length of derivatization time to ensure that the metabolites of the xylem tissue were being completely derivatized. In order to study the effects of derivatization time, four derivatization times were chosen and observed for any effects in the number or relative concentration of polar phase metabolites from the xylem tissue. Also, two model compounds (arabinose and caffeate), metabolites identified through mass spectra, were taken and run under identical derivatization conditions to determine the efficacy of derivatization.

Arabinose and caffeate were derivatized individually, as well as in a combined mixture to simulate the derivatization process. Results from the series of tests revealed that a derivatization time of greater than 2 hours was sufficient to achieve quantitative derivatization of the individual metabolites or the mixture of metabolites. Similar results were found, when performing the polar phase derivatization on xylem tissue, wherein a maximum in relative concentrations area was observed at between 2.5 and 3.5 hours of reaction time.

Metabolic Profiling of Loblolly Pine Xylem Tissue

Figures 3 and 4 show metabolite profiles for the lipid and polar phases obtained from xylem tissue of Loblolly pine. It can be seen that a large number of metabolites are observed, and although a majority of them have been identified, several remain unknown. Identification of these unknown metabolites is currently underway. However, the power of metabolic profiling is not in the identification of all metabolites present, but rather in only those responsible for an observed phenotypic change or difference. We are using

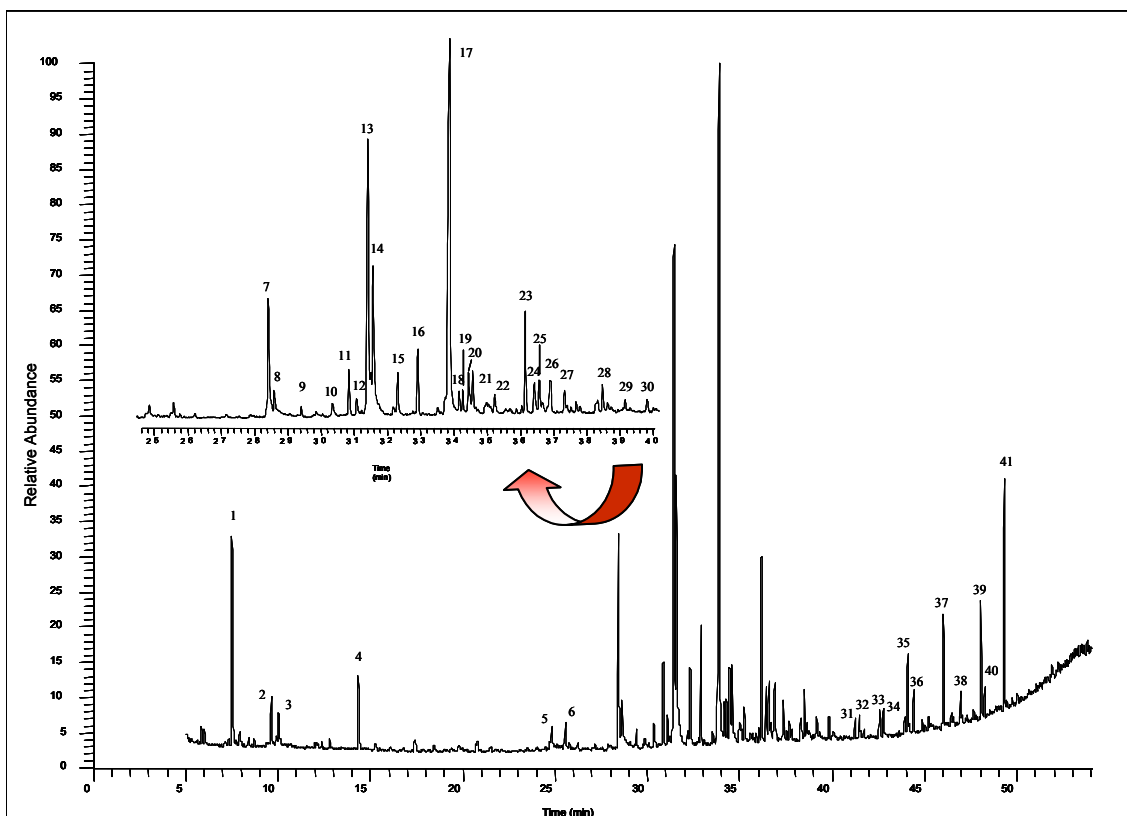
metabolic profiling to study genotypic differences in naturally occurring trees with observed differences in phenotype, e.g. cellulose content.

Due to the large amounts of data generated by metabolic profiles, it is very difficult to interpret the meaning of the enormous amount of chromatographic information. However, PCA is a useful tool that is used to categorize large amounts of data. To exhibit the utility of PCA with the metabolic profiling of xylem tissue, this statistical analysis was applied to xylem tissue has been collected from thirteen trees involved in a breeding program at NC State University that exhibit extreme phenotypic differences in cellulose content. Metabolic profiling in combination with PCA in this study is used to 1) separate the samples into metabolic phenotypic clusters as seen in figure 5 and 2) indicate the influence that the metabolites have on the clustering results, which is shown in figure 6.

Figure 5 illustrates the phenotypic cluster created by PCA from the SAS JMP 4.04 software. In this statistical representation, principal components 1, 2, and 4 are used to express 74% of the information compiled from the polar metabolites of thirteen individual trees that have been shown to have different levels of alpha cellulose content. The clusters presented are not as tight as one might expect; however, this could be due to the fact that the trees are grown on plots of land in a seed orchard and not a controlled environment, such as a green house. Due to this fact micro-environmental variations are introduced: i.e. changes in soil conditions across the plot, temperature gradients, and slopes on the land, which effects ground water flow, can contribute to variation in the presence of metabolites or metabolite concentration, which could be responsible for the larger phenotypic clusters.

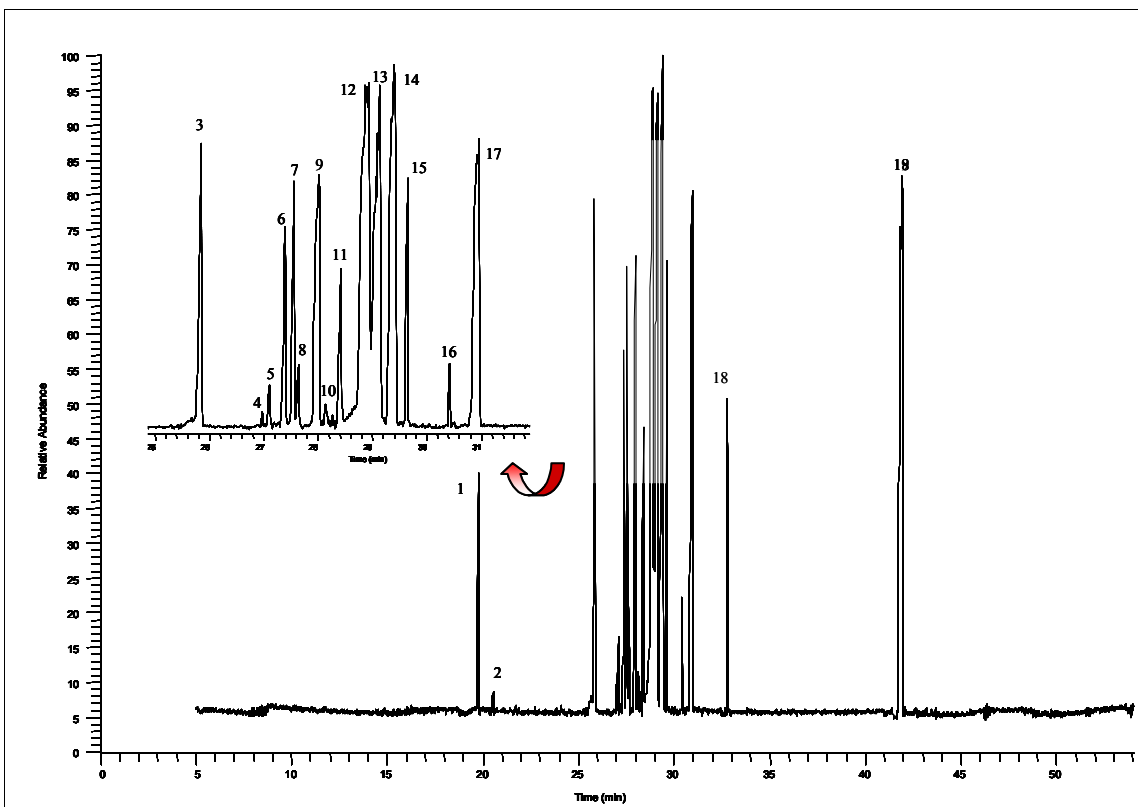
Plotting principal component 1 versus principal component 2 in Figure 6, provides the relationship between individual metabolites and clusters formed in Figure 5. This graph shows how each individual polar metabolite is utilized in the calculation of the principal component. Metabolites further away from zero have a greater influence on the calculation of the principal component and those closer to zero have less influence on the calculation. Therefore, the metabolites found further from the point of zero for principal component 1 are responsible for the phenotypic differences seen between the trees containing high (50-52%) and medium (45-48%) alpha cellulose content.

The developed protocol shows that a wide range of metabolites, both polar and lipophilic can be obtained from wood specific xylem tissue. Metabolites that play a major role in wood production have found, e.g. shikimic acid, which is a precursor to the lignin biosynthesis pathway, and citric acid, which is an intermediate of the citric acid cycle that generates ATP and also serves as a major source of building blocks for the biosynthesis of Amino Acids. When coupled to PCA, metabolic profiling can be used to display phenotypic differences as well as the metabolites responsible for those differences.



Peak Number	Compound Name	Retention Time	Peak Number	Compound Name	Retention Time
L1	Propanoic Acid	7.50	L22	Phenanthrene Carboxylic Acid	35.23
L2	Unknown	9.63	L23	Dehydroabetic Acid	36.16
L3	Unknown	10.03	L24	Hexanedioic Acid	36.42
L4	Glycerol	14.35	L25	Unknown	36.57
L5	Unknown	24.84	L26	Unknown	36.90
L6	1 OH-14:0	25.58	L27	Unknown	37.34
L7	Hexadecanoic Acid ME	28.42	L28	1,2-Benzenedicarboxylic Acid, diisooctyl ester	38.48
L8	Dibutyl Phthalate	28.59	L29	Unknown	39.16
L9	Unknown	29.41	L30	Unknown	39.82
L10	Unknown	30.35	L31	Unknown	41.22
L11	Hexadecanoic Acid TMS ester	30.85	L32	Unknown	41.49
L12	5,8,11,14 Eicosatetraenoic Acid, ethyl ether	31.08	L33	Unknown	42.54
L13	Trans 9,12-Octadecadienoic Acid ME	31.42	L34	Unknown	42.77
L14	9-Octadecadienoic Acid ME	31.57	L35	Unknown	44.07
L15	Hexadecanoic Acid, 2-[(trimethylsilyl)oxy]-,ME	32.31	L36	Unknown	44.42
L16	1 OH-18:0	32.92	L37	Unknown	46.02
L17	Internal Standard	33.89	L38	Unknown	46.95
L18	Unknown	34.17	L39	Unknown	48.06
L19	Octadecanoic Acid	34.27	L40	Unknown	48.24
L20	Methyl (Z)-5,11,17,17-eicosatetraenoate	34.44	L41	Unknown	49.30
L21	Unknown	34.57			

Figure 2.4.3. Lipid Phase Metabolite Profile



Peak Number	Compound Name	Retention Time	Peak Number	Compound Name	Retention Time
P1	Malic Acid	19.77	P11	Quinic Acid	28.41
P2	4-Amino Butyric Acid	20.56	P12	D-Fructose	28.92
P3	Internal Standard	25.66	P13	D-Fructose	29.14
P4	Shikimic Acid	26.98	P14	D-Glucose	29.38
P5	Citric Acid	27.11	P15	D-Glucose	29.65
P6	Unknown	27.40	P16	Ononitol	30.41
P7	Unknown	27.56	P17	β -D-Glucopyranose	30.95
P8	Unknown	27.64	P18	Inositol	32.79
P9	Unknown	28.00	P19	Sucrose	41.91
P10	Unknown	28.13			

Figure 2.4.4. Polar Phase Metabolite Profile

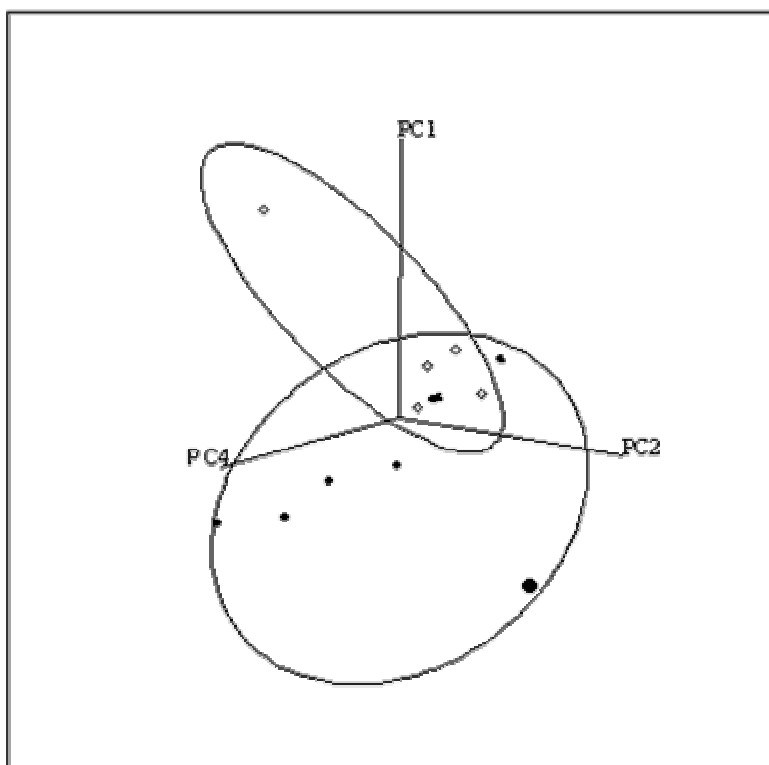


Figure 2.4.5. Metabolic phenotype clustering. Representation of clusters created after PCA of polar metabolite data taken from thirteen samples, which were split into two groups representing separate families. In order to provide, the best clusters between families, principal components 1, 2 and 4 were chosen. These principal components contain 74% of the 'total information content derived from metabolite variances'.

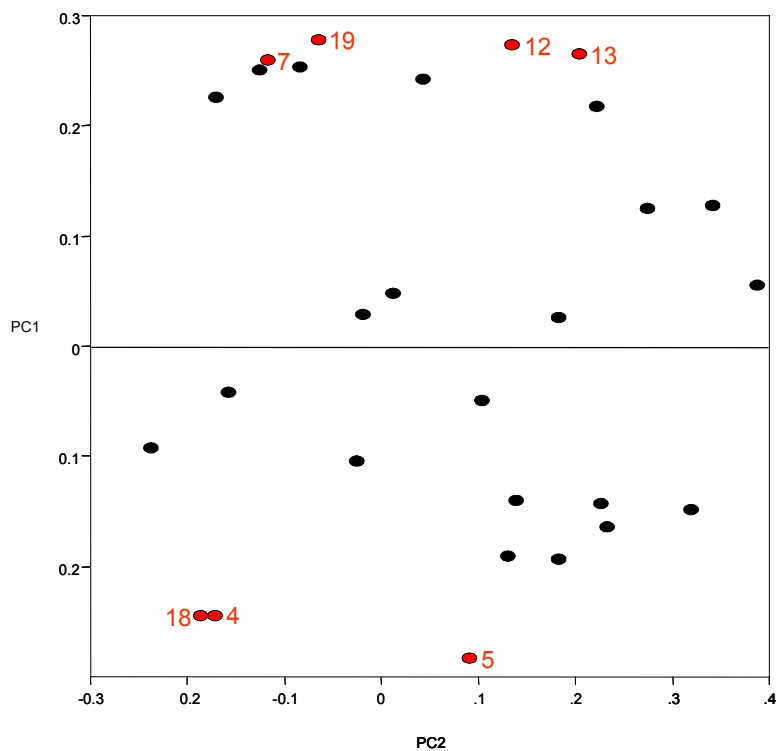


Figure 2.4.6. Principal Component 1 vs. Principal Component 2. This provides a graphical representation of the impact that individual polar metabolites have on the clustering results. The metabolites that are further away from zero have a greater impact on the linear combination that is used to calculate the principal component vector. Likewise, the metabolites closer to zero have less of an impact on the linear combination. The labeled metabolites, which are further from zero are responsible for the phenotypic differences seen between the two families of trees: 4 = shikimic acid; 5 = citric acid; 7 = unknown; 12 and 13 = D-Fructose; 18 = Ononitol; 19 = Glucopyranose

ACKNOWLEDGEMENTS

Financial support from USDA-IFAFS (Grant # 2001-52104-11224) is gratefully acknowledged.

2.5 REFERENCES

- Adams, R. F. (1974). "Determination of amino acid profiles in biological samples by gas chromatography." J Chromatography **95**: 189-212.
- Fiehn, O. (2000). Protocol for Plant Leaf Metabolite Profiling.
- Fiehn, O., J. Kopka, et al. (2000). "Metabolite profiling for plant functional genomics." Nature Biotechnology **18**(11): 1157-1161.
- Fiehn, O., J. Kopka, et al. (2000). "Identification of Uncommon Plant Metabolites Based on the Calculation of Elemental Compositions Using Gas Chromatography and Quadrupole Mass Spectrometry." Anal. Chemistry **72**: 3573-3580.
- Fraser, P. D., M. Elisabete, et al. (2000). "Application of High-Performance Liquid Chromatography with Photodiode Array Detection to the Metabolic Profiling of Plant Isoprenoids." The Plant Journal **24**(4): 551-558.
- Garner, R. C. and D. Leong (2000). "Pushing the accelerator - speeding up drug research with accelerator mass spectrometry." Nuclear Instruments & Methods in Physics Research Section B- Beam Interactions With Materials and Atoms **172**: 892-898.
- Glassbrook, N., C. Beecher, et al. (2000). "Metabolic profiling on the right path." Nature Biotechnology **18**(11): 1142-1143.
- Holland, J. F., J. J. Leary, et al. (1986). "Advanced Instrumentation and Strategies For Metabolic Profiling." Journal of Chromatography **379**: 3-26.
- Hong, J. F. and R. P. Baldwin (1997). "Profiling clinically important metabolites in human urine by capillary electrophoresis and electrochemical detection." Journal of Capillary Electrophoresis **4**(2): 65-71.
- Le Gall, G., I. J. Colquhoun, et al. (2003). "Metabolite Profiling of Tomato (*Lycopersicon esculentum*) Using ¹H NMR Spectroscopy as a Tool to Detect Potential Unintended Effects Following a Genetic Modification." J. Agric. Food Chem. **51**(9): 2447-2456.
- Lim, H. K., Stellingweif, S., Sisenwine, S., Chan, K.W. (1999). "Rapid drug metabolite profiling using fast liquid chromatography, automated multiple-stage mass spectrometry and receptor binding." J Chromatogr A **831**: 227-241.
- Matsumoto, I. and T. Kuhara (1996). "A new chemical diagnostic method for inborn errors of metabolism by mass spectrometry - Rapid, practical, and simultaneous urinary metabolites analysis." Mass Spectrometry Reviews **15**(1): 43-57.
- Ning, C., T. Kuhara, et al. (1996). "Gas chromatographic mass spectrometric metabolic profiling of patients with fatal infantile mitochondrial myopathy with de Toni-Fanconi-Debre syndrome." Acta Paediatrica Japonica **38**(6): 661-666.
- Niwa, T. (1986). "Metabolic Profiling with Gas Chromatography-Mass Spectrometry and its Applications to Clinical Medicine." J Chromatography **379**: 313-345.

- Rashed, M. S., M. P. Bucknall, et al. (1997). "Screening blood spots for inborn errors of metabolism by electrospray tandem mass spectrometry with a microplate batch process and a computer algorithm for automated flagging of abnormal profiles." Clinical Chemistry **43**(7): 1129-1141.
- Sauter, H., M. Lauer, et al. (1991). "Metabolic Profiling of Plants - a New Diagnostic-Technique." Acs Symposium Series **443**: 288-299.
- Tayek, J. A., E. A. Bergner, et al. (1990). "Correction of Glucose Carbon Recycling For the Determination of True Hepatic Glucose-Production (Hgp) Rates By 1-13c-Glucose." Clinical Research **38**(2): A570-A570.
- Toshimitsu, N. (1986). "Metabolic Profiling with Gas-Chromatography-Mass Spectroscopy and Its Applications to Clinical Medicine." Journal of Chromatography **379**: 313-345.
- Trethewey, R. N. (1999). Metabolic Genomics Abstract 6098. XVI International Botanical Congress.

Variation of α -Cellulose Content and Related Metabolites During Wood Formation in Loblolly Pine

Cameron R. Morris¹, Barry Goldfarb², Fikret Isik², Bailian Li², Chris Smith³, Hou-min Chang¹, Ronald R. Sederoff⁴, and John F. Kadla^{5*}

¹ Wood Chemistry, ²Forestry, ³ Bioinformatics, ⁴ Forest Biotechnology,
North Carolina State University, Raleigh, NC 27695

⁵Advanced Biomaterials Chemistry
University of British Columbia, Vancouver, B.C. V6T 1Z4

* Telephone number (604) 827-5254, Fax number (604) 822-9104

email address john.kadla@ubc.ca

3.1 Abstract

In this study, observations were conducted to study the variation in α -cellulose content through a tree breeding program. In the breeding program, fourteen families were planted in South Carolina. Four of these families were chosen for this study based on their α -cellulose content. Two of the families were classified as high α -cellulose while the other two contained low α -cellulose. Upon the completion of the metabolite analysis, metabolite concentrations were normalized for subsequent statistical analysis. Metabolite concentration along with α -cellulose content were utilized to construct a model of α -cellulose prediction to identify the metabolites that are associated with this variation in α -cellulose content. Here we show that the concentration levels of glucose, fructose, and sucrose play a major role in the metabolic biosynthesis of cellulose in loblolly pine breeding.

Keywords: Metabolic profiling, α -cellulose, loblolly pine, step-wise regression, xylem tissue, tree breeding

3.2 INTRODUCTION

The world demand for wood and fiber is predicted to increase considerably due to population growth and an improvement in the general standard of living around the world. However, the land available for future wood production will be restricted or even reduced by competing land uses for the economic viability of urbanization. As a result, research must focus its efforts on increasing the potential of trees to produce more and better wood through physiology, tree breeding and biotechnology (Sykes, Isik et al. 2003).

In order to meet the future demands of wood, tree will have to be harvested at a much earlier age, the wood will have a higher proportion of juvenile wood. Since juvenile wood could become the main source of wood and fiber for softwoods, the strategy of tree breeding and silviculture should be to improve the quality of juvenile wood to fit its targeted end products. Genetic variation in physiochemical wood traits of loblolly pine was investigated utilizing 14 full-sib families generated by a 6-parent half-diallel mating design (Sykes, Isik et al. 2003). Average fiber length, fiber coarseness, lignin content and α -cellulose content were measured on the earlywood and latewood of rings three and eight for each full-sib family. α -Cellulose content varied from 38 percent to about 44 percent between the families, which supports the suggestion that there is genetic variation for wood properties in juvenile wood (Zobel and Jett 1995; Zobel and Sprague 1998; Sykes, Isik et al. 2003). Therefore, the variation in α -cellulose content in breeding populations' needs to studied further.

Cellulose is a polymer of β -1, 4 linked anhydroglucopyranose monomers. It makes up 40-45% of the tree's dry weight (Delmer and Haigler 2002). Cellulose

production in plants is thought to occur in a three step process. The first step is the delivery of UDP-glucose to a cellulose synthase complex (Amor, Haigler et al. 1995; Babb and Haigler 2001; Doblin, Kurek et al. 2002). The second step of the process polymerizes glucose monomers into glucose rosettes (Doblin, Kurek et al. 2002). The final step of the process is the conversion of the glucose chain into a cellulose microfibril by KORRIGAN (KOR), which monitors cellulose microfibril production (Doblin, Kurek et al. 2002).

The main goal of this work is to identify the metabolic differences associated with the variation in α -cellulose content in loblolly pine populations. Also, this work will determine whether the metabolic differences are related to known metabolic cellulose biosynthesis pathways. In this study, 14 full-sib families were grown on a seed orchard as part of a tree breeding experiment.

Of the 14 families, four families were chosen based upon their differences in α -cellulose. From the four families, two expressed high levels of α -cellulose while the other two families expressed low levels. For the metabolic profiling experiments, differentiating xylem tissue was collected from individual trees of the four families. At the conclusion of the growing season, cores were collected from the trees to measure the α -cellulose content of tissue from the same year that tissue was collected for the metabolic profiling experiments. Results from the analysis of variance (ANOVA) and multi-variant analysis (MVA) show significant accumulations of fructose and sucrose play a major role in low α -cellulose individuals and glucose in the high α -cellulose individuals.

3.3 MATERIALS AND METHODS

Materials. Materials and reagents were of analytical reagent grade (Aldrich Chemicals Co.) and used as received.

Methods. Xylem tissue was collected, in May 2003, from 53 individuals planted on a seed orchard in Rock Hill, S.C. at an age of 13. The tissue was immediately frozen in liquid N₂ and stored on dry ice for transport back to the lab where it was stored in a -80°C freezer, until the performance of metabolic profiling experiments. The profiling method was conducted according to Morris et al. (Morris, Scott et al. 2004).

Software. Principal Component Analysis (PCA) was carried out in SAS JMP 4.04. Stepwise regression was completed using SAS Release 8.02. The following software was used to create a browser-based interface to a profiling database: Apache web server, PHP and the PostgreSQL database.

Profiling Database

Each of the 53 samples was exposed to the extraction method and run on the gas chromatography, GC, after derivatization. To store the gas chromatography data, and to help in its analysis, a browser-based interface to a database was created. The software used was the open-source Apache web server, PHP and the PostgreSQL database. Text output files containing the results of gas chromatography experiments can be imported into the database through a web browser. After import the raw data can be viewed. The internal standard peak is selected by the import process but can be changed. The sample data can be marked as unusable, if necessary, and is then excluded from further processing. Simple statistics and a graph of the internal standard peak times can be plotted to help in the detection of outliers, which may indicate a bad sample. A table of

peaks aligned across all samples can be displayed and exported to a text file suitable for reading into Microsoft Excel or other applications.

The algorithm used for aligning peaks normalizes the data by dividing by the standard peak time within each sample. It then orders all peaks across all samples by normalized peak time and looks for gaps in the list of times created. These gaps are assumed to delineate the peaks. Quite frequently we see that two values are revealed from the same sample before a gap is revealed. Since it is known that two values from the same sample represent different peaks this is also used to distinguish between aligned peaks. Finally, an iterative process is run over the candidate peaks to move any values that look out of place according to the average normalized time in each aligned peak.

The organized peak data was then exported into Excel for normalization. Normalization was used to correct for the environmental effects created by the trees location on different replications within the site. The environmental effects of replication could be due to directness of sunlight, or slopes in the ground, which affects water transport. Both of which could change the metabolite levels of a genetically identical tree planted on a different replication. As a result, each metabolite was normalized based on the trees replication using the following equation:

$$M_N = M - M_{\text{Replication Avg.}} \quad (1)$$

M_N = Normalized relative concentration of metabolite

M = Relative concentration of metabolite

$M_{\text{Replication Avg.}}$ = Average relative concentration of metabolite for all trees on a single replication.

Once the relative concentrations for each metabolite were normalized for each sample, the data was then imported into SAS JMP 4 for Principal Component Analysis (PCA). PCA is a statistical tool used to convert a number of possibly correlated variables into a smaller number of uncorrelated variables known as principal components. Therefore, PCA is used to reduce the dimensionality of the data set and identify the new meaningful underlying variables.

Mass Spectra of Significant Metabolites

4-Amino Hydroxybutyric Acid TMS MS m/z (rel. int) 304(40), 246(5), 216(25), 174(100), 147(42), 131(10), 100(13), 73(62)

Asparagine, N, N, O-TMS MS m/z (rel. int.) 258(4), 231(8), 215(5), 202(5), 188(7), 172(2), 159(5), 147(7), 141(6), 116(61), 100(11), 75(22), 74(13), 77(100)

Citric Acid TMS MS m/z (rel. int.) 465(4), 375(14), 363(13), 347(25), 305(7), 273(100), 257(34), 217(42), 147(82), 73(80)

Coniferin TMS MS m/z (rel. int.) 507(9), 450(20), 361(47), 324(98), 293(10), 271(22), 243(34), 217(42), 204(21), 169(33), 147(43), 129(19), 103(10), 73(100)

Fructose TMS MS m/z (rel. int.) 570 (14), 480(14), 364(10), 307(20), 277(10), 217(100), 205(2), 147(10), 103(14), 73(50)

Glucose TMS MS m/z (rel. int.) 570 (32), 480(27), 319(100), 229(12), 217(32), 205(11), 157(43), 147(21), 129(53), 73(80)

Inositol TMS MS m/z (rel. int.) 507(5), 432(16), 393(5), 367(9), 343(4), 318(67), 305(83), 284(11), 265(23), 217(86), 204(25), 191(41), 147(75), 129(16), 103(8), 73(100)

Malic Acid TMS MS m/z (rel. int.) 335(5), 319(4), 307(5), 265(4), 245(11), 233(23), 217(6), 189(9), 175(7), 147(80), 133(20), 117(9), 101(11), 73(100)

Ononitol TMS MS m/z (rel. int.) 496(5), 437(6), 349(17), 319(4), 259(57), 217(30), 204(61), 189(8), 157(9), 147(31), 129(18), 117(6), 103(28), 89(5), 73(100)

Pinitol TMS MS m/z (rel. int.) 627(30), 555(5), 507(14), 449(31), 433(42), 405(5), 375(36), 359(13), 343(41), 318(27), 305(21), 285(9), 267(30), 260(47), 247(16), 217(46), 207(33), 191(67), 179(10), 161(12), 159(21), 142(34), 134(31), 129(36), 103(19), 89(9), 73(100)

Quinic Acid TMS MS m/z (rel. int.) 553(13), 537(11), 419(8), 345(100), 334(6), 255(66), 204(9), 191(11), 147(25), 73(66)

Shikimic Acid TMS MS m/z (rel. int.) 462(3), 372(7), 357(8), 282(17), 254(19), 204(100), 189(17), 147(25), 73(85)

Unknown 7 (U7) MS m/z (rel. int.) 176.1(9), 175.1(18), 174.1(100), 133(11), 100(15), 86.0(16.95), 73.0(45), 59.1(22)

Unknown 8 (U8) MS m/z (rel. int.) 286(4), 285(32.71), 207(7), 185(13), 184(100), 174(21), 147(13), 134(60), 133(13), 130(11), 86(10), 79(46), 78(15), 77(64), 73(62), 52(58)

Unknown 10 (U10) MS m/z (rel. int.) 335(1), 333(10), 243(6), 218(11) 217(60), 204(43), 190(14), 149(7.67), 147(32.60), 133(15), 117(4), 75(13), 74(10), 73(100), 59(4)

Unknown 17 (U17) MS m/z (rel. int.) 451(2), 450(4), 334(3), 333(10), 294(1), 293(2), 292(10), 219(9), 218(21), 217(100), 149(8), 148(6), 147(38), 143(8), 129(10), 73(73), 59(2)

Unknown 20 (U20) MS m/z (rel. int.) 554(1), 493(3), 403(2), 402(10.83), 401(4), 330(4), 319(6), 312(20), 300(7), 229(2), 217(4), 205(23), 204(8), 189(32), 149(21), 148(15), 147(86), 130(20), 129(13), 117(10), 103(9), 89(10), 73(100) 59(8)

Unknown 26 (U26) MS m/z (rel. int.) 217(6), 215(2), 175(4), 173(3), 170(5), 150(5), 149(23), 147(100), 134(4), 131(15), 128(9), 103(6), 74(8), 73(35), 59(4)

Unknown 33 (U33) MS m/z (rel. int.) 478(2), 477(2), 452(5), 451(7), 450(17), 364(3), 363(13), 362(27), 361(80), 332(3), 331(12), 324(3), 295(20), 294(62), 293(5), 271(27), 218(12), 217(32), 205(10), 204(26), 203(7), 191(10), 169(32), 149(10), 147(44), 129(18), 103(11), 75(14), 73(100)

Unknown 34 (U34) MS m/z (rel. int.) 440(2), 438(11), 437(34), 401(4), 364(5), 363(15), 362(32), 361(100), 282(5), 281(13), 267(8), 217(15), 209(13), 208(13), 207(79), 204(12), 191(15), 169(14), 147(11), 133(9), 75.1(8), 73(23.32)

Xylose TMS MS m/z (rel. int.) 364(5), 333(4), 288(4), 263(23), 217(4), 205(43), 173(25), 147(61), 133(14), 117(27), 103(100), 73(100)

3.4 RESULTS AND DISCUSSION

Of the 14 full-sib families, 53 samples from four of the families were collected to conduct metabolic profiles experiments. The four full-sib families were chosen based on their α -cellulose production (Sykes, Isik et al. 2003). As illustrated in Figure 1, Families D and I were families that had higher levels of average α -cellulose, while families N and G measured lower levels of α -cellulose in growth rings from year 8.

To get a more accurate assessment of the metabolic changes in response to α -cellulose content, α -cellulose content measurements were taken for each of the 51 samples on the 13th year early wood ring. This is the same age of the tissue in which the metabolic profiling xylem tissue was taken for the study. This was accomplished by returning to the site to collect cores later in the year after the spring wood had formed.

The range of α -cellulose content varied from 39.8 to 49.8 percent on an extractive free basis.

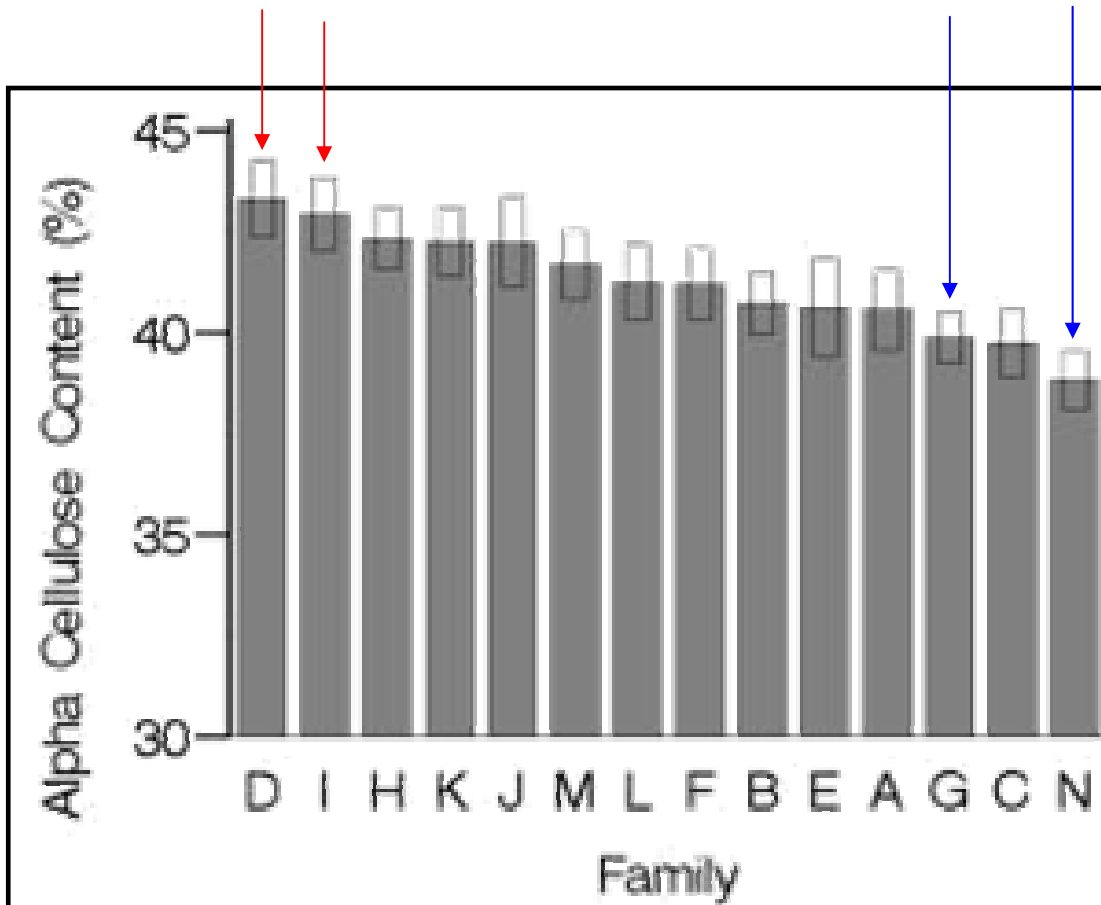


Figure 3.4.1. Cellulose Production of the 14 Full-sib Families (Sykes, Isik et al. 2003)

This study investigated only the trees that exhibited extremes levels of α -cellulose content. Trees that contained an α -cellulose content higher than 46.7 percent were defined as high in α -cellulose. Those samples that had an α -cellulose content less than 42.5 percent were classified as low α -cellulose. By separating the sample set into two extreme groups, the number of individual trees was decreased from 51 to 22.

Figure 2 displays the PCA outcome achieved from the metabolic profiling experiment for the 13th year α -cellulose content results. Principal components 1, 2, and 3 represent 57 percent of the total metabolic variation. There are distinct phenotypic clusters; however, these clusters have a mixture of high and low α -cellulose content trees. Other groups of principal components were plotted on a 3-dimensional graph to try and distinguish any kind of distinct clusters; nonetheless, none of the phenotypic clusters could separate out the high and low α -cellulose content trees. Although there are two distinct cluster produced by PCA, α -cellulose was not the phenotypic variable responsible for the cluster separation.

Since the PCA results were not decisive in separating out the high and low α -cellulose content samples, a stepwise regression model was used to pinpoint those metabolites that play a major role in cellulose production. The basic procedure of the stepwise regression involves the following steps:

1. creating a preliminary model
2. iteratively “stepping” -- altering the model at the preceding step by adding or removing a variable in accordance with the “stepping criteria”

3. terminating the search when stepping is no longer possible given the “stepping criteria” or when a specified maximum number of steps has been reached

The ANOVA table (Table 1) illustrates the results of the stepwise regression model. As shown in the table the p value of model is less than 0.001. Table 2 provides statistical information on the specific variables as well as identifies the metabolites associated with the model.

Figure 3 illustrates the α -cellulose prediction model and provides a graph of the predicted α -cellulose content versus measured α -cellulose content. Out of 57 polar phase metabolites, 14 were selected for the model based on the criteria set forth by the SAS statistical software, which utilizes a p value of 0.15 for introducing metabolites into the model. There are several unknown metabolites included in the model such as unknowns 7, 8, 14, 17, 20, 26, 33, and 34. Unknowns 7 and 8 are eluted at a retention time generally associated with amino acids. Unknowns 14, 17, and 20 are eluted in the region of carbohydrate retention times. Unknown 20 could possibly be galactose; however, it is difficult to distinguish between mannose which has an overlapping retention time and similar mass spectra. Lastly, unknowns 33 and 34 are eluted in the area of disaccharides or glucosides.

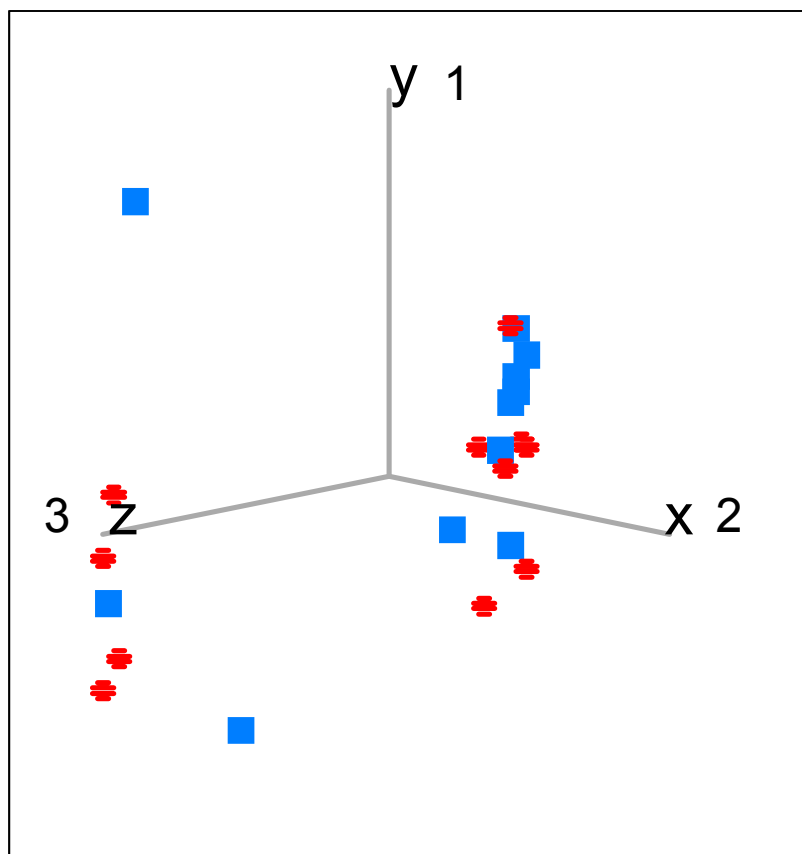


Figure 3.4.2. Metabolic phenotypic clustering results of extreme high and low α -cellulose for the 13th year. Principal components 1, 2, and 3 contain 57% of the total information substance resulting from the metabolite variances.

Table 3.4.1. ANOVA Table of Stepwise Regression Model

Analysis of Variance

Source Model	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	245.92626	245.92626	126.74	<.0001
Error	49	95.08055	1.94942		
Corected Total	50	341.00681			

Table 3.4.2. Stepwise Regression Model Information

Parameter Variable	Metabolite	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept		44.908084	0.2757	103187	39069.3	<.0001
F9	unknown 7	-328.9733	95.91228	31.07149	11.76	0.0015
F11	unknown 8	1126.87436	273.94898	44.68902	16.92	0.0002
F17	4-Amino Hydroxybutyric Acid	178.19233	72.19081	16.09175	6.09	0.0185
F21	L-Asparagine	-1947.00378	770.81031	16.85108	6.38	0.0161
F22	Xylose	696.25509	288.10863	15.42458	5.84	0.0209
F24	unknown 14	-649.98401	161.66939	42.69126	16.16	0.0003
F28	unknown 17	-1294.56437	403.55788	27.17846	10.29	0.0028
F29	Shikimic Acid	159.47089	28.80698	80.93874	30.65	<.0001
F35	Glucose	2.07201	0.61268	30.20673	11.44	0.0017
F37	unknown 20	-101.70388	33.58437	24.22088	9.17	0.0045
F46	unknown 26	-107.86637	44.13067	15.77903	5.97	0.0195
F48	Sucrose	-4.68727	0.99713	58.36137	22.1	<.0001
F54	unknown 33	338.99773	184.47743	8.91859	3.38	0.0744
F56	unknown 34	-98.99674	66.96858	5.7715	2.19	0.148

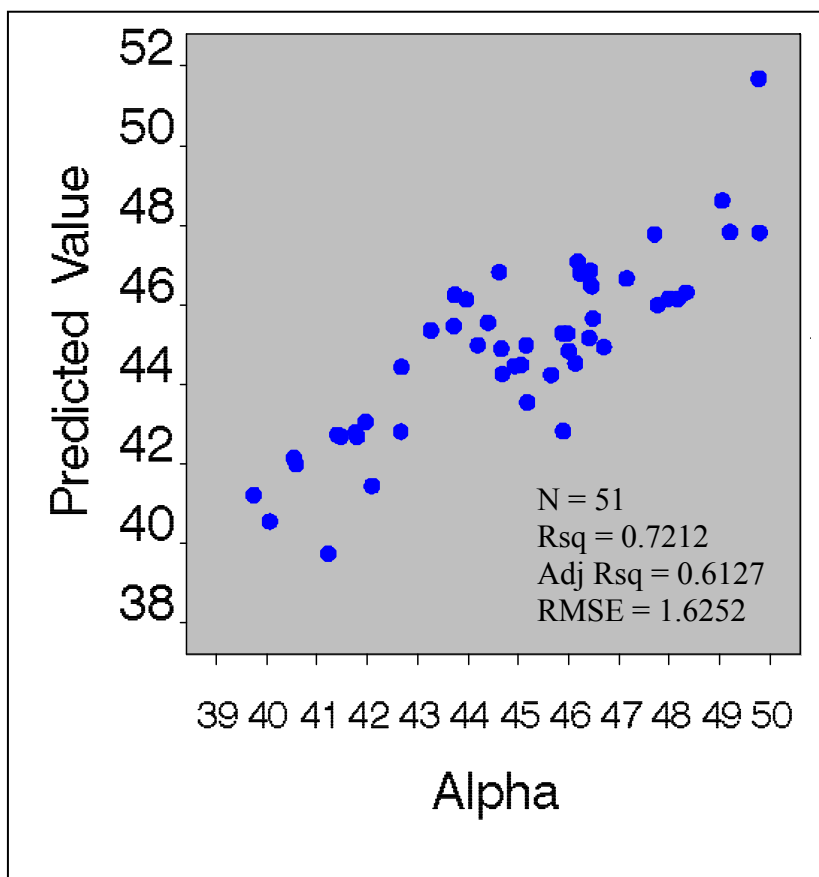


Figure 3.4.3. Stepwise Regression Model for α -cellulose Content Prediction (ACY = 44.981 - 328.97 F9 + 1126.9 F11 + 178.19 F17-1947 F21 + 696.26 F22 - 649.98 F24 - 1294.6 F28 + 159.47 F29 + 2.072 F35 - 101.7 F37 - 107.87 F46 - 4.6873 F48 + 339 F54 - 98.997 F56)

The known metabolites included in the model are 4-amino hydroxybutyric acid, L-asparagine, xylose, shikimic acid, glucose and sucrose. Based on the model 4-amino hydroxybutyric acid, xylose, shikimic acid and glucose have a positive correlation with α -cellulose content, indicating an increase in glucose concentration results in higher α -cellulose content. By contrast, L-asparagine and sucrose have a negative correlation with α -cellulose content indicating that the accumulation in sucrose concentration will decrease α -cellulose content.

Figure 4 provides the results of the analysis of variance for some of the known metabolites. The figure indicates whether the accumulation of individual metabolites occur in low or high α -cellulose individual trees. Included in the known metabolites are those that are directly associated with the metabolic biosynthesis of cellulose. For the high α -cellulose individuals, there is a large accumulation of glucose. Consequently, there are significant accumulations of sucrose and fructose in the low α -cellulose individuals.

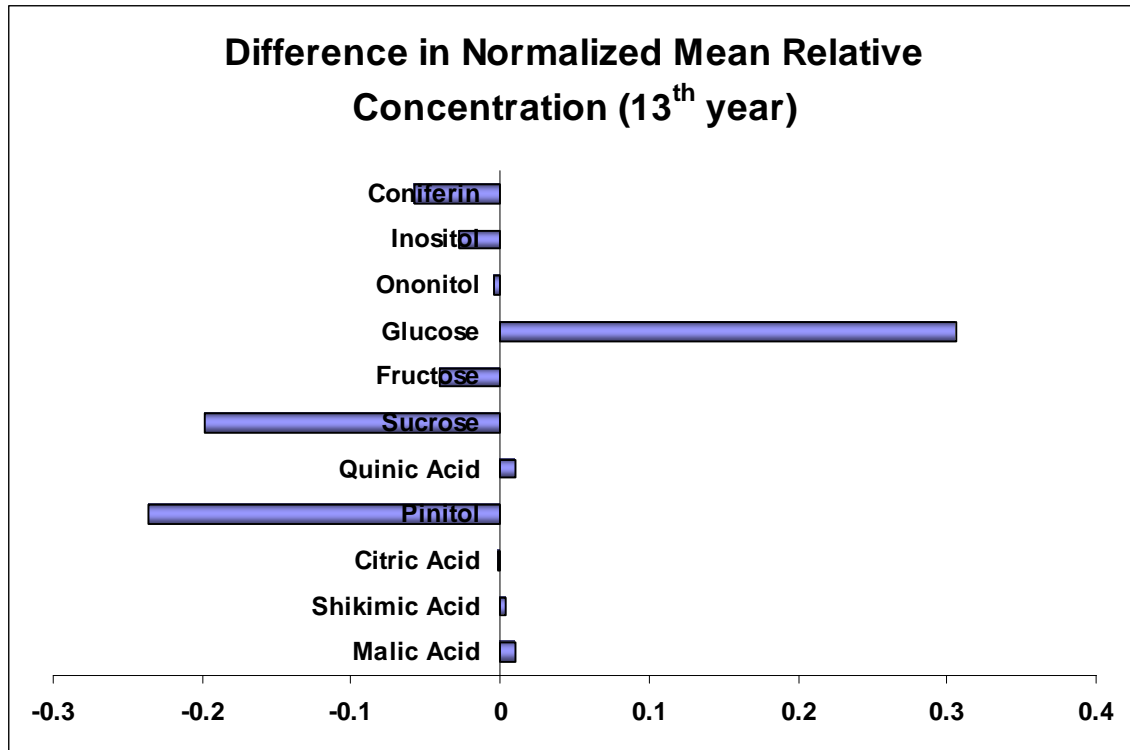


Figure 3.4.4. Graph of the differences between normalized mean relative concentrations for the 13th year. Metabolites that have positive integers are high in cellulose content. Metabolites that have negative integers are part of the low cellulose classification.

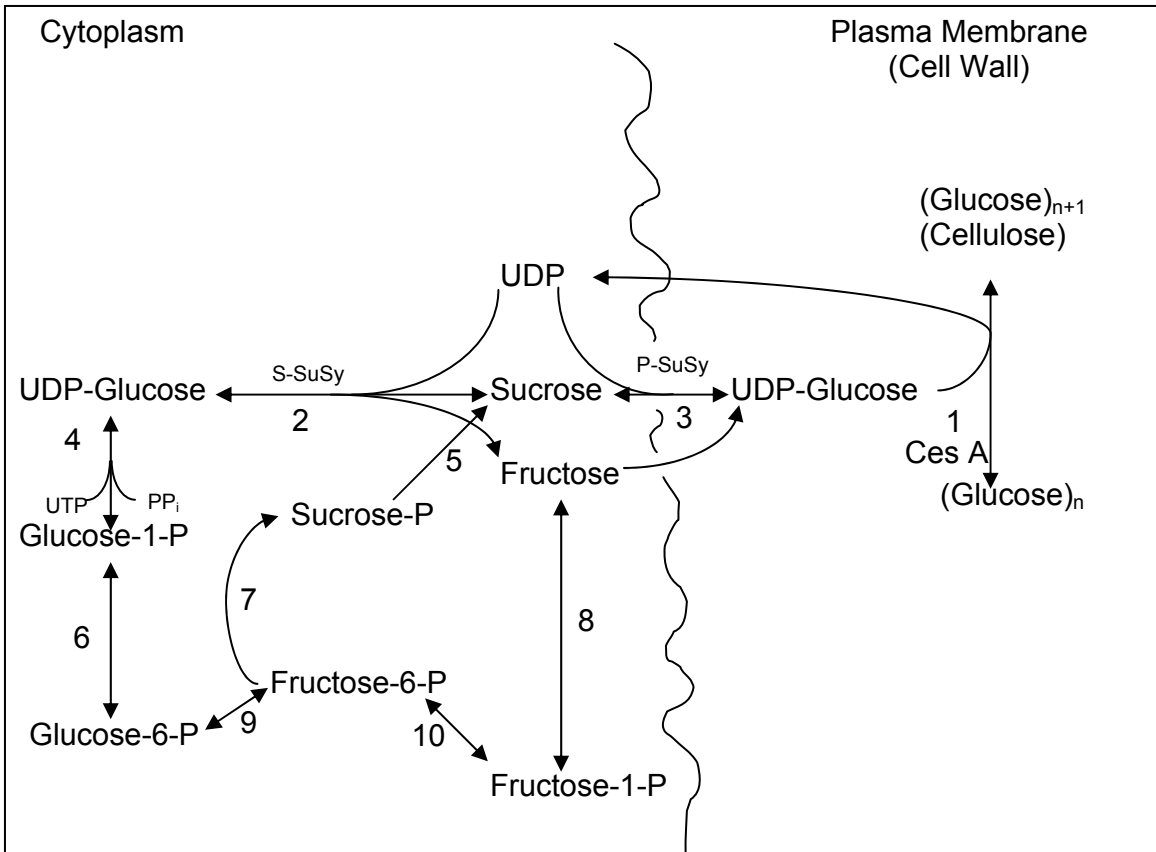


Figure 3.4.5. Cellulose Metabolic Biosynthetic Pathway (Delmer and Haigler 2002)

To gain a further understanding of the variation of α -cellulose content in breeding populations of loblolly pine, a greater understanding of the metabolic biosynthetic pathway for cellulose is necessary. Models for the biosynthetic pathway for cellulose suggest that the sources of UDP-glucose are complex, and involve several numerous possibilities for the recycling of intermediates (Delmer 1999; Haigler, Ivanova-Datcheva et al. 2001; Delmer and Haigler 2002). There are two reversible enzymatic reactions that may synthesize UDP-glucose. Figure 5 shows that reaction 4 is catalyzed by UDP-glucose pyrophosphorylase to convert glucose-1-phosphate to UDP glucose. The second proposed reaction of importance in the production of UDP-glucose is catalyzed by a form of sucrose synthase (SuSy) in reaction 2 or 3 to convert sucrose to UDP-glucose (Delmer and Haigler 2002).

However, there are two forms of SuSy in two different intracellular locations that could lead to two discrete pools of UDP-glucose (Delmer and Haigler 2002). Therefore, SuSy could interact with the inner bilayer of the plasma membrane and/or the cytoplasmic portion of the secondary wall cellulose synthase (Salnikov, Grisom et al. 2001). The two forms of SuSy are known as S-SuSy, which is the soluble form of SuSy (Salnikov, Grisom et al. 2001) and P-SuSy, the membrane-associated form of SuSy. P-SuSy is thought to be the enzyme responsible for the flow of carbon from sucrose to cellulose synthase via reaction 1 (Naki, Tonouchi et al. 1999; Delmer and Haigler 2002).

Fructose is the second product formed from the conversion of sucrose to UDP-glucose. In vascular plants, fructose can act as an inhibitor of SuSy (Winter, Huber et al. 1998; Haigler, Ivanova-Datcheva et al. 2001; Delmer and Haigler 2002). For the proposed pathway to operate efficiently there must be some mechanism to prevent the

accumulation of fructose near the site of cellulose synthesis. Some fructose produced by sucrose synthase (SuSy) can be involved in other biosynthetic pathways used for respiration or stored in the vacuole (Delmer and Haigler 2002). In cotton, the fructose created by the conversion of sucrose to UDP-glucose is recycled back to sucrose to be converted by P-SuSy again (Delmer 1999).

Sucrose, which is a key intermediate in the synthesis of cellulose, was found to accumulate in the low α -cellulose content trees as well as shown to have a negative correlation with α -cellulose content in the model. The pooling of sucrose coincides with the accumulation of fructose for the low α -cellulose content trees. Since an amassing of fructose inhibits the activity of SuSy, the conversion of sucrose to UDP-glucose will be repressed; thus causing pools of sucrose to accumulate in the cell. We suggest it is the lack of UDP-glucose that reduces cellulose biosynthesis.

3.5 CONCLUSION

Results from the metabolite analysis display accumulations in a number of metabolites between individuals in the high and low α -cellulose classifications which include fructose, glucose, inositol, and sucrose. Of the known metabolites directly associated with cellulose biosynthesis, glucose accumulates in the high α -cellulose category, while fructose and sucrose are increased in the low α -cellulose class. As in cotton, the accumulation of fructose may possibly inhibit P-SuSy, which converts sucrose to UDP-glucose. Therefore, not only is there an accumulation of fructose, sucrose could accumulate as a result of a possible inhibition of P-SuSy. Thus, possibly decreasing the amount of UDP-glucose produced which may result in lower amounts of cellulose being produced in the tree.

3.6 REFERENCES

- Amor, Y., C. Haigler, et al. (1995). "A membrane-associated form of sucrose synthase and its potential role in synthesis of cellulose and callose in plants." Proc. Natl. Acad. Sci. **92**: 9353-9357.
- Babb, V. M. and C. Haigler (2001). "Sucrose Phosphate Synthase Activity Rises in Correlation with High-Rate Cellulose Synthesis in Three Heterotrophic Systems." Plant Physiology **127**: 1234-1242.
- Delmer, D. P. (1999). Cellulose biosynthesis in developing cotton fibers. In "Cotton Fibers". A. S. Basra. New York, Hayworth/Food Sci. Press: 85-102.
- Delmer, D. P. and C. H. Haigler (2002). "The Regulation of Metabolic Flux to Cellulose, a Major Sink for Carbon Plants." Metabolic Engineering **4**: 22-28.
- Doblin, M. S., I. Kurek, et al. (2002). "Cellulose Biosynthesis in Plants: from Genes to Rosettes." Plant Cell Physiology **43**(12): 1407-1420.
- Haigler, C. H., M. Ivanova-Datcheva, et al. (2001). "Carbon partitioning to cellulose synthesis." Plant Mol. Biol. **47**: 29-51.
- Morris, C. R., J. T. Scott, et al. (2004). "Metabolic Profiling: a new tool in the study of wood formation." J. Agric. Food Chem. **52**(6): 1427-1434.
- Naki, T., N. Tonouchi, et al. (1999). "Enhancement of cellulose production by expression of sucrose synthase in *Acetobacter xylinum*." Proc. Natl. Acad. Sci. **96**: 14-18.
- Salnikov, V. V., M. J. Grisom, et al. (2001). "Sucrose synthase localizes to cellulose synthesis sites in tracheary elements." Phytochemistry **57**: 823-833.
- Sykes, R., F. Isik, et al. (2003). "Genetic Variation of Physiochemical Wood Properties in Loblolly Pine (*Pinus taeda* L.)." TAPPI **86**(12): 3-8.
- Winter, H., J. L. Huber, et al. (1998). "Membrane association of sucrose synthase: Changes during graviresponse and possible control by protein phosphorylation." FEBS Lett. **420**: 151-155.
- Zobel, B. and J. R. Sprague (1998). Juvenile Wood in Forest Trees. Berlin ; New York, Springer.
- Zobel, B. J. and J. B. Jett (1995). Genetics of Wood Production. Berlin, Springer-Verlag.

Metabolic Profiling and Microarray Analysis of a Loblolly Pine Growth Study

Cameron R. Morris¹, Deborah Craig⁵, Barry Goldfarb², Fikret Isik², Chris Smith³,
Hou-min Chang¹, Jason Osborne⁴, Ronald R. Sederoff⁵, and John F. Kadla^{6*}

¹ Wood Chemistry, ²Forestry, ³ Bioinformatics, ⁴Statistics, ⁵ Forest Biotechnology,
North Carolina State University, Raleigh, NC 27695

⁶Advanced Biomaterials Chemistry
University of British Columbia, Vancouver, B.C. V6T 1Z4

* Telephone number (604) 827-5254, Fax number (604) 822-9104

email address john.kadla@ubc.ca

4.1 Abstract

In this study, observations were made to investigate the molecular basis of differences in primary growth between two families. The heights of the trees were measured at 15 years old and were found to be significantly different between the families. Metabolic profiling was utilized to identify the metabolic differences between slow and fast growing families. Also, microarray experiments were conducted to identify the genes that have significant differences in transcript levels between the fast and slow growing families. L-glutamine was identified as a metabolite that has a positive correlation with tree height. From the microarray results, a transcript for a gene encoding a homolog for a NAM (no apical meristem)-like protein was found to increase in abundance in the fast growing family. Many other gene transcripts showed significant differences in abundance between the two families.

Keywords: Metabolic profiling, NAM (no apical meristem), loblolly pine, step-wise regression

4.2 INTRODUCTION

Over the course of a 100-year lifespan, the loblolly pine reaches an average height of 25-40 meters. Tree growth depends on specific centers known as meristems. Meristems are the location of repetitive cell division of unspecialized cells during the growing season, May to early September.

In trees, two types of meristems exist - apical and lateral. Apical meristems are located at the tips of each growing shoot and are responsible for vertical growth (elongation) and producing the cells that form new meristems; thus, giving the tree its primary growth. The lateral meristems are located in the cambium zone between the xylem and phloem. These meristems account for the secondary growth of the tree providing the tree with its bulk.

In this study, variability in tree growth will be investigated using metabolic profiling and microarray analysis. The families chosen were grown on an experimental site in Raleigh, N.C. Xylem tissue was collected from seven individual trees from each family for the profiling and microarray experiments. At the conclusion of the growing season, cores were taken from each tree to measure the lignin and α -cellulose content.

The main objective of this study is to observe the genetic and metabolic variability in tree height amongst families. Metabolic profiling will be used in an attempt to identify those metabolites whose concentration is correlated with tree growth. At the same time, microarray analysis has been done to obtain information on transcript level variation associated with tree growth or metabolite concentration. Correlations between tree height and lignin and α -cellulose content were also studied in this experiment.

Results show that the fast growing family has lower lignin content; however, there was no correlation between tree height and lignin content. The multi-variant analysis (MVA) of the metabolic profiling experiments indicates accumulations of L-glutamine in the fast growing family. Microarray analysis identifies many specific transcript levels are significantly different. One of these is a homolog for NAM (no apical meristem)-like protein, with an increased transcript level in the fast growing family.

4.3 MATERIALS AND METHODS

Materials. Materials and reagents were of analytical reagent grade (Aldrich Chemicals Co.) and used as received.

Methods. Xylem tissue was collected, in May 2003, from 7 individual trees from two families planted in Schenk Forest in Raleigh, N.C. for the microarray and metabolic profiling experiments. Cores were collected, in October, from each tree for subsequent measurements. Lignin and α -cellulose content were measured based upon the procedures constructed by Yeh et al. (Yeh, Chang et al. 2004). The profiling method was done as described in an earlier publication. (Morris, Scott et al. 2004).

Software. For metabolite analysis Principal Component Analysis (PCA) was carried out in SAS JMP 4.04. Stepwise linear regression was completed using SAS Release 8.02. The following software was used to create a browser-based interface to a profiling database: Apache web server, PHP and the PostgreSQL database.

Sample Collection

Seven loblolly pine families were grown on an experimental genetic site in Raleigh, North Carolina. At the end of fifteen years of growth, the height was taken for

each individual tree, and tree height average was taken for each family. Each family average was classified into ‘good’ and ‘poor’ growth categories for the experimental site. Of the ten families of trees studied, one family that exhibited fast growth, B, and one family that demonstrated slow growth, C, were chosen to investigate metabolite and gene expression levels related to height. The heights of both families in the study were found to be significantly different using a student t-test (p-value = 0.02). A total of 14 trees (seven for each family) were used for the study.

Xylem tissue was obtained by cutting away a small window of bark and phloem, typically 4” x 8”. Once the window was removed, the differentiating xylem tissue was rapidly scrapped into a pre-chilled centrifuged tube. The tissue was then immediately immersed in liquid nitrogen and transferred to a dry ice chest for transport to the lab to be stored at -80 °C. About 1g of xylem tissue from each tree was ground into powder using a pestle and mortar and liquid nitrogen. The RNA was extracted from the powder using a RNeasy plant mini kit (Qiagen, Valencia, CA). The integrity of the RNA was determined on a 1.0% agarose gel, while the purity and concentration was determined by UV absorbance at 260/280 nm.

cDNA selection

The 2,109 distinctive ESTs used for this study were acquired from cDNA libraries of *Pinus taeda* related to wood formation. The ESTs were selected closest to the 3’ end of the transcript of the respective contig, based on the website provided by University of Minnesota (<http://web.ahc.umn.edu/biodata/nsfpine>). Genes on the current array, are grouped into 17 different functional categories, as proposed for *Arabidopsis thaliana* (Kirst, Johnson et al. 2003; Stasolla, Zyl et al. 2003).

cDNA labeling and hybridization

The RNA isolated using the RNaseasy plant mini kit was amplified using a Message Amp aRNA kit by Ambion based on the RNA amplification developed by Dr. James Eberwine (Van-Gelder, Xastrow et al. 1990). The method utilizes reverse transcription with an oligo (dT) primer bearing a T7 promoter and *in vitro* transcription of the DNA product with T7 RNA polymerase to produce a large number of antisense RNA copies of each mRNA. The next step involves the aminoallyl labeling of the RNA (Hasseman 2004). The method reverse transcribes RNA using aminoallyl nucleotides to synthesize cDNA which can then be fluorescently labeled by coupling the aminoallyl groups to either Cyanine 3 or 5 (Hasseman 2004). Upon the completion of the coupling reaction, the product is purified using Qiagen PCR Purification Kit to remove the uncoupled dye. To begin the hybridization, 80 μ L of the labeled cDNA (target) in 50% deionized formamide was applied to the slides. A cover slip was placed on top of the slide. The slides were placed in Corning microarray slide chambers and incubated in a water bath overnight at 42 °C. The first post hybridization wash was 4 minutes at 42 °C in 1 X SSC, 0.2% SDS. The second wash was 4 minutes at 42 °C in 0.1 X SSC, 0.2% SDS and the last wash was 4 minutes at 42 °C three times in 0.1 x SSC.

Array Slide Preparation

For slide preparation, PCR amplification products of 2109 *Pinus taeda* cDNA clones were printed on aminosilane-coated glass microscope slides (Corning INC.; Corning, NY, USA) using a LUCIDEA printer (Amersham, NJ). After the slides were printed and dried, the spotted cDNAs were UV-crosslinked using a Stratalinker (Stratagene, LA Jolla, CA, USA), then baked for 2 hours at 75 °C and stored in the dark

at room temperature until use. Details of materials and methods for array preparation have been described (Stasolla, Zyl et al. 2003).

Experimental Design

A balanced (equal number of individuals) reciprocal loop design was used in this experiment to compare the 14 individuals in these families (Kerr and Churchill 2001). The balanced loop design allows for pair-wise comparisons, pairing the fast growth and slow growth individuals on the same array. This maximizes the power of the comparisons. The reciprocal portion of the design provides a technical replication and allows for correction for dye specific effects. There are 28 slides in total.

For microarray experiments, the statistical evaluation of differences in the transcript levels of genes is done by analysis of variance (ANOVA) (Kerr, Martin et al. 2000; Wolfinger, Gibson et al. 2001). This statistical method allows for flexibility in the experimental design, and generates probability values, which express the significance levels of the differences in transcript abundance. Two linear mixed models were performed in sequence, the normalization model and the gene model. The normalization model is a global normalization process that corrects for slide effects, dye effects and slide-dye interactions (Jin, Riley et al. 2001; Wolfinger, Gibson et al. 2001). The gene model was applied separately to each gene using residuals from the normalization model to infer the transcript level for each gene (Stasolla, Bozhkov et al. 2004).

Normalization Model

$$\text{Log}_2(Y_{ijkl}) = \mu_{ij} + D_k + S_l + DS_{kl} + e_{ijkl} \quad (1)$$

Y_{ijkl} = intensity of the s^{th} spot in the l^{th} slide with the k^{th} dye applying the j^{th} treatment for the i^{th} tree

μ_{ij} , D_k , S_l , DS_{kl} = the mean effect of the j^{th} treatment in the i^{th} tree, the k^{th} dye effect, the l^{th} slide random effect and the random interaction effect of the k^{th} dye in the l^{th} slide

e_{ijkl} = stochastic error term

Gene Model

$$R_{ijkl} = \mu_{ij} + T_k + S_l + SS_{ls} + e_{ijkl} \quad (2)$$

R_{ijkl} = residual of the g^{th} gene from the normalization model

μ_{ij} , T_k , and S_l = similar roles as μ_{ij} , D_k , S_l , and DS_{kl} except they are specific for the g^{th} gene

SS_{ls} = represents the spot by slide random effect for the g^{th} gene

e_{ijkl} = stochastic error term

Note. All random terms are assumed to be normally distributed and mutually independent within each model.

Mass Spectra of Significant Metabolites

4-Amino Hydroxybutyric Acid TMS MS m/z (rel. int) 304(40), 246(5), 216(25), 174(100), 147(42), 131(10), 100(13), 73(62)

Citric Acid TMS MS m/z (rel. int.) 465(4), 375(14), 363(13), 347(25), 305(7), 273(100), 257(34), 217(42), 147(82), 73(80)

Glutamine, N, N, N, O-TMS MS m/z (rel. int.) 419(5), 344(5), 329(4), 317(12), 301(4), 227(37), 216(9), 203(25), 188(5), 172(4), 156(12), 147(48), 128(15), 113(6), 100(9), 73(100)

Ononitol TMS MS m/z (rel. int.) 496(5), 437(6), 349(17), 319(4), 259(57), 217(30), 204(61), 189(8), 157(9), 147(31), 129(18), 117(6), 103(28), 89(5), 73(100)

Pinitol TMS MS m/z (rel. int.) 627(30), 555(5), 507(14), 449(31), 433(42), 405(5), 375(36), 359(13), 343(41), 318(27), 305(21), 285(9), 267(30), 260(47), 247(16), 217(46), 207(33), 191(67), 179(10), 161(12), 159(21), 142(34), 134(31), 129(36), 103(19), 89(9), 73(100)

Pyroglutamic Acid 2TMS MS m/z (rel. int.) 263(4), 258(9), 230(11), 214(4), 205(4), 156(100), 147(20), 140(4), 133(5), 117(73)

Unknown 11 (U11) MS m/z (rel. int.) 334(4), 333(7), 247(9), 246(21), 245(100), 217(13), 191(5), 189(16), 163(11), 149(42), 147(61), 117(13), 97(10), 83(26), 70(7), 69(87)

Unknown 12 (U12) MS m/z (rel. int.) 333(2), 307(8), 260(7), 229(19), 217(31), 205(15), 189(32), 185(5), 157(24), 150(10), 149(10), 148(11), 147(71), 129(30), 117(23), 103(9), 101(7), 73(100), 59(6)

Unknown 13 (U13) MS m/z (rel. int.) 391(2), 378(6), 349(6), 302(4), 277(9), 246(7), 205(5), 204(21), 202(10), 201(44), 188(10), 187(12), 186(57), 172(21), 149(46), 148(9), 147(63), 117(18), 89(9), 73(100)

Unknown 20 (U20) MS m/z (rel. int.) 433(4), 360(2), 359(6), 334(6), 333(21), 305(12), 292(15), 277(4), 221(7), 217(33), 191(8), 189(13), 157(10), 149(33), 147(65), 129(23), 117(7), 73(100)

Unknown 23 (U23) MS m/z (rel. int.) 375(2), 374(8), 359(14), 269(3), 256(5), 242(22), 241(100), 213(8), 199(9), 185(18), 159(12), 131(14), 109(17), 91(13), 67(6)

Unknown 25 (U25) MS m/z (rel. int.) 372(2), 357(4), 250(10), 240(21), 239(100), 197(8), 173(8), 141(10), 117(7)

Unknown 27 (U27) MS m/z (rel. int.) 279(4), 217(2), 168(4), 167(36), 150(11), 149(100)

Unknown 28 (U28) MS m/z (rel. int.) 480(3), 439(1), 438(5), 437(11), 390(5), 363(8), 362(14), 361(37), 300(8), 271(9), 257(9), 217(100), 218(22), 219(8), 169(28), 149(23), 147(26), 129(15), 89(7), 73(88)

4.4 RESULTS AND DISCUSSION

Prior to the investigation of metabolite and gene expression levels, measurements of lignin content and α -cellulose content were measured for each tree. Lignin and α -cellulose content was measured by a non-destructive method through Near-Infrared (NIR) Spectroscopy (Yeh, Chang et al. 2004), to determine if any correlation could be made with tree heights. As shown in Figures 1 and 2, for these samples there is no correlation between tree height and lignin content or tree height and α -cellulose content. However, in taking the averages of lignin content of the fast growing trees was 27.9 percent while the average lignin content for the slow growing trees was 28.9 percent. Therefore, there was a 1 percent difference in lignin content between the fast and slow growing families which could account for the differences seen in height. It has been argued that a decrease in lignin content is associated with better tree growth (Kirst, Myburg et al. 2004). Despite having lower lignin content, the fast growing family on average had lower α -cellulose content than the slow growing family. On average, the fast

growing family had an α -cellulose content of 48.8 percent but the slow growing family had an α -cellulose content of 49.8 percent.

Metabolic Profiling Results

Figure 3 illustrates a principal component analysis of the metabolic data. The first three principal components contain 65.36% of the total information content derived from the metabolic variances. The Hierarchical Cluster Analysis (HCA) in Figure 4 shows that there is an overlap between the fast growth and slow growth families. Height did not separate these samples into two distinct clusters. This result was expected because the two families do not vary that much genetically due to the continuous variation of individual trees. As a result, the phenotypic cluster results of the metabolic analysis might not vary significantly.

In an analysis of variance of the metabolic data related to tree height, three metabolites were found to be significantly different (Figure 5). All three were higher in concentration in the fast growing family. These significantly different metabolites were unknowns 13 (possibly an aldonic acid), 28 (possibly a disaccharide) and pyroglutamic acid. Pyroglutamic acid is an intermediate for three isoforms of glutamic acid. It is also a component in the pathway of glutathione metabolism.

As a result of the lack of clustering in the principal component analysis, a stepwise linear regression model was applied to the metabolic data to predict tree height. The model was utilized to identify metabolites that play a major role in the prediction of tree height. The basic procedure of the stepwise regression model involves the following steps:

1. constructing a preliminary model

2. iteratively “stepping” -- altering the model at the preceding step by adding or removing metabolic variables in accordance with the “stepping criteria”
3. terminating the search when stepping is no longer possible given the “stepping criteria” or when a specified maximum number of steps has been reached

Table 1 illustrates the ANOVA table of the stepwise regression model results.

Table 2 illustrates the tree height prediction model resulting from the stepwise regression. The tree height prediction model is comprised of 12 of the 55 individual metabolites, all strongly significant in the tree height prediction model (Table 2). Six of the 12 metabolites contain positive parameter estimates, while the other six metabolites contain negative parameter estimates. Of the 12 strongly significant metabolites, only five are known metabolites, which include: 4-amino hydroxybutyric acid, citric acid, L-Glutamine, pinitol, and ononitol. As the concentration of unknowns 11, 20 (possibly an aldonic acid, gluconic acid or galactonic acid), 27, and 28 (possibly a disaccharide), 4-amino hydroxybutyric acid, and pinitol increases the predicted tree height decreases. Likewise, increases in the concentration of unknowns 12, 23, and 25, citric acid, L-glutamine and ononitol increase the tree height prediction.

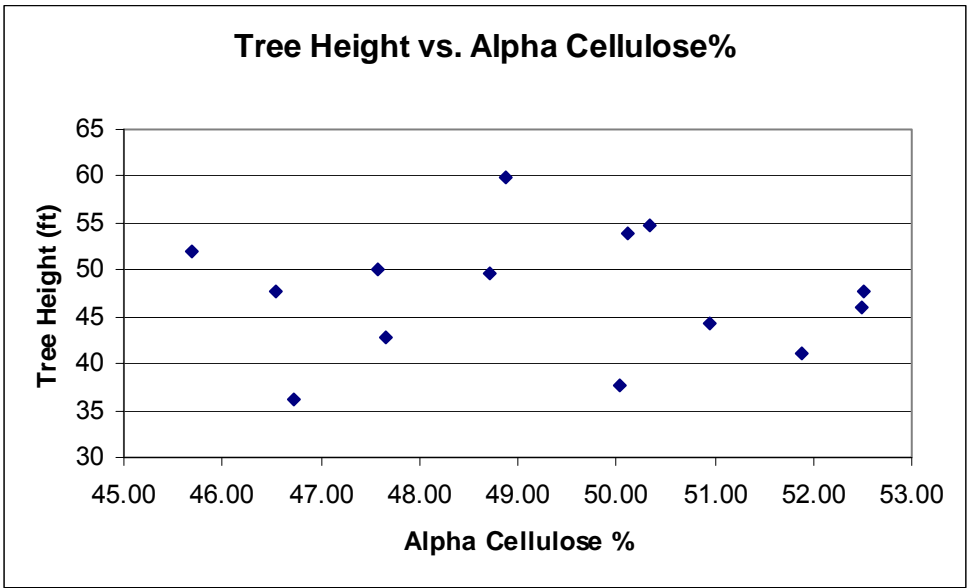


Figure 4.4.1. Graph of Tree Height versus α -Cellulose Content

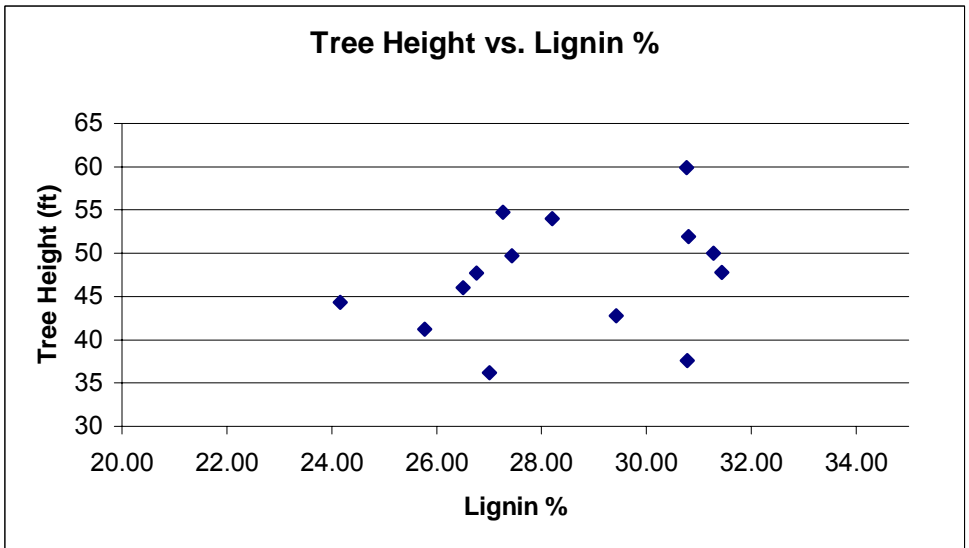


Figure 4.4.2. Graph of Tree Height versus Lignin Content

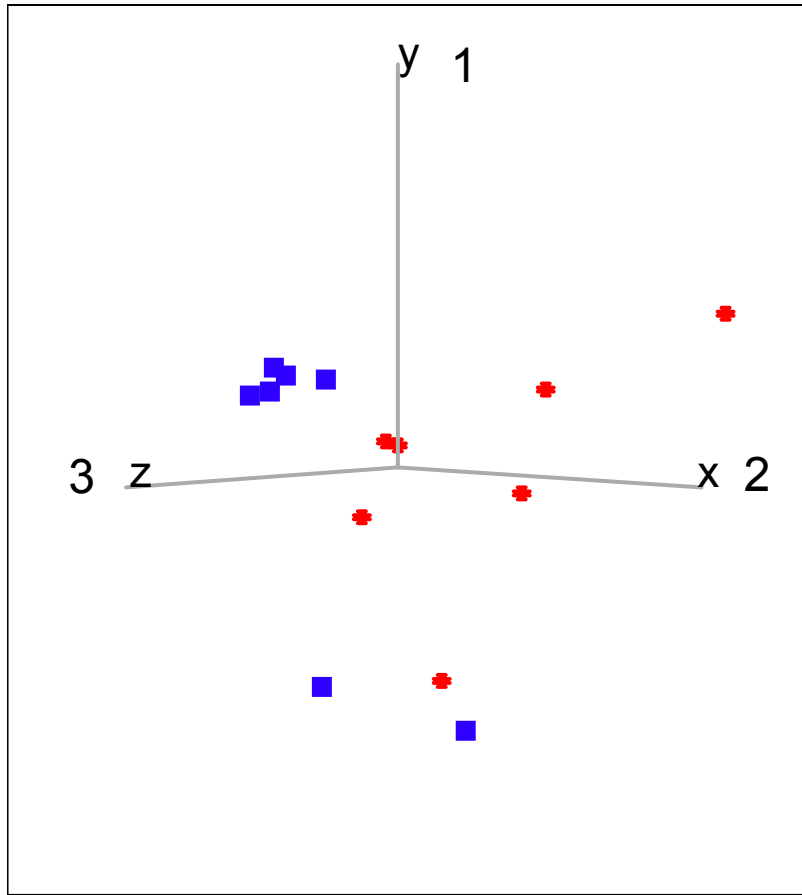


Figure 4.4.3. Phenotypic Clustering Analysis (PCA) Fast-Slow Growing Trees. Principal components 1, 2, and 3 contain 65.36% of the total information resulting from the metabolite variances. The filled squares represent trees from the slow growing family. The circles represent trees from the fast growing family.

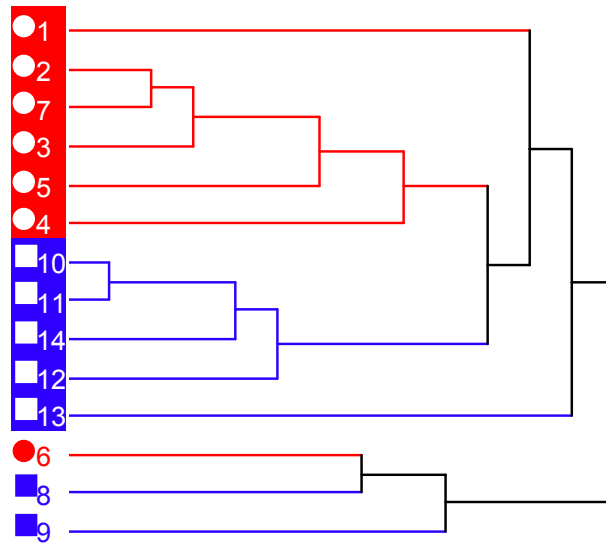


Figure 4.4.4. Hierarchical Cluster Analysis (HCA) of Fast-Slow Growing Trees. Results of this clustering analysis indicate that there is an overlap between trees in the fast and slow growing families.

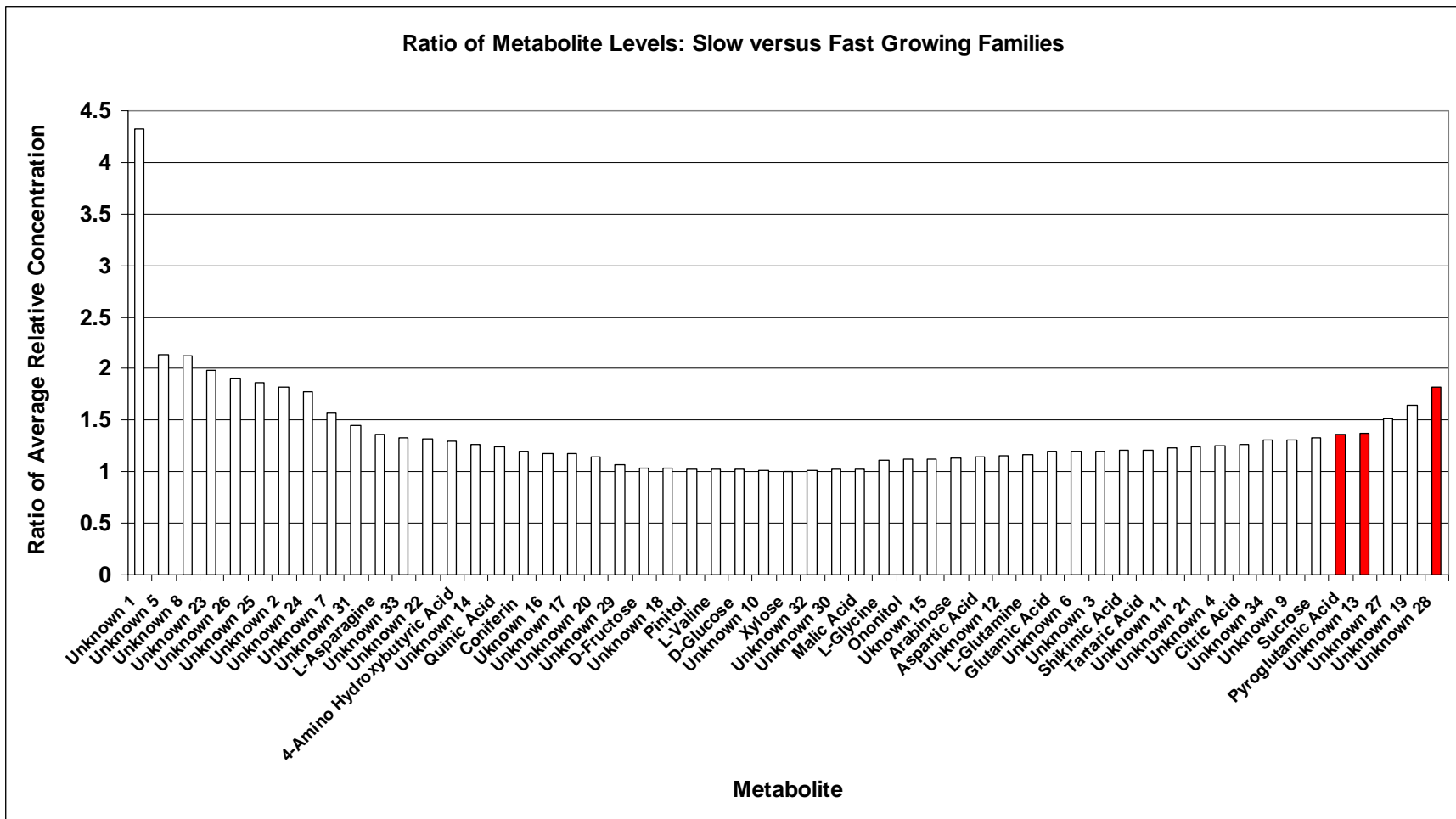


Figure 4.4.5. Ratio of Metabolite Levels in Fast and Slow Growing Families. Metabolites on the right side of the graph are high in concentrations for the fast growing family. Metabolites on the left side of the graph represent metabolites higher in concentrations for slow growing family.

Table 4.4.1. ANOVA Table of Stepwise Regression Model

Analysis of Variance

Source	DF	Sum of Squares	Mean of Square	F Value	Pr > F
Model	1	578.29714	578.29714	6.02E+11	<.0001
Error	12	1.15E-08	9.61E-10		
Corrected Totals	13	578.29714			

Table 4.4.2. Stepwise Regression Model Information

Variable	Metabolite	Parameter Estimate	Standard Error	Type II	F value	Pr > F
Intercept		69.89588	0.00029138	662.83873	5.75E+10	0.002
F15	4-Amino Hydroxybutyric Acid	-72.6432	0.01271	0.37641	3.27E+07	0.002
F22	Unknown 11	-38.56919	0.05663	0.00534	463938	0.002
F23	Unknown 12	3710.65628	0.06574	36.70151	3.19E+09	0.002
F26	L-Glutamine	1359.51945	0.05939	6.0366	5.24E+08	0.002
F30	Citric Acid	385.77912	0.00515	64.61947	5.61+09	0.002
F31	Pinitol	-28.26651	0.00025062	146.53227	1.27E+10	0.002
F38	Ononitol	226.5664	0.0034	51.15496	4.44E+09	0.002
F42	Unknown 20	-353.5664	0.01216	9.7321	8.45E+08	0.002
F45	Unknown 23	69.20054	0.01411	0.27716	2.41E+07	0.002
F47	Unknown 25	2.34668	0.08361	0.00000907	7.88E+02	0.0227
F49	Unknown 27	-217.46332	0.01241	3.53679	3.07E+08	0.002
F53	Unknown 28	-2038.89433	0.02245	95.02798	8.25+09	0.002

Microarray Results

Due to the complexity of statistical model preparation for data analysis in microarrays, the results from two models were taken and included in the results. The first model essentially uses a statistical t-test to determine which genes are differentially expressed between the fast and slow growing families. The results of this analysis indicate that 54 genes transcripts were differentially abundant. A volcano plot compares the fast growing family versus the slow growing family (Figure 6). Table 3 displays the genes that were up-regulated for the fast growing family. Table 4 shows the genes that were up-regulated in the slow growing family.

A cellulose synthase transcript level is increased in the fast growing trees (Table 3), which could result in higher cellulose content for the fast growing trees; although, this result was not consistent with the Near-IR α -cellulose content measurements. However, there is some difficulty in relating the control of mRNA transcript levels to overall phenotype. In the case where there were a larger number of individuals in the study, the cellulose content could possibly be higher in the fast growing family.

In Table 4, an important gene that is up-regulated in the slow growing families is laccase. Laccase is a gene that is associated with lignin formation (Sterjiades, Dean et al. 1992; Bao, O'Malley et al. 1993; O'Malley, Whetten et al. 1993; Dean and Eriksson 1994; Ranocha, Chabannes et al. 2002). The up-regulation of this gene would suggest an increase in lignin production, which is consistent with lignin content measurements in which the slow growing trees have higher levels. A reduction in lignin production is sometimes associated with increased tree (Hu, Harding et al. 1999; Kirst, Myburg et al. 2004).

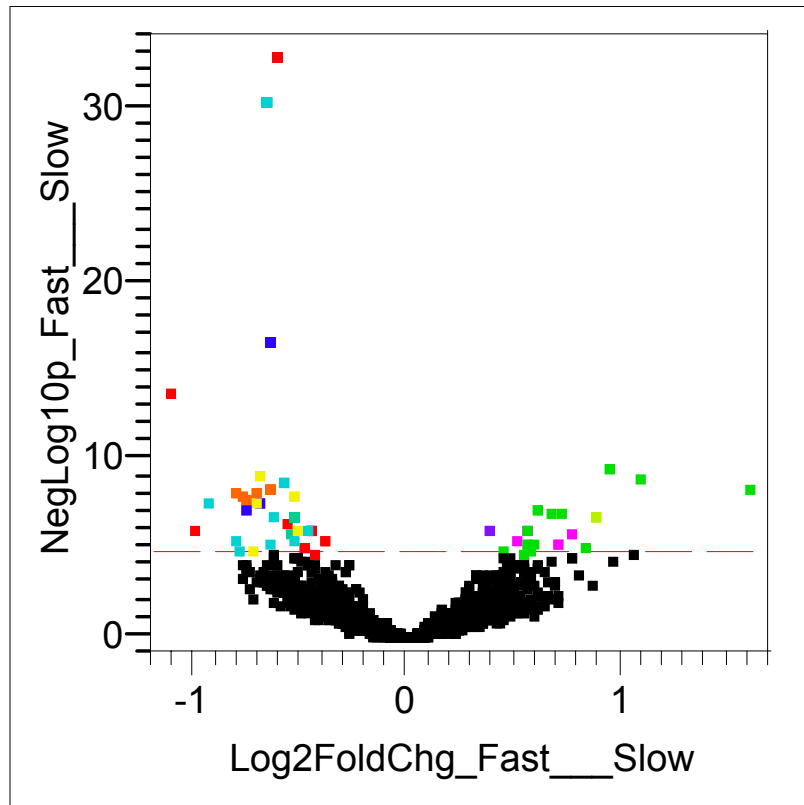


Figure 4.4.6. Volcano plot comparing fast growing family versus slow growing family. Each point of the graph represents a gene. The y-axis is $-\log_{10}$ (P-value) of the comparison. The x-axis is the estimated fold change on a \log_2 scale. Horizontal line in the graph is obtained from Bonferroni correction across the multiple hypothesis tests. Genes above the line are statistically significant under strict control of false positive rate.

Table 4.4.3. Significant Genes Up-regulated in the Fast Growing Family (Model 1)

CloneID	Putative Function	Bonferroni p-value	False Discovery Rate p-value
NXSI_105_A06	no hit	2.44471E-11	6.11178E-12
06 G07	ATP-dependent CLP protease subunit CLPP	1.16738E-07	2.33476E-08
03 F07	hypothetical 15.7 kd protein	1.59866E-06	1.77628E-07
NXPV_097_B11_F	EMB:Q9QZM6 sodium/calcium/potassium exch	1.58466E-06	1.77628E-07
NXNV047B02	EMB:Q64139 calcium binding protein, 140	2.52152E-06	2.29229E-07
30 G05	envelope glycoprotein precursor	3.92488E-06	3.27073E-07
NXPV_096_D02_F	SWP:P2CG_BOVIN 1b) (magnesium-dependent	5.19027E-06	3.70733E-07
06 B06	Fimbrial Assembly protein (Serogroup E1)	5.99704E-06	3.99802E-07
06 A10	NADH dehydrogenase subunit 2	1.39465E-05	7.43559E-07
NXSI_058_G02	EMB:Q9XEL3 Q9XEL3 PUTATIVE DEHYDRIN. >GE	1.41276E-05	7.43559E-07
NXR075_C07_F	EMB:Q9X571 putative cellulose synthase	0.000100011	4.76243E-06
07 E12	PRTD protein	0.000254376	9.78371E-06
29 C05	hyp pro	0.000327096	1.1682E-05
NXSI_139_G02	SWP:HS13_ARATH P19037 18.2 KDA CLASS I H	0.000763232	2.54411E-05
NXSI_104_B11	SWP:FER_ARATH P16972 FERREDOXIN PRECURSO	0.000973275	3.1396E-05
05 H03	putativeWD-40 repeat protein, MSI4	0.001483526	4.63602E-05
NXNV_186_D04	putative receptor-like protein kinase ER	0.002400526	6.66813E-05
16 H01	putative gdsI motif ligase/hydrolase	0.002546063	6.88125E-05
NXNV_186_D12	EMB:Q9M8N4 Q9M8N4 PUTATIVE RING ZINC FIN	0.005659061	0.000148923
22 A06	homeobox proteinPpHB7	0.006612023	0.000168127
03 B07	intracellular pathogenesis related prote	0.007946683	0.000193822
34 G05	hyp pro	0.008978867	0.000213783
34 C03	hyp pro	0.009651794	0.000221152
NXSI_055_F10	EMB:O04052 O04052 TONOPLAST INTRINSIC PR	0.009820135	0.000221152
NXCI_155_E06	PIR:T09541 T09541 transketolase (EC 2.2.	0.015088968	0.000314353
06 F05	putative beta-Ketoacyl-COA synthase	0.018427701	0.000368554
NXSI_021_A09	leucine-rich receptor-like protein kinas	0.021863475	0.000428696
NXSI_080_G02	KNAT1 homeobox-like protein(STM)	0.032116787	0.000605977
ST_06_F05	Blast for putative function	0.034043701	0.000630439
NXNV_154_G11	NAM (no apical meristem)-like protein	0.041086677	0.000733475
NXCI_060_A12	PIR:A48426 A48426 heat shock protein HSP	0.047164644	0.000799401
NXNV_156_B02	auxin response factor 6 (ARF6)	0.049757266	0.000815693

Table 4.4.4. Significant Genes Up-regulated in the Slow Growing Family (Model 1)

CloneID	Putative Function	Bonferroni p-value	False Discovery Rate p-value
NXCI_055_D03	no hit	2.61968E-07	4.36614E-08
12_G02	splicing coactivator subunit SRM300	2.16453E-06	2.16453E-07
NXNV_096_C08	EMB:O81134 O81134 INTRACELLULAR PATHOGEN	9.15404E-06	5.72128E-07
ST_21_H06	Blast for putative function	7.3182E-05	3.6591E-06
NXCI_020_A02	PIR:T09264 T09264 embryonic abundant pro	0.000113995	5.18159E-06
NXNV_076_A05	no hit	0.000207496	9.02155E-06
NXNV_086_C07	SWP:E13B_WHEAT P52409 GLUCAN ENDO-1,3-BE	0.000225946	9.41442E-06
NXSI_078_A03	EMB:Q9XEH3 Q9XEH3 FLAX RUST RESISTANCE P	0.000326848	1.1682E-05
NXNV_064_E06	EMB:BAB09857 BAB09857 Phi-1-like protein	0.000475773	1.6406E-05
NXNV_127_G04	Blast for putative function	0.001714956	5.19684E-05
ST_12_D06	Blast for putative function	0.002118843	6.23189E-05
NXNV027B07	EMB:Q9XHP6 cellulose synthase catalytic	0.002275123	6.50035E-05
NXNV_008_F05	SWP:METK_PINBN P50300 S-ADENOSYLMETHIONI	0.006725063	0.000168127
NXSI_103_C04	EMB:Q9LLZ4 Q9LLZ4 PUTATIVE ARABINOGLACT	0.009951822	0.000221152
NXNV_072_C01	no hit	0.010249267	0.00022281
NXSI_114_A04	EMB:Q9LLZ5 Q9LLZ5 PUTATIVE ARABINOGLACT	0.011174285	0.000237751
NXNV_106_F12_F	EMB:Q07524 endoglucanase 1 (ec 3.2.1.4)	0.016759962	0.00034204
NXCI_045_C01	no hit	0.027002928	0.000519287
NXSI_121_D05_F	EMB:Q07524 endoglucanase 1 (ec 3.2.1.4)	0.037846667	0.000688121
NXNV_143_B10	NAM (no apical meristem)-like protein	0.041808076	0.000733475
NXCI_006_A10	EMB:O24044 O24044 LACCASE (EC 1.10.3.2).	0.043776532	0.000754768
NXNV_147_H04	no hit	0.049593529	0.000815693

The second model applied to the microarray data considers individual trees as a random effect. This model takes into account the variability between individual trees. The application of this random effect to the model decreases the number of differentially expressed genes from 54 genes to 13 genes. In this model, there were 13 differentially expressed genes based on the q values < 0.05 (Storey 2002; Storey and Tibshirani 2003; Storey, Taylor et al. 2004). The use of the Bonferroni correction p values results in only 2 genes that are differentially expressed between the fast and slow growing families.

In Table 5, the fold change gives an indication of up-regulation for the families. A fold-change greater than 1 indicates that the particular gene is up-regulated in the fast growing family. A fold change less than 1 indicates a gene that is up-regulated in the slow growing family.

Of the 13 genes, NAM (no apical meristem)-like protein stands out being specifically related to tree growth. In the first model, two different forms of this protein are up-regulated in both the slow growing and fast growing family. However, the introduction of individual trees as a random effect into the model eliminated the association of this gene with slow growth in these families.

Table 4.4.5. Differentially Expressed Genes for Fast/Slow Growing Families (Model 2). Fold change greater than 1 represents genes up-regulated in the fast growing trees. Fold change less than 1 represent genes down-regulated in slow growing trees. Levels of significance are based on John Storey's False Discovery Rate q-value.

CloneID	Putative Function	Fold Change	p Value	q Value
NXSI058G02	putative cold-acclimation protein	0.4967683	1.56E-07	0.00025
ST33D01	unknown protein	0.5379863	5.83E-07	0.00047
NXCI100G08	cyclin	0.6978672	9.78938E-05	0.02609
NXSI029D04	glyoxalase I	0.7665418	0.000332624	0.04414
ST03G03	mitochondrial chaperonin 10	0.7725145	0.000251104	0.04414
NXSI109B06	60S ribosomal protein L37A like	0.7768720	0.000353661	0.04414
NXSI116G04	no hit	1.2383089	0.000298833	0.04414
NXNV135G10	no hit	1.3449366	4.05504E-05	0.01621
NXNV143B10	NAM (no apical meristem)-like protein	1.3455154	3.80186E-05	0.01621
NXNV155A09	no hit	1.5047351	0.000345084	0.04414
NXNV154G09	putative protein	1.7040549	0.000358823	0.04414
NXNV086C07	beta-1,3-glucanase	1.7295354	9.52877E-05	0.02609
NXSI078A03	no hit	1.7412166	0.00011534	0.02635

Results from the MVA of the metabolic profiling experiments, indicates a positive correlation of L-glutamine with tree height. In a field trial of transgenic trees, trees that were over-expressed in glutamine synthetase were found to have improved growth (Jing, Gallardo et al. 2004). Glutamine synthetase is regarded as a point of control in synthesis and the mobilization of nitrogen compounds in plants (Miflin and Lea 1980). This ATP dependent enzyme condenses ammonia with glutamic acid to yield L-glutamine. Therefore, an over-expression of glutamine synthetase could possibly result in an accumulation of L-glutamine. However, this result is not evident in the microarray analysis.

Overall, the observed phenotypic variation includes genotypic variation plus environmental variation in addition to a cross of genotypic by environmental variation. Due to the complexity of the steps involved in the expression of genotypes to transcripts to the production of polypeptides and their quaternary structures that act on these metabolites, it is difficult to determine how much control is at what level. Therefore, it may be difficult to find links between transcript levels of mRNA and metabolite pools. Developmental, physiological and environmental effects may all play a role in the formation of cellular metabolite pools.

From the microarray analysis, a NAM (no apical meristem)-like protein homolog is differentially expressed in the fast growing family. NAM is a protein found in Arabidopsis that regulates apical meristem formation in shoots (Ooka, Satoh et al. 2003; Ko, Han et al. 2004). Apical meristems are located at the tips of each growing shoot and are responsible for vertical growth (elongation) and producing the cells that form new

meristems; thus, giving the tree its primary growth. This increase in transcript levels for NAM could be directly related to greater tree heights for the fast growing family.

4.5 CONCLUSION

The step-wise linear regression height prediction model from the profiling results indicates that L-glutamine is positively correlated with tree height. It seems that this accumulation of L-glutamine could be due to an environmental effect such as extra nitrogen in the soil. Microarray analysis results from the first model showed an increase in transcript levels for a laccase, a lignin-related gene, for the slow growing family and an increase in transcripts for cellulose synthase for the fast growing family. The second model, which includes individual trees as a random variable, implicates a NAM-like protein as being up-regulated in the fast growing family.

4.6 REFERENCES

- Bao, W., D. M. O'Malley, et al. (1993). "A laccase associated with lignification in loblolly pine xylem." Science **260**: 672-674.
- Dean, J. F. D. and K.-E. L. Eriksson (1994). "Laccase and the deposition of lignin in vascular plants." Holzforschung **48**: 21-33.
- Hasseman, J. (2004). "Aminoallyl Labeling of RNA for Microarrays." The Institute For Genomic Research Standard Operating Procedure.
- Hu, W. J., S. A. Harding, et al. (1999). "Repression of lignin biosynthesis promotes cellulose accumulation and growth in transgenic trees." Nature Biotechnology **17**(8): 808-812.
- Jin, W., R. M. Riley, et al. (2001). "The contributions of sex, genotype and age to transcriptional variance in *Drosophila melanogaster*." NAT. Genet. **29**: 389-395.
- Jing, Z. P., F. Gallardo, et al. (2004). "Improved growth in a field trial transgenic hybrid poplar overexpressing glutamine synthetase." New Phytologist **164**: 137-145.
- Kerr, M. K. and G. A. Churchill (2001). "Experimental design for gene expression microarrays." Biostatistics **2**: 183-201.
- Kerr, M. K., M. Martin, et al. (2000). "Analysis of variance for gene expression microarray data." J. Comput. Bioi **7**: 819-837.

- Kirst, M., A. F. Johnson, et al. (2003). "Apparent homology of expressed genes from wood-forming tissues of loblolly pine (*Pinus taeda*) with *Arabidopsis thaliana*." Proc. Natl. Acad. Sci. USA **100**: 7383-7388.
- Kirst, M., A. A. Myburg, et al. (2004). "Coordinated Genetic Regulation Growth and Lignin Revealed by Quantitative Trait Locus Analysis of cDNA Microarray Data in an Interspecific Backcross of Eucalyptus." Plant Physiology **135**: 2368-2378.
- Ko, J.-H., K.-H. Han, et al. (2004). "Plant Body Weight-Induced Secondary Growth in *Arabidopsis* and Its Transcription Phenotype Revealed by Whole-Transcriptome Profiling." Plant Physiology **135**: 1069-1083.
- Mifflin, B. J. and P. J. Lea (1980). Ammonia assimilation. The biochemistry of plants. B. J. Mifflin. San Diego, American Press. **5**: 169-202.
- Morris, C. R., J. T. Scott, et al. (2004). "Metabolic Profiling: a new tool in the study of wood formation." J. Agric. Food Chem. **52**(6): 1427-1434.
- O'Malley, D. M., R. Whetten, et al. (1993). "The role of laccase in lignification." Plant J. **4**: 751-757.
- Ooka, H., K. Satoh, et al. (2003). "Comprehensive Analysis of NAC Family Genes in *Oryza sativa* and *Arabidopsis thaliana*." DNA Research **10**: 239-247.
- Ranocha, P., M. Chabannes, et al. (2002). "Laccase Down-Regulation Causes Alterations in Phenolic Metabolism and Cell Wall Structure in Poplar." Plant Physiology **129**: 145-155.
- Stasolla, C., P. V. Bozhkov, et al. (2004). "Variation in transcript abundance during somatic embryogenesis in gymnosperms." Tree Physiology **24**: 1073-1085.
- Stasolla, C., L. v. Zyl, et al. (2003). "The effects of the polyethylene glycol on gene expression of developing white spruce somatic embryos." Plant Physiology **131**: 49-60.
- Sterjiades, R., J. F. D. Dean, et al. (1992). "Laccase from sycamore maple (*Acer pseudoplatanus*) polymerizes monolignols." Plant Physiology **99**: 1162-1168.
- Storey, J. D. (2002). "A direct approach to false discovery rates." Journal of the Royal Statistical Society, Series B **64**: 479-498.
- Storey, J. D., J. E. Taylor, et al. (2004). "Strong control, conservative point estimation, and simultaneous conservative consistency of false discovery rates: A unified approach." Journal of the Royal Statistical Society, Series B **66**: 187-205.
- Storey, J. D. and R. Tibshirani (2003). "Statistical significance for genome-wide experiments." Proceeding of the National Academy of Sciences **100**: 9440-9445.
- Van-Gelder, R. N., M. E. Xastrow, et al. (1990). "Amplified RNA synthesized from limited quantities of heterogeneous cDNA." Proc. Natl. Acad. Sci. USA **87**: 1663-1667.
- Wolfinger, R. D., G. Gibson, et al. (2001). "Assessing gene significance from cDNA microarray expression data via mixed models." J. Comput. Bioi **8**: 625-637.
- Yeh, T.-F., H.-m. Chang, et al. (2004). "Rapid Prediction of Solid Wood Lignin Content Using Transmittance Near-Infrared Spectroscopy." J. Agric. Food Chem. **52**(6): 1435-1439.

A Study of Metabolic Changes in Fraser Fir in Response to Balsam Woolly Adelgid (BWA) Infestation

Cameron R. Morris¹, Barry Goldfarb², John Frampton², Bailian Li², Hou-min Chang¹,
and John F. Kadla^{3*}

¹Wood Chemistry, ²Forestry
North Carolina State University, Raleigh, NC 27695

³Advanced Biomaterials Chemistry
University of British Columbia, Vancouver, B.C. V6T 1Z4

* Telephone number (604) 827-5254, Fax number (604) 822-9104

email address [**john.kadla@ubc.ca**](mailto:john.kadla@ubc.ca)

5.1 Abstract

The supply of Fraser Fir trees to the Christmas tree industry is slowly being depleted as a result of infestation by the BWA. Consequently, changes occur to the tree's metabolism in reaction to infestation that can be observed by metabolic profiling. Substantial increases in concentration for polyols, shikimic acid, and some amino acids are observed in response to infestation. The accumulation of the polyols and amino acids indicate that the trees are under a state of water stress in response to the infestation by the BWA. The 'rotholz' formed by the BWA infestation restricts the flow of water and nutrients throughout the trees; thus, causing the tree to die.

Keywords: Fraser Fir, metabolic profiling, rotholz, Balsam Woolly Adelgid (BWA)

5.2 INTRODUCTION

Fraser Fir (*Abies fraseri*), also known as southern balsam Fir, is the only Fir species that is indigenous to the Southern Appalachians (Burns and Honkala 1990). The species was named after John Fraser, a Scottish botanist who explored the southern Appalachian Mountains in the late 1700's. Fraser Fir covers a large portion of the Appalachian Mountains. It serves as one of the major forest cover types for a popular recreation area. The Great Smoky Mountains, the Black Mountains and the Balsam Mountains are visited by millions of tourists each year.

Fraser Fir also has a great economic impact in the Southern Appalachians. Due to its shape, dense green foliage, fragrance, and needles that are well retained, this species is an excellent Christmas tree (Thor 1966; Thor and Barnett 1974). North Carolina is the second largest producer of Christmas trees in the United States. A total of 5.5 to 6.0 million Christmas trees are harvested annually, of which 98 percent are Fraser Fir, that generates in an estimated annual revenue industry of over \$100 million for North Carolina (Frampton 2001).

However, the Fraser Fir species is threatened by the Balsam Woolly Adelgid (BWA). Native to central Europe, the BWA is a small, blackish purple soft-bodied insect that has become a serious pest of Fraser Fir, while creating economic problems for the Christmas tree industry (Thor 1966; Footitt and Mackauer 1983; Hollingsworth and Hain 1994; Hollingsworth and Hain 1994). A basic criterion for Christmas trees is to have a straight top which gives the tree its pyramid shape. Infestation by BWA causes the tops of trees to bend. Also, infested trees lose needles faster than non-infested trees (Sidebottom 1998).

The BWA's life cycle consists of three different stages: egg, three nymphal instars, and adult wingless female (Hollingsworth and Hain 1991). The first nymphal instar is coined a "crawler" which is only present during the months of April through October (Arthur and Hain 1984; Hollingsworth and Hain 1991; Sidebottom 1998). During this period of the year, the BWA can be transported from tree to tree to create new infestations. Although wingless, the crawler can be carried to new trees by gusts of wind or when new Fir seedlings are relocated (Sidebottom 1998).

The crawler has two days to find a suitable feeding location or else the insect dies (Sidebottom 1998). Once an appropriate feeding location has been established, the insect feeds by inserting its stylet into the bark of the tree while becoming permanently attached to the host at this location for the rest of its life (Arthur and Hain 1984; Arthur and Hain 1986; Hollingsworth and Hain 1991; Sidebottom 1998). During the insertion of the stylet, the BWA releases salivary compounds. These compounds aid in the feeding process and at the same cause chemical and structural changes to occur within the bark (Hollingsworth and Hain 1991; Nicholas, Eager et al. 1998).

Parenchyma cells in the phloem tissue, surrounding the BWA's stylet, will increase in size and number (Balch, Clarke et al. 1964). As a result, the parenchyma cells within this feeding zone will eventually disintegrate and become infiltrated with resin (Hollingsworth and Hain 1991). Another site affected by this feeding process is the cambium. At this location, an atypical xylem tissue is formed, known as "rotholz," which is German for red wood (Arthur and Hain 1986; Hollingsworth and Hain 1994). Fraser Fir produces rotholz in response to injury in attempt to protect itself. However, it is thought that the tree tends to overreact or over-protect itself in response to the BWA's

feeding process. This reaction wood is very dense and severely restricts the movement of water, nutrients, and hormones through the tree. Ultimately, the tree enters a constant state of physiological drought, resulting in a reduction of photosynthesis to near zero and eventually causing the death of the host.

In recent years, there have been great strides made in the realm of plant genomics. Microarray analysis and metabolic profiling are two of the newest techniques in the plant genomics community. Metabolic profiling itself is a technique that can analyze changes in gene expression and its effects on metabolites and phenotypes during development or as a response to environmental or chemical stresses. Metabolic profiling utilizes chromatographic separation, such as gas chromatography (GC) or high-performance liquid chromatography (HPLC), and detection techniques such as mass spectrometry (MS) or nuclear magnetic resonance spectroscopy (NMR). A combination of these techniques provides a comprehensive approach to understand the flux of metabolites through the biological pathways as a result of changes in gene expression or post transcriptional regulation. Therefore, metabolic profiling will be utilized to gain important information on the accumulation of intermediate metabolites that form as a result of infestation by the BWA.

The overall goal of this study is to utilize metabolic profiling of xylem tissue in Fraser Fir as a means to detect the metabolic changes that occur in response to BWA infestation. Also, metabolic profiling of phloem tissue will be used to distinguish the differences between the Veitch Fir and Fraser Fir species. Profiling of the phloem tissue will help provide information on whether there is something on a metabolic level in

Veitch Fir phloem tissue that discourages the BWA from feeding or is there something in the Fraser Fir phloem tissue that attracts the BWA to feed.

In order to study the metabolic response of BWA infestation on Fraser Fir, xylem tissue was collected from infested and non-infested Fraser Fir trees planted in Laurel Spring, N.C. virtually a couple of feet from each other. Results from this comparison indicate an accumulation of shikimic acid, amino acids, and polyols in the infested samples. To study the differences between species, phloem tissue was collected from Veitch and non-infested Fraser Fir planted on the same location. In the comparison, there were significant accumulations of polyols, shikimic acid, and citric acid in the non-infested Fraser Fir samples. For the Veitch Fir samples there was a significant accumulation of a couple of unknowns.

5.3 MATERIALS AND METHODS

Materials. Materials and reagents were of analytical reagent grade (Aldrich Chemicals Co.) and used as received.

Methods. Xylem tissue and phloem tissue was collected from infested and non-infested Fraser Fir trees located in Laurel Spring, N.C. and Veitch Fir and non-infested Fraser Fir located in Averi County, N.C. for metabolic comparisons. The metabolic profiling method utilized is described by Morris et al. in a previous publication (Morris, Scott et al. 2004).

Software. Principal Component Analysis (PCA) was carried out in SAS JMP 5.1. Statistical t-tests were conducted using SAS Software Release 8.2.

Mass Spectra of Significant Metabolites

Xylem

Asparagine, N, N, O-TMS MS m/z (rel. int.) 258(4), 231(8), 215(5), 202(5), 188(7), 172(2), 159(5), 147(7), 141(6), 116(61), 100(11), 75(22), 74(13), 77(100)

Glutamine, N, N, N, O-TMS MS m/z (rel. int.) 419(5), 344(5), 329(4), 317(12), 301(4), 227(37), 216(9), 203(25), 188(5), 172(4), 156(12), 147(48), 128(15), 113(6), 100(9), 73(100)

Glycine, N, N, O-TMS MS m/z (rel. int.) 276(5), 248(20), 202(4), 188(6), 174(98), 158(9), 147(26), 133(20), 117(8), 100(10), 86(17), 73(100)

Phosphoric Acid, O, O, O-TMS MS m/z (rel. int.) 314(13), 299(71), 283(9), 269(5), 225(6), 211(20), 207(9), 191(11), 181(9), 165(5), 147(7), 133(17), 115(5), 103(4), 73(100)

Pinitol TMS MS m/z (rel. int.) 627(30), 555(5), 507(14), 449(31), 433(42), 405(5), 375(36), 359(13), 343(41), 318(27), 305(21), 285(9), 267(30), 260(47), 247(16), 217(46), 207(33), 191(67), 179(10), 161(12), 159(21), 142(34), 134(31), 129(36), 103(19), 89(9), 73(100)

Shikimic Acid TMS MS m/z (rel. int.) 462(3), 372(7), 357(8), 282(17), 254(19), 204(100), 189(17), 147(25), 73(85)

Serine, N, O, O-TMS MS m/z (rel. int.) 306(5), 278(7), 218(20), 204(30), 188(7), 172(4), 147(10), 133(7), 116(6), 100(13), 73(100)

Tartaric Acid, O, O, O, O-TMS MS m/z (rel. int.) 423(5), 351(4), 333(4), 321(4), 305(6), 292(16), 277(4), 263(4), 219(12), 189(10), 175(4), 163(3), 147(40), 133(10), 117(6), 102(8), 73(100)

Unknown 1 (U1) MS m/z (rel. int.) 295(1), 232(6), 218(13), 217(22), 207(12), 173(8), 163(17), 149(45), 148(16), 147(100), 133(30), 129(14), 115(6), 75(11), 72(92), 59(8)

Unknown 3 (U3) MS m/z (rel. int.) 178(2), 177(16), 161(7), 149(32), 148(17), 147(100), 133(9), 103(2), 75(6), 73(42)

Unknown 7 (U7) MS m/z (rel. int.) 221(1), 176(7), 175(17), 174(92), 149(7), 148(5), 147(37), 133(15), 131(15), 130(15), 100(20), 86(34), 73(100), 59(35)

Unknown 12 (U12) MS m/z (rel. int.) 409(2), 319(4), 293(7), 292(26), 245(4), 220(12), 217(20), 189(5), 149(20), 148(11), 147(84), 102(22), 73(100)

Unknown 15 (U15) MS m/z (rel. int.) 307(7), 277(4), 245(5), 231(11), 217(39), 189(9), 156(4), 149(15), 148(8), 147(40), 129(14), 75(12), 74(10), 73(100)

Unknown 20 (U20) MS m/z (rel. int.) 394(4), 333(5), 319(5), 305(5), 245(4), 221(8), 218(17), 217(80), 205(9), 189(13), 149(27), 148(16), 147(100), 129(17), 117(12), 73(80)

Unknown 22 (U22) MS m/z (rel. int.) 333(2), 305(4), 292(7), 245(8), 221(11), 217(42), 189(23), 175(8), 147(49), 133(13), 129(28), 81(8), 75(8), 74(9), 73(100)

Unknown 23 (U23) MS m/z (rel. int.) 422(1), 421(2), 347(3), 333(7), 307(3), 292(22), 277(4), 258(2), 257(5), 217(33), 205(2), 204(5), 189(10), 147(50), 129(12), 117(5), 103(6), 102(7), 73(100)

Unknown 24 (U24) MS m/z (rel. int.) 347(2), 333(8), 332(2), 331(5), 307(3), 293(8), 292(23), 257(6), 219(4), 218(6), 217(35), 189(13), 149(15), 148(9), 147(53), 129(11), 103(7), 74(9), 73(100)

Unknown 29 (U29) MS m/z (rel. int.) 321(1), 320(4), 319(11), 269(5), 268(13), 267(47), 217(10), 205(9), 193(6), 157(15), 149(11), 148(8), 147(50), 131(6), 130(4), 129(25), 117(9), 73(100)

Unknown 30 (U30) MS m/z (rel. int.) 321(5), 320(6), 319(22), 245(7), 217(16), 204(5), 157(26), 148(8), 147(60), 129(35), 73(100)

Unknown 38 (U38) MS m/z (rel. int.) 279(4), 217(2), 168(3), 167(30), 150(10), 149(100), 121(4), 104(6), 71(6), 57(6), 55(11)

Unknown 41 (U41) MS m/z (rel. int.) 637(1), 636(11), 635(21), 634(33), 563(8), 562(18), 546(17), 545(34), 544(68), 474(5), 473(17), 472(37), 450(6), 413(11), 381(12), 362(10), 361(32), 271(20), 243(23), 217(38), 179(8), 169(38), 147(38), 129(31), 74(11), 73(100)

Unknown 43 (U43) MS m/z (rel. int.) 438(5), 437(13), 364(3), 363(17), 362(29), 361(100), 347(4), 319(5), 271(15), 257(14), 243(17), 218(11), 217(34), 204(20), 191(10), 169(39), 147(25), 129(22), 103(6), 73(69)

Unknown 45 (U45) MS m/z (rel. int.) 450(11), 412(6), 386(10), 385(21), 384(60), 361(22), 338(13), 323(14), 298(23), 297(94), 271(11), 243(18), 225(51), 217(29), 206(11), 169(29), 147(25), 129(20), 103(6), 73(100)

Phloem

Inositol TMS MS m/z (rel. int.) 507(5), 432(16), 393(5), 367(9), 343(4), 318(67), 305(83), 284(11), 265(23), 217(86), 204(25), 191(41), 147(75), 129(16), 103(8), 73(100)

Ononitol TMS MS m/z (rel. int.) 496(5), 437(6), 349(17), 319(4), 259(57), 217(30), 204(61), 189(8), 157(9), 147(31), 129(18), 117(6), 103(28), 89(5), 73(100)

Unknown 15 (U15) MS m/z (rel. int.) 420(2), 333(6), 305(15), 292(19), 219(9), 217(9), 216(12), 191(12), 147(31), 131(15), 130(13), 129(10), 75(24), 73(100)

Unknown 17 (U17) MS m/z (rel. int.) 309(5), 308(3), 307(13), 277(5), 256(4), 219(7), 218(14), 217(98), 189(8), 149(6), 148(4), 147(24), 129(10), 103(15), 74(7), 73(100)

Unknown 23 (U23) MS m/z (rel. int.) 422(1), 421(2), 347(3), 333(7), 307(3), 292(22), 277(4), 258(2), 257(5), 217(33), 205(2), 204(5), 189(10), 147(50), 129(12), 117(5), 103(6), 102(7), 73(100)

Unknown 27 (U27) MS m/z (rel. int.) 447(7), 419(8), 389(10), 373(21), 361(6), 333(13), 305(12), 292(24), 277(16), 276(22), 275(95), 257(8), 247(28), 219(11), 207(10), 204(33), 189(17), 148(10), 147(33), 131(16), 116(18), 113(20), 85(18), 74(16), 73(100)

Unknown 37 (U37) MS m/z (rel. int.) 361(4), 319(6), 290(11), 243(3), 219(3), 218(3), 217(20), 207(5), 149(14), 148(6), 147(37), 133(14), 129(13), 75(10), 74(9), 73(100)

5.4 RESULTS AND DISCUSSION

Typically we utilize, differentiating xylem tissue when conducting metabolic profiling of wood formation in trees; however, in this case phloem was used as part of the profiling experiment due to the insertion of the BWA's stylet into the bark. It is our belief that the profiling of the bark tissue may give an indication of the metabolic changes that occur due to the insertion of the stylet and the release of the insect's saliva. In addition, profiles of the differentiating xylem tissue will provide an illustration of the

changes in the metabolic network of the differentiating xylem caused in reaction to saliva released into the tree bark.

In this study, six trees were chosen to investigate the differences between BWA infested and non-infested Fraser Fir trees to obtain information on why these infested trees are dying in response to BWA infestation. Immature xylem tissue and phloem was collected from each individual tree in Laurel Springs, North Carolina for metabolite profiles. Also, included in the study was a selection of xylem and phloem from three Veitch Fir located in Avery County, North Carolina. The Veitch Fir is a species that is thought to be resistant to the BWA, although the mechanism of resistance is not known. The intent of introducing the Veitch Fir to the study is to determine if there is something related to the metabolites in the phloem tissue of Veitch Fir that prevents infestation by the BWA or if whether there is something related to metabolites that attracts BWA in Fraser Fir phloem tissue.

The first step in the analysis of the metabolite data is to apply principal components analysis to all of the chromatographic data to produce phenotypic clusters (Fiehn, Kopka et al. 2000). The next step is the production of X by Y plots of the principal components (PC) that were used to create the phenotypic clusters. These graphs assist in the identification of individual metabolites that are responsible for the phenotypic clustering results. The last step of the process is the creation of the ratio of relative concentration graphs for each individual metabolite. Unlike the X by Y plots, these graphs provide a visualization of the differences in the relative concentrations of each individual metabolite between samples in a study. Also, this graph can illustrate metabolites that are significantly different between samples. All three steps of the data

analysis provide insight into the biological changes that occur between the samples due to environmental factors, genetic variation, or in this case, insect infestation.

Figures 1 and 2 illustrate the polar phase phenotypic cluster results achieved by the metabolic profiles of the xylem and phloem tissue samples respectively. Figure 1 displays distinct clustering between all four sample groups using the first three principal components. The BWA infested Fraser Fir samples from Laurel Springs shows separation from the remaining samples along the First principal component. While the non-infested Fraser Fir and Veitch Fir is separated by the second and third principal components.

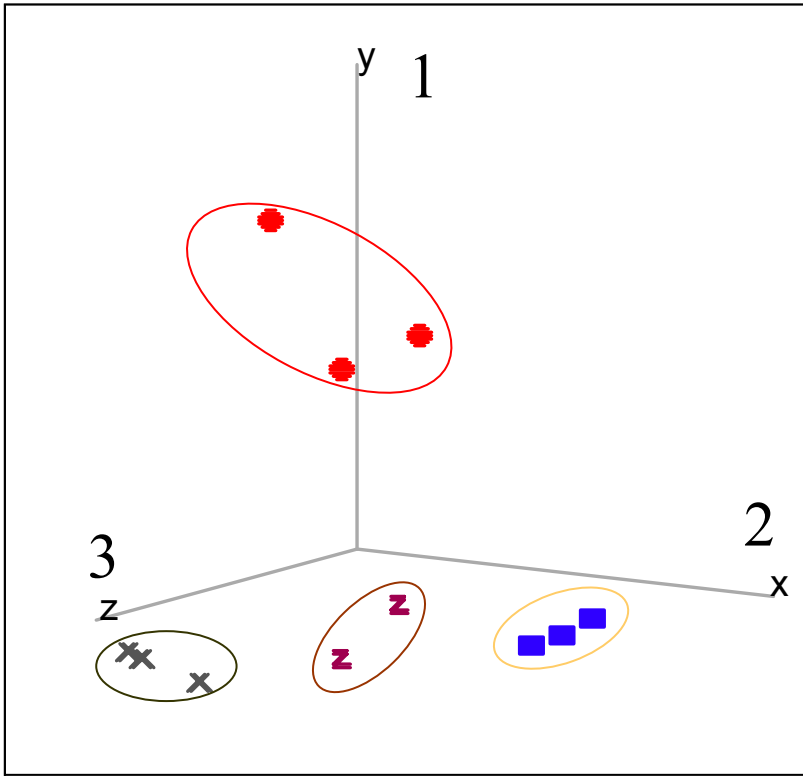


Figure 5.4.1. Metabolic Phenotypic Clustering Results for Xylem Tissue

circle = infested Fraser Fir

square = non-infested Fraser Fir

x = Veitch Fir

z = infested Fraser Fir from Veitch Fir

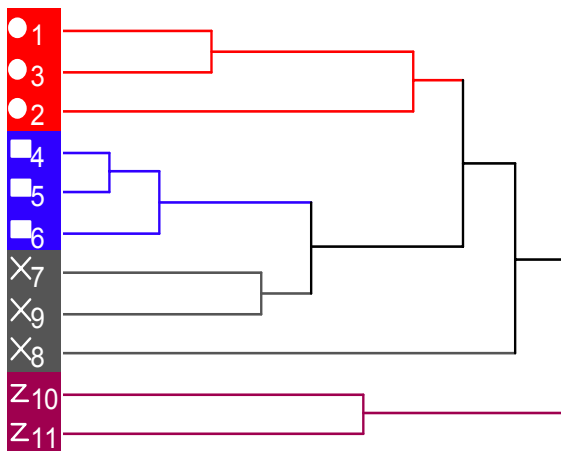
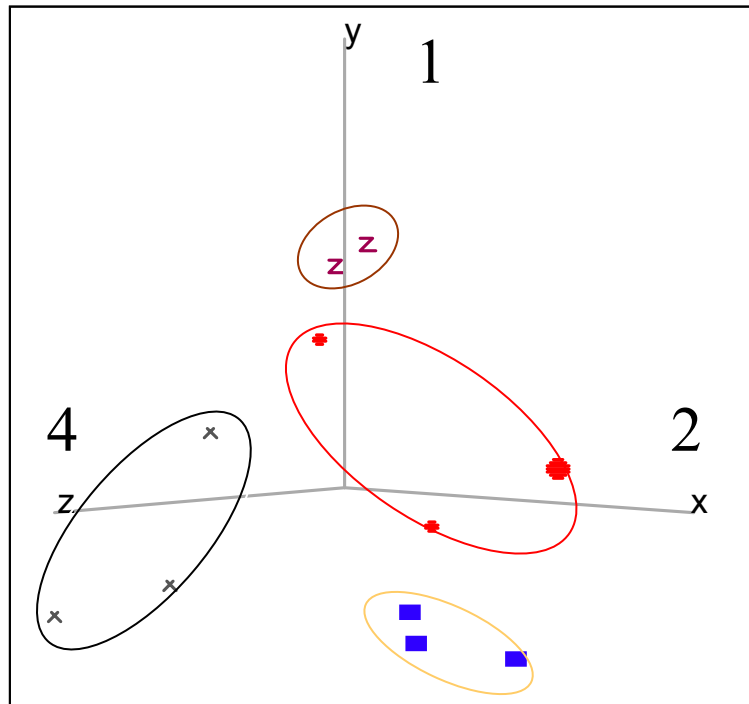


Figure 5.4.2. Metabolic Phenotypic Clustering Results for Phloem Tissue.

circle = infested Fraser Fir

square = non-infested Fraser Fir

x = Veitch Fir

z = infested Fraser Fir from Veitch Fir

Xylem Tissue

Figure 3 displays the X by Y plot of the principal components that separate out the phenotypic clusters for the xylem tissue. For instance, a plot of PC1 versus PC2 which are the principal components that separate BWA infested and non-infested Fraser Fir identify the metabolites that are responsible for the clustering results between the two samples. According to the graph, there are over twenty metabolites that are responsible for the infested phenotypic cluster.

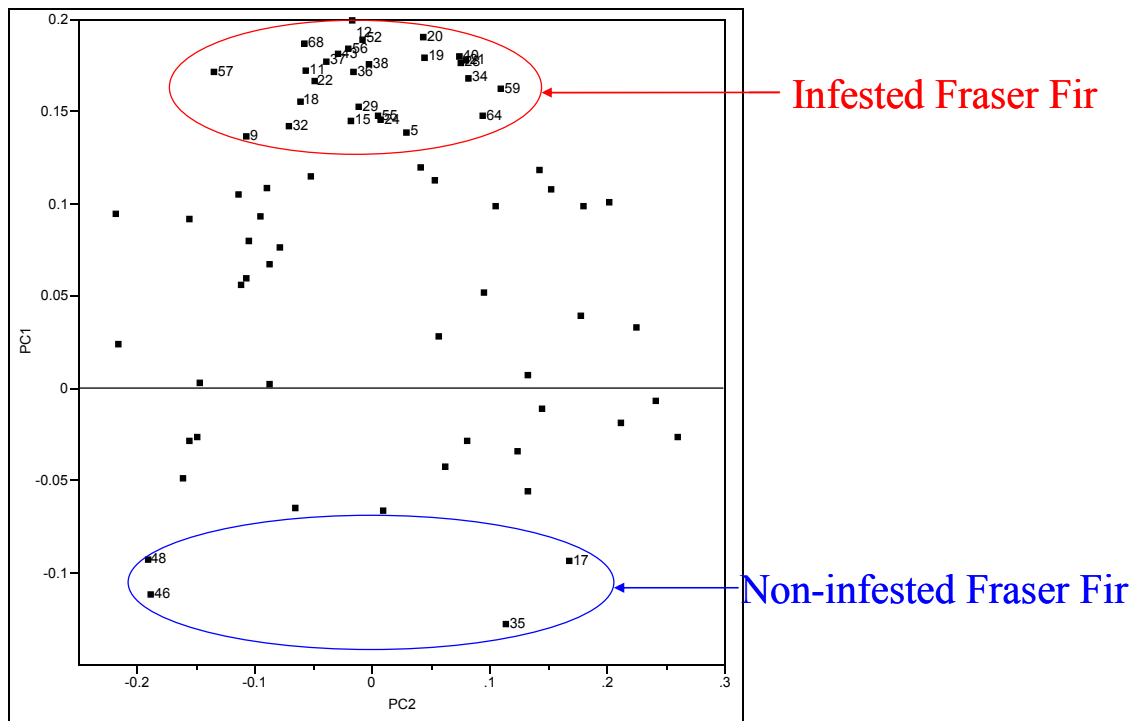


Figure 5.4.3. Plot of PC 1 versus PC 2 separating infested and non-infested Fraser Fir xylem tissue samples. Metabolites further away from zero play a greater role in the clustering results.

Coupled with this X by Y plot is a graph generated in Excel that displays the ratios of the relative concentrations of individual metabolites. Figure 4 shows the individual metabolite breakdown for the infested versus non-infested Fraser Fir. The graph pinpoints the metabolites that are significantly higher in relative concentrations. At a first glance, the figure illustrates that there are differences in relative concentration for a number of amino acids, a polyol, organic acids, and phosphoric acid. Three metabolites not shown in figure 4 are seen in figure 5 because their ratios were too high and would offset the scale of figure 4. One metabolite with a high ratio for the non-infested Fir was unknown 45 (possibly a glucoside), while unknown 30 (possibly a carbohydrate) and tartaric acid contain high ratios for the infested samples. Also, there were a number of metabolites that were only present in the BWA infested trees, such as unknowns 3 (possibly an organic acid), 20 (possibly a carbohydrate or aldonic acid), 38, 41, and 43 (possibly a disaccharide), serine, and asparagine.

After the completion of t-tests for each individual metabolite, less than half of the metabolites were found to be significantly different in concentration between the infested and non-infested samples. Unknown 22 (possibly an aldonic acid) and D-fructose were found to significantly accumulate in the xylem tissue for the non-infested Fraser Fir samples. On the other hand, phosphoric acid, shikimic acid, pinitol, glycine, and several unknowns 1, 7, 12, 15 (possibly a carbohydrate), 23 (possibly an aldonic acid), 24, and 29 (possibly a carbohydrate) were significantly higher in concentration for the infested Fraser Fir samples.

To observe the biological significance of these changes in metabolite concentration in response to infestation, there was an increase in the pools of shikimic

acid, which indicates a possible increase in extractives production with regards to flavonoids and tannins which may pool in response to infestation. Flavonoids, which are commonly present in heartwood, sapwood, and bark, are thought to provide resistance to fungal and insect attack. Also, the increase in shikimic acid production could lead to amplification in lignin production in response to the BWA feeding process. In fact, chemical analysis of Rotholz wood analysis revealed approximately 15 percent higher lignin content in comparison to normal Fraser Fir wood (Balakshin, Capanema et al. 2004).

In addition to the increase in pools of shikimic acid, there was a significant increase in the relative concentration of glutamine in the infested samples. Also, there was the detectable presence of serine and asparagine in BWA infested samples which were not found in non-infested samples. Concentrations of both glutamine and asparagine have been found to accumulate in trees during water drought (Cyr, Buxton et al. 1990).

Another metabolite that has a significantly higher relative concentration for the BWA infested samples was pinitol. Pinitol is a cyclic polyol, which is a polyhydric alcohol (Jefferies 1980; Gorham, Hughes et al. 1981; Bohnert and Shen 1999; Ashraf 2004). Polyols increase in concentration in plant species as a result of drought stress (Ford 1982; Ford 1984; Bieleski 1994; Popp and Smirnoff 1995; Nelson, Koukoumanos et al. 1999). More specifically, pinitol operates as a scavenger or antioxidant in the removal of free radicals and plays a vital role in intracellular osmotic adjustment between the vacuole and the cytoplasm (Ashraf 2004; Griffin, Ranney et al. 2004). In the case of this study, the accumulation in pinitol concentration observed from the profiling results

could be in response to water deficiency (Ford and Wilson 1981; Ford 1984; Griffin, Ranney et al. 2004).

The results of the xylem tissue profiling of the BWA infested and non-infested trees indicate that the infested trees are experiencing water stress. This result is evident by the increases in relative concentrations in pinitol and a couple of amino acids such as asparagine, and glutamine. The water stress that the trees are experiencing could be in response to the BWA's feeding process, creating a dense reaction wood that restricts the movement of water and nutrients through the tree.

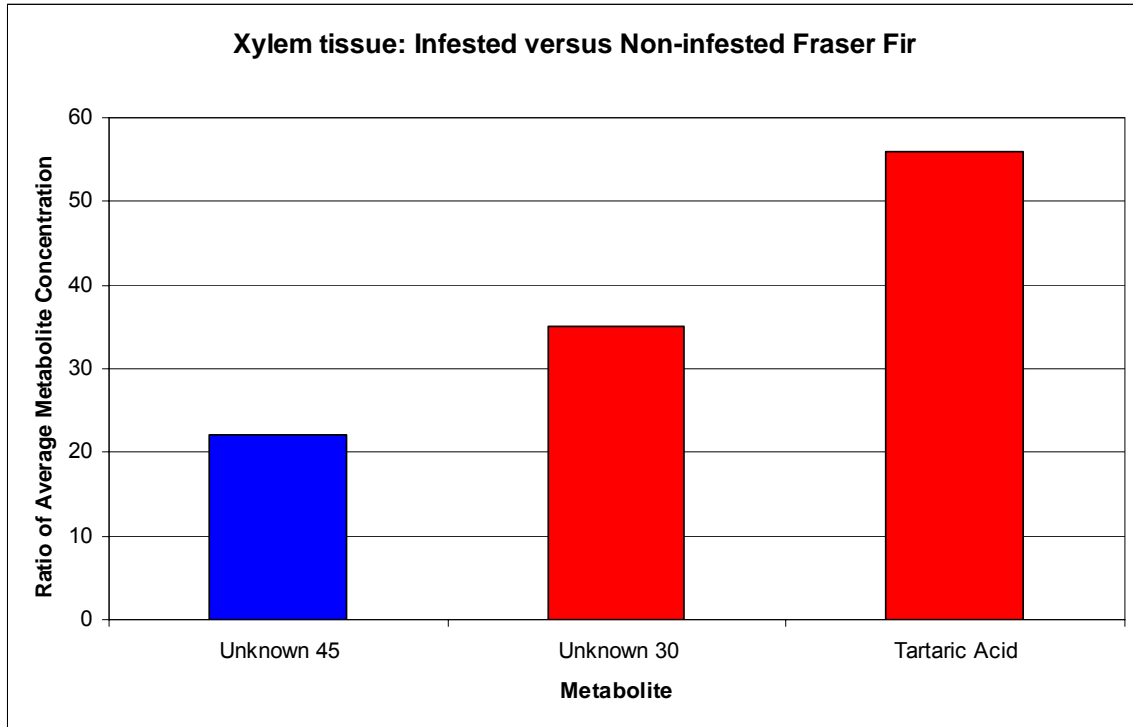


Figure 5.5.5. Graph of the ratio of the average relative concentration of individual metabolites: unknowns 45, 30 and tartaric acid. Unknown 45 accumulates in non-infested Fraser Fir and unknown 30 and tartaric acid accumulate in infested Fraser Fir.

Phloem Tissue

The BWA inserts its stylet into the bark and releases salivary compounds. These compounds causes the parenchyma cells in the phloem tissue surrounding the stylet to increase in size and number (Balch, Clarke et al. 1964; Hollingsworth and Hain 1991; Nicholas, Eager et al. 1998). This feeding process eventually causes the parenchyma cells to disintegrate and to be filled with resin (Hollingsworth and Hain 1991). As a result, profiling of the phloem tissue was conducted to investigate the differences in profiles as a result of the BWA feeding process. The hierarchal clustering analysis shows the phloem profiles separate into four distinct clusters (Figure 2).

Figure 6 illustrates the differences in phloem tissue metabolic profiles between the non-infested Fraser Fir and Veitch Fir species. There is a significant accumulation of organic acids, polyols, and a couple of unknowns 15, 27, and 37 (possibly a disaccharide) for Fraser Fir phloem tissue in comparison to Veitch Fir. However, the Veitch Fir shows a significant increase in a couple of unknowns, 17 (possibly a carbohydrate) and 23 (possibly an aldonic acid). Therefore, between the species, there seems to be some differences between several metabolites. These differences could either play a part in the attraction of the BWA to the Fraser Fir or to the resistance of the BWA by the Veitch Fir. Further test should be run to determine the attraction of the BWA to Fraser Fir and not Veitch Fir.

5.5 CONCLUSION

According to the results, PCA of metabolite metabolic profiling created distinct phenotypic clusters between infested and non-infested Fraser Fir, and Veitch Fir. Findings from the immature xylem tissue profiling illustrate the fact that the infested trees are encountering a state of water stress as a result of the BWA's feeding process. A number of amino acids and polyols have been suggested to accumulate under water deficiency. The pools of shikimic acid suggest amplification in lignin as well as extractives. Therefore, it seems that the infestation of the BWA could restrict the flow of water and nutrients throughout the tree; thus causing the tree to produce 'rotholz wood' and eventually resulting in death.

The results of the phloem tissue profiling provide a starting point to observe the differences between Fraser Fir and Veitch Fir. These metabolic differences could give an indication of the attraction of the BWA to feed on the Fraser Fir. Likewise, it could give an indication of the metabolic properties of the Veitch fir that discourage the BWA from landing or feeding. Due to significant accumulation of several unknowns, further analysis of future phloem metabolite samples may be required to gain a further understanding in the differences between the Fir species.

5.6 REFERENCES

- Arthur, F. H. and F. P. Hain (1984). "Seasonal History of the Balsam Woolly Adelgid (Homoptera: Adelgidae) in Natural Stands and Plantations of Fraser Fir." Journal of Economic Entomology **77**(5): 1154-1158.
- Arthur, F. H. and F. P. Hain (1986). "Water Potential of Fraser Fir Infested with Balsam Woolly Adelgid (Homoptera: Adelgidae)." Environmental Entomology **15**(4): 911-913.

- Ashraf, M. (2004). "Potential biochemical indicators of salinity tolerance in plants." Plant Science **166**: 3-16.
- Balakshin, M. Y., E. A. Capanema, et al. (2004). "NMR Studies on Fraser Fir Lignins." Plant Science **166**: 17-24.
- Balch, R. E., J. Clarke, et al. (1964). "Hormonal action in products of tumors and compression wood in *Abies balsamea* by an (*Adelges piceae*)." Nature **202**: 721-722.
- Bieleski, R. L. (1994). "Pinitol is a major carbohydrate in leaves of some coastal plants in indigenous to New Zealand." New Zealand J. Bot. **32**: 73-78.
- Bohnert, H. J. and B. Shen (1999). "Transformation and compatible solutes." Scient. Hort. **78**: 237-260.
- Burns, R. M. and B. H. Honkala (1990). Conifers. Ag. Handbook. U.S.D.A. For. Serv, Washington, D.C.
- Cyr, D. R., G. F. Buxton, et al. (1990). "Accumulation of free amino acids in the shoots and roots of three northern conifers during drought." Tree Physiology **6**(3): 293-303.
- Fiehn, O., J. Kopka, et al. (2000). "Metabolite profiling for plant functional genomics." Nature Biotechnology **18**(11): 1157-1161.
- Footitt, R. G. and M. Mackauer (1983). "Subspecies of Balsam Woolly Aphid, *Adelges piceae* (Homoptera: Adelgidae), in North America." Annals of the Entomological Society of America **76**(2): 299-304.
- Ford, C. W. (1982). "Accumulation of O-methyl-inositols in water-stressed *Vigna* species." Phytochemistry **21**: 1149-1151.
- Ford, C. W. (1984). "Accumulation of low molecular weight solutes in water-stressed tropical legumes." Phytochemistry **23**: 1007-1015.
- Ford, C. W. and J. R. Wilson (1981). "Changes in levels of solute during osmotic adjustment to water stress in leaves of four tropical pasture species." Aust. J. Plant Physiol. **1981**: 77-91.
- Frampton, J. (2001). North Carolina's Christmas Tree Genetics Program. Proc. 26th South. For. Tree Improv. Conf., Univ. of GA.
- Gorham, J., L. L. Hughes, et al. (1981). "Low-molecular-weight carbohydrates in some salt-stressed plants." Physiol. Plant **53**: 1-6.
- Griffin, J. J., T. G. Ranney, et al. (2004). "Heat and Drought Influence Photosynthesis, Water Relations, and Soluble Carbohydrates of Two Ecotypes of Redbud (*Cercis canadensis*)." J. Amer. Soc. Hort. Sci. **129**(4): 497-502.
- Hollingsworth, R. G. and F. P. Hain (1991). "Balsam Woolly Adelgid (Homoptera: Adelgidae) and Spruce-Fir Decline in the Southern Appalachians: Assessing Pest Relevance in a Damaged Ecosystem." Florida Entomologist **74**(2): 1-6.
- Hollingsworth, R. G. and F. P. Hain (1994). "Balsam Woolly Adelgid (Homoptera: Adelgidae): effects on growth and abnormal wood production in Fraser fir seedlings as influenced by seedling genetics, insect source, and soil source." Can. J. For. Res. **24**: 2284-2294.
- Hollingsworth, R. G. and F. P. Hain (1994). "Effect of drought stress and infestation by the balsam woolly adelgid (Homoptera: Adelgidae) on abnormal wood production in Fraser fir." Can. J. For. Res. **24**: 2295-2297.
- Jefferies, R. L. (1980). The role of organic solutes in osmoregulation in halophytic higher plants. Genetic Engineering of Osmoregulation: Impact on Plant Productivity for

- Food, Chemicals and Energy. D. W. Rains, R. C. Valentine and A. Hollander. New York, Plenum Press: 135-159.
- Morris, C. R., J. T. Scott, et al. (2004). "Metabolic Profiling: a new tool in the study of wood formation." J. Agric. Food Chem. **52**(6): 1427-1434.
- Nelson, D. E., M. Koukoumanos, et al. (1999). "Myo-inositol-dependent sodium uptake in ice plant." Plant Physiol. **119**: 165-172.
- Nicholas, N. S., C. Eager, et al. (1998). Threatened Ecosystem High Elevation Spruce-Fir Forest. Ecosystem Management for Sustainability: principles and practices illustrated by a regional biosphere reserve cooperative. J. D. Peine. New York, Lewis Publishers: 431-454.
- Popp, M. and N. Smirnoff (1995). Polyol accumulation and metabolism during water deficit. Environment and Plant Metabolism: Flexibility and Acclimation. N. Smirnoff. Oxford, Bios Scientific: 199-215.
- Sidebottom, J. (1998). The Balsam Woolly Adelgid, North Carolina Cooperative Extension Service.
- Thor, E. (1966). "Christmas tree research in Tennessee." American Christmas Tree Journal **10**(3): 7-12.
- Thor, E. and P. E. Barnett (1974). "Taxonomy of *Abies* in Southern Appalachians: variation in balsam monoterpenes and wood properties." Forest Science **20**(1): 32-40.

Metabolite Analysis of Clones in a Cellulose Study

Cameron R. Morris¹, Barry Goldfarb², Fikret Isik², Bialian Li², Hou-min Chang¹, Ronald R. Sederoff³, and John F. Kadla^{4*}

¹ Wood Chemistry, ²Forestry, ³ Forest Biotechnology,
North Carolina State University, Raleigh, NC 27695

⁴Advanced Biomaterials Chemistry
University of British Columbia, Vancouver, B.C. V6T 1Z4

* Telephone number (604) 827-5254, Fax number (604) 822-9104

email address john.kadla@ubc.ca

6.1 Abstract

In a current clonal study, clones of two different families were determined to have high and low cellulose content on two different sites. From each family, one clone contains high cellulose while the other clone contains low cellulose. Metabolite analysis was conducted to examine the differences in metabolic content between the cellulose classifications. Since the clones were planted on two different sites, variations between clones were observed on the same site and between sites allowing a comparison of environmental and genetic differences. Results related to genetics differences indicate an accumulation of phosphoric acid, inositol, and free sugars, glucose, fructose, and mannose, in the high cellulose clone. The accumulation of fructose suggests that it is not recycled back or is slow in recycling back to the cellulose biosynthesis pathway to produce sucrose for subsequent conversion to UDP-glucose.

Keywords: Metabolic profiling, α -cellulose, clones, loblolly pine, xylem tissue, cellulose biosynthesis

6.2 INTRODUCTION

Cellulose is the major structural component of the plant cell wall. Cellulose provides trees with a framework that makes them stronger, which allows them to stand upright (Cano-Delgado, Penfield et al. 2003). One can envision the plant cell wall as having comparable strength of reinforced concrete (Haigler, Ivanova-Datcheva et al. 2001). In growing tissue, cells must be flexible to adjust for cell growth and changes in cell shape. Therefore, cellulose orientation in the cell wall and its ability to break and reform chemical bonds between hemicellulose allow this flexibility to occur with the growth of cell walls.

Cellulose makes up as much as 45 percent of a tree's dry weight. Due to the large volume of cellulose, it is an important component in converting wood to lumber, beams, posts and telephone poles. Also, through various pulping processes cellulose produced by trees can be isolated and used as fibers. For example, cellulose fibers are isolated from trees in the form of pulp and used to produce various forms of paper and paperboard.

Since cellulose plays such an important role in the pulp and paper industry, any relative increase in cellulose content in the trees used by mills would enhance the paper production of the industry tremendously. By starting with more cellulose in the tree, additional amounts of cellulose or fibers should be obtained at the conclusion of the pulping process. Thus, the higher cellulose amounts will increase the amount of pulp yield.

In a previous study, the average α -cellulose content for 14 full-sib loblolly pine families was varied from 39 to 43.1 percent (Sykes, Isik et al. 2003). Individuals from

four of the 14 families were chosen to conduct metabolic profiling experiments. Two of the families exhibited high α -cellulose content while the other two families displayed values of low α -cellulose.

Although principal component analysis was unable to separate the groups into two distinct clusters based on α -cellulose content, a α -cellulose model was constructed utilizing metabolite analysis. This model was able to provide information on individual metabolites significance in the α -cellulose calculation. From the model, significant metabolites were determined to be positively or negatively correlated with cellulose.

Also observed were the differences between individual metabolites for those samples that displayed extreme levels of high and low α -cellulose content. Results from those comparisons indicate an accumulation of sucrose and fructose in the low α -cellulose samples. The conversion of sucrose to UDP-glucose by P-SuSy was thought to be inhibited by accumulations of fructose. This idea maybe reflected in the profiling results by increased fructose, which could be inhibiting the conversion of sucrose to UDP-glucose, potentially reducing the production level of cellulose. The hypothesis is consistent with accumulation of sucrose in the low cellulose samples, which indicates that sucrose is not being rapidly converted into UDP-glucose by P-SuSy for cellulose production.

To take the study a step further, clonal samples from two families exhibiting high and low α -cellulose content were chosen from two different sites. The previous study used trees from seedlings, in which the effects were confounded. Clones were used in this study to decompose the genetic and environmental effects of cellulose production in loblolly pine. The main objective of this study is to gain a further understanding of

cellulose biosynthesis in trees using metabolite analysis. By using clones as samples in this study, information can be gained about the variation within clones on the same site and variation within clones on different sites. Also, observations will be made on the variation between families on the same site and different sites.

6.3 MATERIALS AND METHODS

Materials. Materials and reagents were of analytical reagent grade (Aldrich Chemicals Co.) and used as received.

Methods. Xylem tissue was collected from clones from two different families, both of which were planted on two locations, Monroe County, Alabama and Nassau County, Florida. For each family samples were collected from two clones, one is a high α -cellulose clone and the other being a low α -cellulose clone. The profiling method on xylem tissue was conducted according to Morris et al. (Morris, Scott et al. 2004).

Metabolite data was normalized based on the experimental design, utilizing site, clone, and site by clone. During the normalization process, site was treated as a fixed effect to remove environmental effects. The new normalized values were applied in the principal component analysis (PCA). Based on the PCA results, there were distinct clusters created for samples from the top and bottoms of the tree. However, there was no distinct clustering on the basis of site and clone.

Software. Principal Component Analysis (PCA) was carried out in SAS JMP 4.04. Analysis of variance was completed using SAS Release 8.02. The following software was used to create a browser-based interface to a profiling database: Apache web server, PHP and the PostgreSQL database.

Mass Spectra of Significant Metabolites

Alanine TMS MS m/z (rel. int.) 304(7), 290(24), 248(93), 232(10), 174(100), 160(7), 147(63), 133(21), 130(19), 117(7), 100(25), 86(36), 73(89)

Coniferin TMS MS m/z (rel. int.) 507(9), 450(20), 361(47), 324(98), 293(10), 271(22), 243(34), 217(42), 204(21), 169(33), 147(43), 129(19), 103(10), 73(100)

Fructose TMS MS m/z (rel. int.) 570 (14), 480(14), 364(10), 307(20), 277(10), 217(100), 205(2), 147(10), 103(14), 73(50)

Glucose TMS MS m/z (rel. int.) 570 (32), 480(27), 319(100), 229(12), 217(32), 205(11), 157(43), 147(21), 129(53), 73(80)

Glutamine, N, N, N, O-TMS MS m/z (rel. int.) 419(5), 344(5), 329(4), 317(12), 301(4), 227(37), 216(9), 203(25), 188(5), 172(4), 156(12), 147(48), 128(15), 113(6), 100(9), 73(100)

Inositol TMS MS m/z (rel. int.) 507(5), 432(16), 393(5), 367(9), 343(4), 318(67), 305(83), 284(11), 265(23), 217(86), 204(25), 191(41), 147(75), 129(16), 103(8), 73(100)

Malic Acid TMS MS m/z (rel. int.) 335(5), 319(4), 307(5), 265(4), 245(11), 233(23), 217(6), 189(9), 175(7), 147(80), 133(20), 117(9), 101(11), 73(100)

Phosphoric Acid, O, O, O-TMS MS m/z (rel. int.) 314(13), 299(71), 283(9), 269(5), 225(6), 211(20), 207(9), 191(11), 181(9), 165(5), 147(7), 133(17), 115(5), 103(4), 73(100)

Quinic Acid TMS MS m/z (rel. int.) 553 (13), 537(11), 419(8), 345(100), 334(6), 255(66), 204(9), 191(11), 147(25), 73(66)

Rythronic Acid TMS MS m/z (rel. int.) 462(3), 372(7), 357(8), 282(17), 254(19),
204(100), 189(17), 147(25), 73(85)

Shikimic Acid TMS MS m/z (rel. int.) 462(3), 372(7), 357(8), 282(17), 254(19),
204(100), 189(17), 147(25), 73(85)

Unknown 2 (U2) MS m/z (rel. int.) 219(2), 192(2), 191(13), 190(5), 149(34),
148(15), 147(100), 117(16), 75(10), 74(70), 73(73)

Unknown 3 (U3) MS m/z (rel. int.) 235(5), 205(3), 190(7), 149(39), 148(17),
147(100), 133(28), 131(15), 117(5), 75(13), 74(9), 73(68), 59(20)

Unknown 5 (U3) MS m/z (rel. int.) 355(2), 285(7), 229(5), 228(10), 185(3),
184(27), 184(27), 136(12), 135(6), 134(54), 110(66), 78(9), 77(76), 76(3), 74(11),
73(100), 69(63), 58(23)

Unknown 9 (U9) MS m/z (rel. int.) 247(4), 150(6), 149(35), 148(14), 147(100),
129(9), 75(18), 73(26)

Unknown 11 (U11) MS m/z (rel. int.) 429(1), 304(9), 200(30), 180(47), 152(18),
149(11), 147(18), 101(6), 100(2), 99(6), 77(13), 75(46), 73(54), 71(23), 59(28)

Unknown 12 (U12) MS m/z (rel. int.) 333(7), 305(18), 292(25), 219(6), 218(1),
217(7), 191(6), 149(15), 148(7), 147(38), 143(15), 131(15), 130(8), 129(16), 117(15),
101(3), 75(30), 73(100), 69(8)

Unknown 13 (U13) MS m/z (rel. int.) 308(2), 307(8), 277(4), 219(4), 218(10),
217(47), 189(5), 149(13), 148(4), 147(38), 129(15), 103(16), 75(8), 74(9), 73(100)

Unknown 16 (U16) MS m/z (rel. int.) 355(3), 334(7), 333(21), 293(6), 292(20),
218(6), 217(17), 208(4), 191(3), 149(13), 148(8), 147(55), 143(15), 129(19), 103(17),
75(8), 74(10), 73(100)

Unknown 18 (U18) MS m/z (rel. int.) 348(3), 347(8), 346(14), 345(31), 319(10), 256(8), 255(24), 219(9), 218(18), 217(87), 191(18), 149(13), 148(10), 147(61), 133(17), 75(12), 74(9), 73(100)

Unknown 19 (U19) MS m/z (rel. int.) 364(1), 363(4), 362(9), 361(34), 332(9), 271(9), 246(2), 245(4), 244(5), 243(11), 217(55), 170(9), 169(19), 147(28), 129(21), 103(12), 75(11), 74(9), 73(100)

Unknown 22 (U22) (MS m/z (rel. int.) 320(4), 319(15), 306(6), 305(11), 218(6), 217(16), 204(30), 180(1), 158(2), 149(12), 148(9), 147(43), 129(18), 103(5), 75(7), 74(8), 73(100)

Unknown 23 (U23) MS m/z (rel. int.) 445(7), 434(10), 432(27), 371(1), 345(9), 344(11), 343(30), 318(12), 271(5), 239(5), 217(5), 204(10), 191(42), 149(23), 148(11), 117(72), 133(11), 73(100)

Unknown 32 (U32) MS m/z (rel. int.) 436(10), 362(9), 361(27), 318(9), 307(8), 257(9), 219(9), 218(21), 217(87), 169(15), 149(21), 147(54), 129(27), 103(14), 75(6), 74(9), 100(100)

Unknown 38 (U38) MS m/z (rel. int.) 449(9), 363(5), 361(27), 326(24), 325(8), 324(35), 272(5), 271(13), 244(5), 243(14), 217(26), 204(10), 169(30), 149(13), 148(6), 147(43), 129(24), 75(11), 74(12), 73(100)

Xylose TMS MS m/z (rel. int.) 364(5), 333(4), 288(4), 263(23), 217(4), 205(43), 173(25), 147(61), 133(14), 117(27), 103(100), 73(100)

6.4 RESULTS AND DISCUSSION

Immature xylem tissue samples were collected from two of the nine full-sib loblolly pine families located in Monroe County, Alabama and Nassau County, Florida.

From the two families, two clones from each family were chosen for a further study of α -cellulose content in trees. For each family, one clone contains high α -cellulose content while the other clone contains low α -cellulose content. Clone 1 from family 35 is classified as having low α -cellulose while clone 75 from family 35 is cataloged as the clone with high α -cellulose content. The low α -cellulose content clone associated with family 34 is clone 24. Clone 65 from family 34 is the clone classified as high α -cellulose content.

Cellulose content is a phenotypic variable that is controlled by genetics and environment. An analysis of variance based on the α -cellulose content for individual trees on both sites was performed using the PROC glm statement, which is a general linear model. The results of the analysis of variance suggested that there was a significant difference in α -cellulose yield (ACY) between the two sites (Table 1). Results also indicate that there was no significant difference in ACY between the four clones or site by clone interaction. However, there was a significant difference between two clones: clone 75, a high α -cellulose clone, and clone 24, a low cellulose clone (Table 2). As a result these two clones were used for subsequent statistical analysis to identify the metabolites that are genetically related to cellulose production in loblolly pine.

Table 6.4.1. Four Clone Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	7	550.226603	78.6038	2.98	0.0078
Error	81	2136.14512	26.372162		
Corrected Total	88	2686.371723			

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Site	1	112.3728128	112.3728128	4.26	0.0422
Clone	3	177.1825735	59.0608578	2.24	0.0899
Site x Clone	3	61.8315059	20.610502	0.78	0.5077

Table 6.4.2. Least Means Squares Analysis for Clonal Effect

Clone	ACY LSMEAN	LSMEAN Number
1	45.6476365	1
24	43.6099688	2
65	46.6850479	3
75	48.8658881	4

Least Squares Means for effect Clone				
Pr > t for H0: LSMean(i)=LSMean(j)				
Dependent Variable: ACY				
i/j	1	2	3	4
1		0.2728	0.5874	0.0739
2	0.2728		0.1658	0.0139
3	0.5874	0.1658		0.3116
4	0.0739	0.0139	0.3116	

To determine the metabolites that are genetically related to cellulose production, an analysis of variance was executed on the two clones, 75 and 24, to determine the metabolites that are significantly different based on site, clone, and site by clone (Table 3). Based on the analysis of variance, there were a total of 21 metabolites that varied based on site. Out of the 21 metabolites, 12 of them were unknowns (2, 5, 9, 12, 13, 18, 19, 23, 32, and 38) which included a number of organic acids, carbohydrates, disaccharides and possible a glucoside. The known metabolites that vary by site include: phosphoric acid, malic acid, rylthronic acid, xylose, glutamine, shikimic acid, quinic acid, fructose, and coniferin.

The second portion of the analysis of variance analyzed the genetic effect on metabolite concentrations. Results from this portion of the statistical analysis, identified 13 metabolites that vary significantly between the clones (Table 3). Eight of the 13 metabolites are unknowns which include unknowns 3, 9, 11, 12, 16, 22, and 24 which include a couple of organic acids and possibly an aldonic acid. The remaining five metabolites that vary based on genetics are known metabolites: phosphoric acid, fructose, glucose, mannose, and inositol.

Table 6.4.3. Significant Metabolites Results from the 2 Clone Analysis of Variance

Metabolite	Site	Clone	Site x Clone
U 2	0.005	0.4533	0.4533
U 3	0.080	0.0071	0.3674
U 5	0.037	0.4707	0.9771
Phosphoric Acid	0.035	0.0061	0.5885
U 9	0.001	0.0117	0.3562
B- Alanine	0.294	0.0415	0.2075
Malic Acid	0.015	0.7010	0.3792
U 11	0.557	0.0456	0.4718
Rrythronic Acid	<.0001	0.5217	0.1015
U 12	0.001	0.0029	0.3820
Xylose	<.0001	0.1089	0.0032
U 13	<.0001	0.5740	0.5922
U 16	0.193	0.0447	0.0586
L-Glutamine	0.010	0.3074	0.3393
Shikimic Acid	0.033	0.3471	0.0449
U 18	0.006	0.9438	0.9438
Quinic Acid	<.0001	0.4375	0.0064
D-Fructose	0.067	0.0395	0.6768
D-Glucose	0.511	0.0076	0.7864
U 19	<.0001	0.5809	0.0084
Mannose	0.418	0.0205	0.4891
Inositol	0.362	0.0058	0.4548
U 22	0.054	0.0034	0.2156
U 23	0.010	0.2979	0.0032
U 24	0.733	0.0342	0.0772
U 25	0.004	0.2108	0.3230
U 27	0.010	0.3955	0.6325
U 32	0.003	0.0405	0.3286
U 38	0.008	0.7211	0.5904
Coniferin	0.002	0.9190	0.6787

Of the 13 metabolites that vary significantly by clone, several also vary significantly by site such as unknowns 9, 12, 32, phosphoric acid, and fructose. To establish whether these metabolites differ based on genetics and not environment, site by clone interactions were also established to investigate metabolite concentrations for each clone across both sites. Metabolites that exhibit no significant interaction or a slope change (no rank change) interaction are suggested to vary based on genetics. For example, fructose illustrates a slope change, which shows that the difference in concentration observed is due in large part to genetics (Figure 1). All five of these metabolites exhibited this slope change interaction. In the event that any of these metabolites exhibited a rank change interaction, those metabolites would have a strong variation based on environment. Results from the analysis of variance detect an accumulation of L-glutamine and an organic acid (unknown 3) in the low α -cellulose clone. In the high cellulose clone, there is a significant accumulation in alanine, phosphoric acid, inositol, free sugars, such as fructose, glucose, and mannose, and several unknowns, which include organic acids, and disaccharides. The only metabolite to be consistent in a previous cellulose study with accumulations in high α -cellulose is glucose.

In cellulose biosynthesis, UDP-glucose is produced via two different paths (Figure 2). This first pathway is the conversion of glucose-6-phosphate to UDP-glucose. Glucose serves as a precursor in the production of UDP-glucose in the first pathway, which is thought to occur in the cytoplasm (Delmer and Haigler 2002). The accumulations of glucose in the high α -cellulose clone could promote a higher production of UDP-glucose. With accumulations in inositol as well, the increased concentrations of

glucose could also be applied in the production of phytic acid, a major source of phosphorus for the tree (Figure 2).

The second pathway converts sucrose to UDP-glucose resulting in the release of fructose from sucrose (Figure 2). In the high α -cellulose clone, the higher amounts of cellulose produced are a result of the conversion of sucrose to UDP-glucose thus creating a greater accumulation of fructose. This accumulation of fructose suggests that fructose is not being recycled or slow to be recycled back to sucrose after its release by the sucrose synthase enzyme (Winter 2004). This recycling of fructose to fructose-6-phosphate for subsequent conversion to sucrose has been found to occur in cotton (Babb and Haigler 2001).

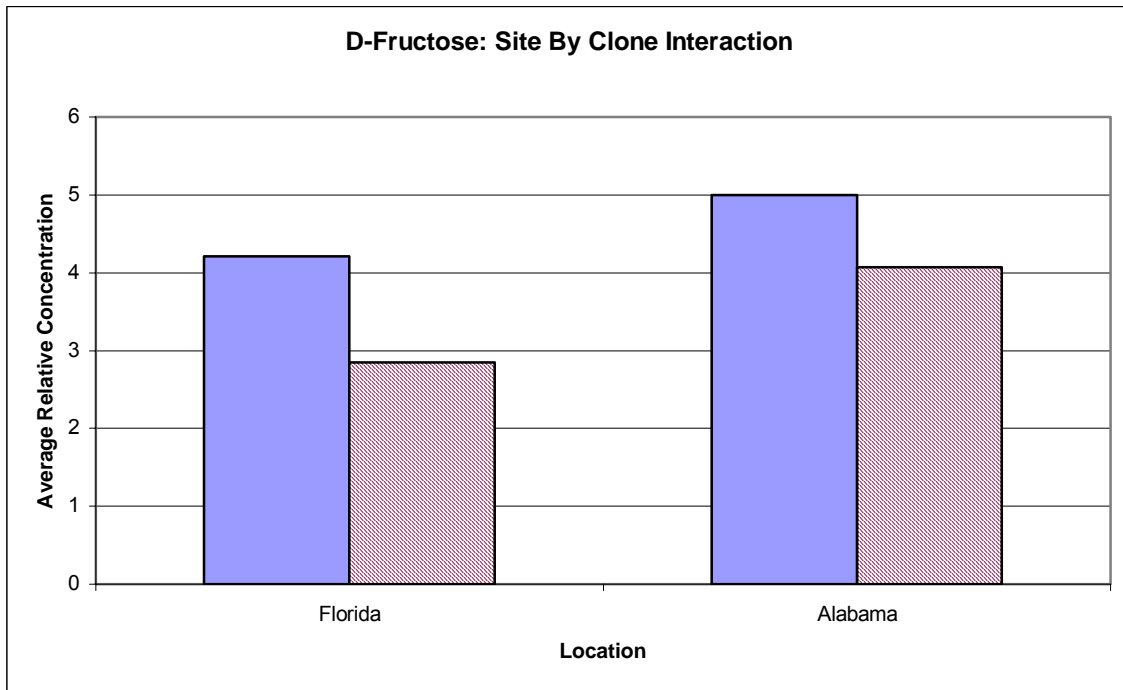


Figure 6.4.1. Site by Clone Interaction for D-Fructose. The solid bars represent clone 75 and the diagonal shaded bars represent clone 24. Results of the graph indicates that D-fructose contains a slope change interaction; thus, suggesting that the changes observed in metabolic concentration levels are due mostly to genetics.

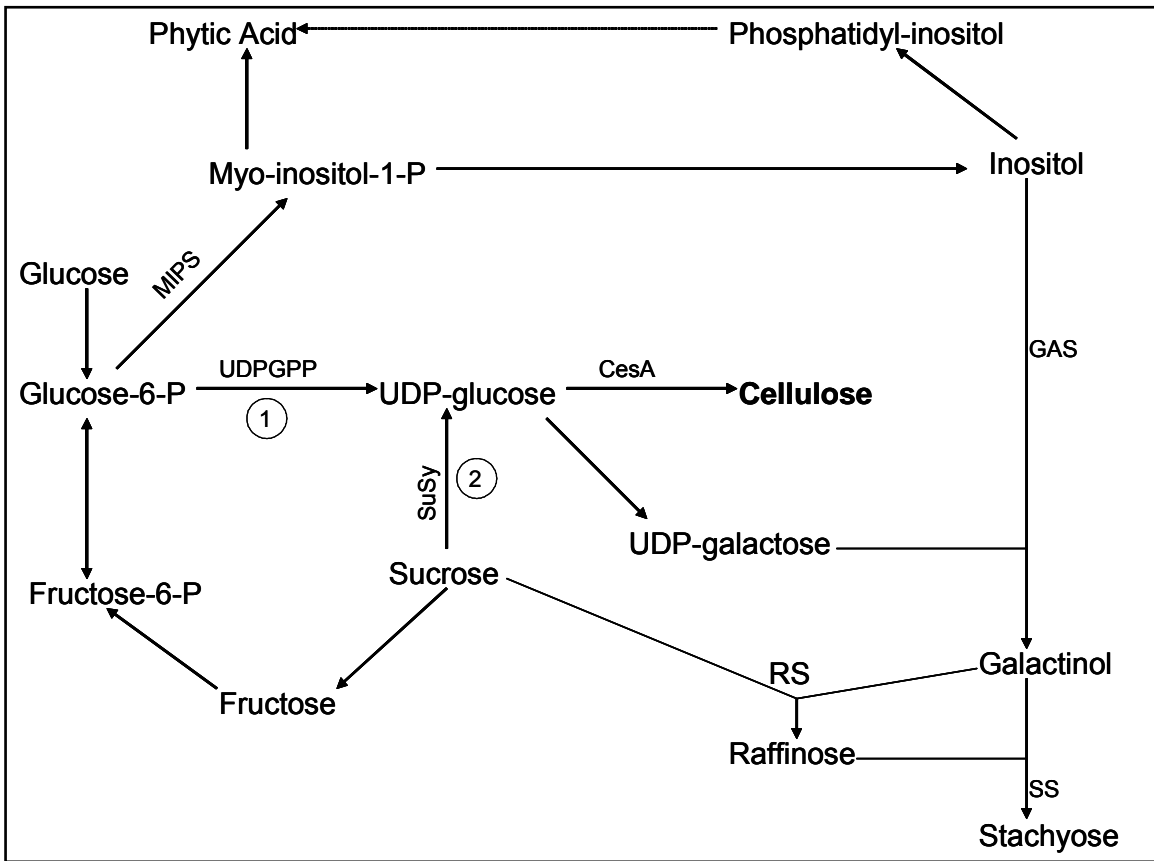


Figure 6.4.2. Cellulose Metabolic Biosynthetic Pathway (Winter 2004)

6.5 CONCLUSION

Through an analysis of variance of the two clones planted on two different sites, metabolic profiling was able to detect the genetic and environmental metabolic differences between the high and low α -cellulose clones. Metabolites such as shikimic acid and coniferin were determined to be effected greatly by environment between the two clones, which may suggest that environment may play an important role in lignin production. In both cases, clone 75 contained higher concentrations on the Florida site, but clone 24 contained higher concentrations on the Alabama site. Genetically, the low α -cellulose clone has significant accumulations in unknown 25 and glutamine. The high α -cellulose clone contains significant accumulations in alanine, phosphoric acid, fructose, glucose, mannose, inositol, and several unknowns. Glucose, fructose, and inositol are both directly associated with cellulose biosynthesis.

6.6 REFERENCES

- Babb, V. M. and C. Haigler (2001). "Sucrose Phosphate Synthase Activity Rises in Correlation with High-Rate Cellulose Synthesis in Three Heterotrophic Systems." Plant Physiology **127**: 1234-1242.
- Cano-Delgado, A., S. Penfield, et al. (2003). "Reduced cellulose synthesis invokes lignification and defense responses in *Arabidopsis thaliana*." The Plant Journal **34**: 351-362.
- Delmer, D. P. and C. H. Haigler (2002). "The Regulation of Metabolic Flux to Cellulose, a Major Sink for Carbon Plants." Metabolic Engineering **4**: 22-28.
- Haigler, C. H., M. Ivanova-Datcheva, et al. (2001). "Carbon partitioning to cellulose synthesis." Plant Mol. Biol. **47**: 29-51.
- Morris, C. R., J. T. Scott, et al. (2004). "Metabolic Profiling: a new tool in the study of wood formation." J. Agric. Food Chem. **52**(6): 1427-1434.
- Sykes, R., F. Isik, et al. (2003). "Genetic Variation of Physiochemical Wood Properties in Loblolly Pine (*Pinus taeda* L.)." TAPPI **86**(12): 3-8.
- Winter, H. (2004). Cellulose biosynthesis. C. Morris.

7. Future Work

A main component of the future work of metabolic profiling should include the addition of other analytical methods to evaluate individual metabolites and their concentrations. One main technique that has not been utilized to its fullest potential is LC/MS. This analytical technique will cut out the derivatization step required with the use of GC/MS. With its expanded mass range LC/MS will also allow better profiling results from the lipid phase by being able to analyze entire triacylglycerols. As shown in the appendix, lipid phase must undergo a transmethylation to cleave the side chains of the triacylglycerols which results in an analysis of side chains and not entire metabolites. These results can be skewed due to the fact that several metabolites in the lipid phase can contain the same side chains thus resulting in a higher concentration for that side chain and not an individual metabolite.

Therefore, a method should be carefully developed to analyze metabolites using the LC/MS system. Some of the main complications will be to develop a method that can analyze as many metabolites for an individual phase at a time. This may be somewhat limited due to the fact analysis of carbohydrates must be done using a refractive index detector, which requires isocratic conditions which may limit the separation of amino acids and other compounds in the polar phase. A solution to this problem may be to select a column that contains the best mode of separation for the classes of compounds normally found in the polar phase and rely on the TIC produced from the MS to identify metabolites by retention time and mass spectra libraries. Also, calculating the area under the curves utilizing MS software can be applied to gain relative and actual concentrations of individual metabolites.

Some recent work in the realm of metabolic profiling have included the utilization of NMR to study the differences between normal and disease states (Aranibar, Singh et al. 2001; Charlton, Allnut et al. 2004). This technique like any other metabolic profiling technique requires a great deal of data processing; however, this NMR techniques were unable to establish the identity and concentration of individual metabolites. Metabolite concentration and identity may be obtained by separating the individual metabolites prior to NMR with the use of LC. This technique will provide NMR spectra which may be useful in the elucidation of unknown metabolites.

One of the problems that statisticians have encountered is the limited number of individual samples used in the experiments. Encompassing more individuals in a study will provide much better statistical results. The additional individuals will provide further metabolic data from the sample set which may provide stronger correlations between phenotypic traits and metabolic data.

Lastly, it is very important to identify as many metabolites as possible. The identification of some of the many unknown metabolites will provide a wealth of information on the metabolic variations occurring during the developmental stages, environmental effects or transgenic applications. One method, however costly, is to run standard samples of known metabolites in various metabolic pathways to build a large and useful library. Another alternative is to separate out individual metabolites observed in samples and run various chemical analysis to elucidate its identity.

REFERENCES

Aranibar, N., B. K. Singh, et al. (2001). "Automated Mode-of Action Detection by Metabolic Profiling." Biochemical and Biophysical Research Communication **286**: 150-155.

Charlton, A., T. Allnut, et al. (2004). "NMR profiling of transgenic peas." Plant Biotechnology Journal **2**: 27-35.

8. APPENDIX

8.1 Metabolic Profiling Protocol

Samples were collected during the height of wood formation (between May and June). Xylem tissue was obtained from loblolly pine (*Pinus taeda L.*) by cutting away a small window of bark and phloem, typically 4" x 8", followed by rapid scraping of the xylem tissue into a pre-chilled sterile centrifuge tube. The tissue was then immediately immersed in liquid nitrogen and transferred to a dry ice chest for transport back to the lab. NOTE: once the tree bark has been scored, the trees defense mechanisms will be triggered and the metabolites present will begin to change, therefore, it is imperative that the tissue be collected as fast as possible, moreover, that consistent collection times be maintained during the course of repetitive sampling. Generally, the entire process takes about 30 seconds to collect enough tissue and place the samples in liquid nitrogen once the bark is removed. At the lab, the samples were placed in a -80°C freezer for storage.

300 mg \pm 30 mg of frozen xylem tissue was ground into a fine powder prior to extraction. Three grinding methods were used: a coffee bean grinder (Gloria Jean's Gourmet Coffee's/Model:202), amalgamator (Zenith/Model:Z14), and pestle and mortar. All three methods exhibited nearly identical results provided that the tissue samples remained frozen during the grinding procedure.

The ground tissue (300 mg \pm 30 mg) was mixed with 1.4-mL of 100% methanol and vortexed for 10 seconds to stop any enzymatic activity. Two internal standards were added, 100- μ L ribitol solution (10 mg/mL H₂O) and 100- μ L nonadecanoic acid methyl ester (10mg/mL CHCl₃). A 100- μ L aliquot of water was added, the mixture was vortexed, and the tissue was extracted by shaking the mixture for 30 minutes at 70°C. The sample was then centrifuged for 3 minutes at 14000g. The supernatant was removed

and diluted with 1.4-mL of H₂O in a second centrifuge tube, vortexed and set aside. The remaining centrifugate was further extracted for 10 minutes at 37°C with 2-mL of CHCl₃. The mixture was centrifuged at 14000g for 3 minutes, and the supernatant transferred to the methanol-water extract. The combined extracts were vortexed and centrifuged at 4000 rpm for 15 minutes.

The polar (upper) phase and the lipid (lower) phase of the combined extract were separated into two 7-mL glass vials. Care was taken not to transfer any of the lipid phase over with the polar phase. Both the polar and lipid phase solutions were filtered into 7-mL vials using a 3-mL syringe and 0.2 μm syringe filter. The polar phase was dried in a speed vac concentrator overnight at 25°C.

The dried polar phase was dissolved in 100-μL of methoxyamine hydrochloride (20mg/mL pyridine) and shaken for 90 minutes at 30°C. 200-μL of N-Methyl-N-(trimethylsilyl) trifluoroacetamide (MSTFA) was then added, the solution was shaken for 30 minutes at 37°C, and then left standing for 2 hours at 25°C. The mixture was then diluted with 200-μL of pyridine, vortexed and directly injected (2-μL) onto the GC/MS.

The lipid phase was transmethylated prior to GC/MS analysis. Accordingly, a 700-μL aliquot of the lipid phase was placed into a 10-mL round bottom flask. To the flask was added 900-μL of CHCl₃ and 1-mL MeOH containing 3%v/v H₂SO₄. The reaction was refluxed for 4 hours. The reaction was then extracted 2 times with water (4 mL), centrifuged at 4000 rpm to promote phase separation, and the water phase was removed. The remaining CHCl₃ phase was dried over anhydrous Na₂SO₄ and concentrated under reduced pressure to ~ 80-μL. The acidic protons were derivatized with 10-μL of MSTFA

in 10- μ L pyridine for 30 minutes at 37°C. After the 30 minutes of derivatization the sample was injected (2- μ L) onto the GC/MS.

The GC/MS system consisted of a ThermoFinnigan TraceGC and PolarisQ ion trap mass spectrometer. The system was controlled by Xcalibur software (version 1.3). GC/MS analyses were conducted using a 25 m x 0.2mm J&W Scientific INC. DB-1 column. The injection temperature was set to 230°C, the interface temperature was set to 200°C, and the ion source was set to 250°C. Helium flow was 1 ml min⁻¹. After a 5 minute solvent delay at 70°C, the oven temperature was increased at 5°C min⁻¹ to 310°C. The oven was then held isocratic for 1 minute and cooled down to 70°C. Mass spectra were recorded from m/z = 50 to 650 at 0.58 s scan⁻¹ with an electron ionization of 70 eV.

8.2 Microarray Protocol

Xylem tissue was obtained by cutting away a small window of bark and phloem, typically 4" x 8". Once the window was removed, the differentiating xylem tissue was rapidly scrapped into a pre-chilled centrifuged tube. The tissue was then immediately immersed in liquid nitrogen and transferred to a dry ice chest for transport to the lab to be stored at -80 °C. About 1g of xylem tissue from each tree was ground into powder using a pestle and mortar and liquid nitrogen. The RNA was extracted from the powder using a RNAeasy plant mini kit (Quiagen, Valencia, CA). The integrity of the RNA was determined by running it on a 1.0% agarose gel, while the purity and concentration was determined by UV absorbance at 260/280 nm.

cDNA selection

The 2,109 distinctive ESTs used for this study were acquired from cDNA libraries of *Pinus taeda* related to wood formation. The ESTs were selected closest to the 3' end of the transcript of the respective contig, based on the website provided by University of Minnesota (<http://web.ahc.umn.edu/biodata/nsfpine>). On the current array, genes are grouped into 17 different categories, as proposed for *Arabidopsis thaliana* (Kirst, Johnson et al. 2003; Stasolla, Zyl et al. 2003).

cDNA labeling and hybridization

The RNA isolated using the RNAeasy plant mini kit was amplified using a Message Amp aRNA kit by Ambion based on the RNA amplification developed by Dr. James Eberwine (Van-Gelder, Xastrow et al. 1990). The method utilizes reverse transcription with an oligo (dT) primer bearing a T7 promoter and *in vitro* transcription of

the DNA product with T7 RNA polymerase to produce a large number of antisense RNA copies of each mRNA. The next step involves the aminoallyl labeling of the RNA, using a procedure from the Institute of Genomic Research. The method reverse transcribes RNA using aminoallyl nucleotides to synthesize cDNA which can then be fluorescently labeled by coupling the aminoallyl groups to either Cyanine 3 or 5. Upon the completion of the coupling reaction, the reaction is purified using Qiagen PCR Purification Kit to remove the uncoupled dye. To begin the hybridization, 80 μ L of the labeled target in 50% deionized formamide was applied to the slides a slip cover was placed on top of the slide. The slides were placed in Corning microarray slide chambers and incubated in a water bath overnight at 42 °C. The first post hybridization wash was 4 minutes at 42 °C in 1 X SSC, 0.2% SDS. The second wash was 4 minutes at 42 °C in 0.1 X SSC, 0.2% SDS and the last wash was 4 minutes at 42 °C three times in 0.1 x SSC.

Array Slide Preparation

For slide preparation, PCR products of 2109 ESTs from *Pinus taeda* cDNA clones were printed on aminosilane-coated glass microscope slides (Corning INC.; Corning, NY, USA) using a LUCIDEA printer (Amersham, NJ). After the slides were printed and dried, the spotted cDNA was UV-crosslinked using a Stratalinker (Stratagene, LA Jolla, CA, USA), then baked for 2 hours at 75 °C and stored in the dark at room temperature until use. Details of materials and methods for array preparation have been previously described (Stasolla, Zyl et al. 2003).

Experimental Design

A balanced (equal number of individuals) reciprocal loop design was used in this experiment to compare the 14 individuals in these families (Kerr and Churchill 2001).

The balanced loop design allows for pair-wise comparisons, paring the fast growth and slow growth individuals on the same array. This maximizes the power of the comparisons. The reciprocal portion of the design helps remove dye effects. There are 28 slides in total.

For microarray experiments, the best statistical evaluation of differences in the transcript levels of genes is applied by analysis of variance (ANOVA) (Kerr, Martin et al. 2000; Wolfinger, Gibson et al. 2001). This statistical method allows for flexibility in the experimental design, and generates probability values, which express the significance levels of the differences in transcript abundance. Two linear mixed models were performed in sequence, the normalization model and the gene model. The normalization model is a global normalization process that corrects for slide effects, dye effects and slide-dye interactions (Jin, Riley et al. 2001; Wolfinger, Gibson et al. 2001). The gene model was applied separately to each gene using residuals from the normalization model to infer the transcript level for each gene (Stasolla, Bozhkov et al. 2004).

Normalization Model

$$\text{Log}_2(Y_{ijkl}) = \mu_{ij} + D_k + S_l + DS_{kl} + e_{ijkl} \quad (1)$$

Y_{ijkl} = intensity of the s^{th} spot in the l^{th} slide with the k^{th} dye applying the j^{th} treatment for the i^{th} tree

μ_{ij} , D_k , S_l , DS_{kl} = the mean effect of the j^{th} treatment in the i^{th} tree, the k^{th} dye effect, the l^{th} slide random effect and the random interaction effect of the k^{th} dye in the l^{th} slide

e_{ijkl} = stochastic error term

Gene Model

$$R_{ijkl} = \mu_{ij} + T_k + S_l + SS_{ls} + e_{ijkl} \quad (2)$$

R_{ijkl} = residual of the g^{th} gene from the normalization model

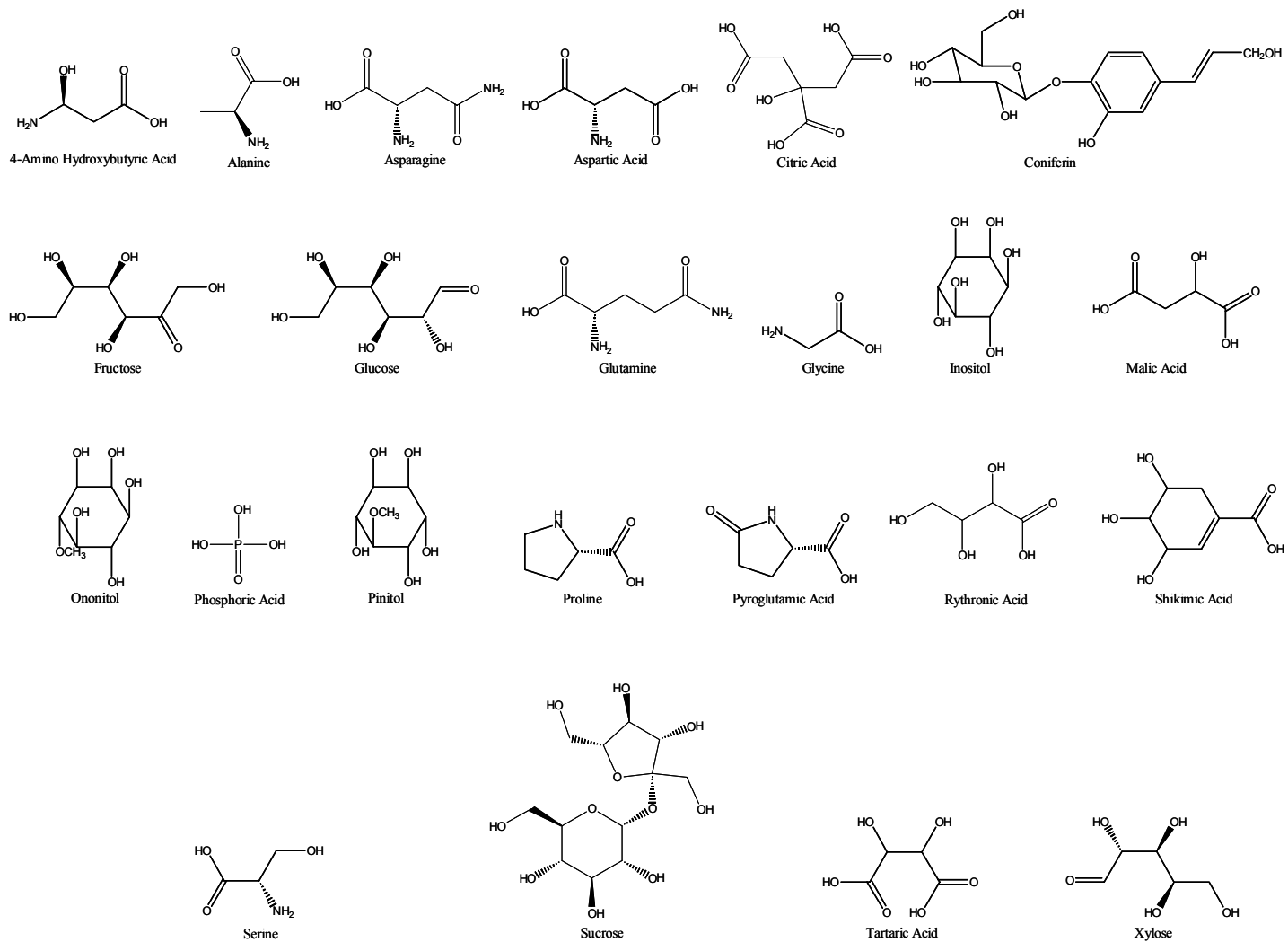
μ_{ij} , T_k , and S_l = similar roles as μ_{ij} , D_k , S_l , and DS_{kl} except they are specific for the g^{th} gene

SS_{ls} = represents the spot by slide random effect for the g^{th} gene

e_{ijkl} = stochastic error term

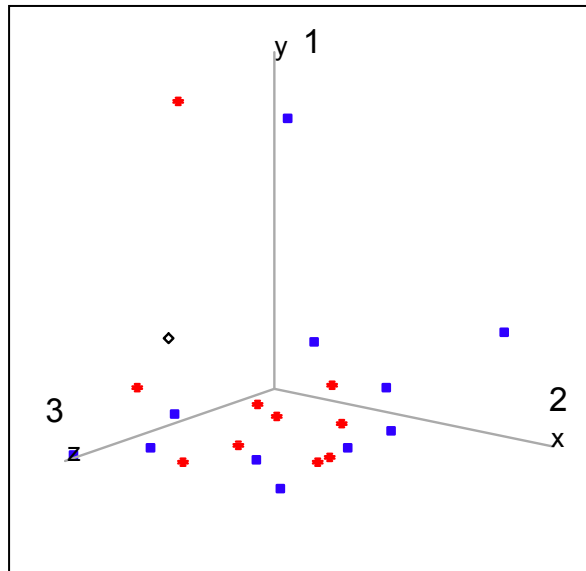
Note. All random terms are assumed to be normally distributed and mutually independent within each model.

8.3 Structure of Significant Known Metabolites



8.4 Lipid Phase of Metabolite Profiles

The lipid phase results have not been included in the dissertation chapters. Due to the lack of m/z range of the GC/MS, the lipid phase must undergo a transmethylation to cleave the side chains of the triacylglycerols extracted in the lipid phase. Transmethylation creates an analysis of side chains and not the entire metabolite in the lipid phase. Therefore, the lipid phase results could be skewed due to the fact that several metabolites in the lipid phase can contain the same side chains thus resulting in a higher concentration for that side chain and not an individual metabolite. Posted in the appendix is an example of the results achieved from the metabolic profiling of the lipid phase from the α -cellulose study in Chapter 2. Included in the appendix are the PCA results and a table of the lipid phase metabolites from the α -cellulose study.



First 3 PC explains 76.6% of the variation of lipid phase metabolites

No.	Metabolite
1	ς-Terpinen
2	Unknown 1
3	Unknown 2
4	Unknown 3
5	Unknown 4
6	Unknown 5
7	Unknown 6
8	Unknown 7
9	Unknown 8
10	Unknown 9
11	Unknown 10
12	Unknown 11
13	Unknown 12
14	Unknown 13
15	Unknown 14
16	Unknown 15
17	Unknown 16
18	Unknown 17
19	Unknown 18
20	Unknown 19
21	Unknown 20
22	Unknown 21
23	Hexadecanoic Acid ME
24	Unknown 22
25	Unknown 23
26	Unknown 24
27	Unknown 25
28	Unknown 26
29	5,8,11,14-Eicosatetraenoic Acid Ethyl Ester
30	Unknown 27
31	Trans,-9,12-Octadecadienoic Acid ME
32	9-Octadecadienoic Acid
33	Unknown 28
34	Methyl pimar-7-en-18-oate
35	Unknown 29
36	Unknown 30
37	Hexadecanoic Acid,2-[TMS oxy], ME
38	Unknown 31
39	Unknown 32
40	Unknown 33
41	1OH-18:0
42	Unknown 34
43	Unknown 35
44	Unknown 36
45	Unknown 37
46	Unknown 38
47	Octadecanoic Acid

48	Methyl (Z) - 5,11,17,17-eicosatetraenoate
49	Unknown 39
50	Unknown 40
51	Unknown 41
52	Unknown 42
53	Unknown 43
54	Unknown 44
55	Phenanthrene Carboxylic Acid
56	Unknown 45
57	Unknown 46
58	Podocarpa-8,11,13-trien-15oic acid, 13 isopropyl-,methyl ester
59	Unknown 47
60	Unknown 48
61	Unknown 49
62	Unknown 50
63	Methyl Abietate
64	Dehydroabietic Acid
65	Unknown 51
66	Unknown 52
67	Unknown 53
68	Unknown 54
69	Abietic Acid
70	Unknown 55
71	Unknown 56
72	Unknown 57
73	Unknown 58
74	Unknown 59
75	Unknown 60
76	Unknown 61
77	Unknown 62
78	Unknown 63
79	Unknown 64
80	Unknown 65
81	Unknown 66
82	Unknown 67
83	Unknown 68
84	Unknown 69
85	Unknown 70
86	Unknown 71
87	Unknown 72
88	Unknown 73
89	Unknown 74
90	Unknown 75
91	Unknown 76
92	Unknown 77
93	Unknown 78
94	Unknown 79
95	Unknown 80

96	Unknown 81
97	Unknown 82
98	Unknown 83
99	Unknown 84
100	Unknown 85
101	Unknown 86
102	Unknown 87
103	Unknown 88