University of Tennessee, Knoxville
TRACE: Tennessee Research and Creative Exchange

Doctoral Dissertations | Graduate School

12-2023

Characterization of Lignin Structural Variability and the Associated Application In Genome Wide Association Studies

Nathan D. Bryant
University of Tennessee, Knoxville, nbryant3@vols.utk.edu

Follow this and additional works at: https://trace.tennessee.edu/utk_graddiss

 Part of the Agricultural Science Commons, Biochemical and Biomolecular Engineering Commons, Biochemistry Commons, Bioresource and Agricultural Engineering Commons, Genetics Commons, Plant Biology Commons, Plant Breeding and Genetics Commons, Plant Pathology Commons, and the Wood Science and Pulp, Paper Technology Commons

Recommended Citation

Bryant, Nathan D., "Characterization of Lignin Structural Variability and the Associated Application In Genome Wide Association Studies. " PhD diss., University of Tennessee, 2023.
https://trace.tennessee.edu/utk_graddiss/9101

This Dissertation is brought to you for free and open access by the Graduate School at TRACE: Tennessee Research and Creative Exchange. It has been accepted for inclusion in Doctoral Dissertations by an authorized administrator of TRACE: Tennessee Research and Creative Exchange. For more information, please contact trace@utk.edu.

To the Graduate Council:

I am submitting herewith a dissertation written by Nathan D. Bryant entitled "Characterization of Lignin Structural Variability and the Associated Application In Genome Wide Association Studies." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Chemical Engineering.

Arthur J. Ragauskas, Major Professor

We have read this dissertation and recommend its acceptance:

Arthur Ragauskas, Mi Li, Paul Frymier, Bamin Khomami

Accepted for the Council:

Dixie L. Thompson

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

Characterization of Lignin Structural Variability and the Associated Application in Genome Wide Association Studies

A Dissertation Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

Nathan D. Bryant

December 2023

COPYRIGHT © 2023 BY NATHAN D. BRYANT
ALL RIGHTS RESERVED.

DEDICATION

To my wife, Laura, for her unwavering support that allowed me to pursue this dream.

To my son, Cameron, who ensured this dissertation was not the only source of many sleepless nights.

To my parents for their endless love, support, and encouragement over the past 34 years.

To my brothers, who ensure I remain humble.

To my beagles, Molly and Luna, who remind me to enjoy the small things in life.

ACKNOWLEDGEMENTS

I would be remiss if I did not first and foremost express my gratitude to my advisor, Dr. Arthur J. Ragauskas, for his continuous guidance and support during my PhD studies. His passion, work ethic, high expectations, and encouragement has set the standard for what it means to be an excellent researcher and is the reason this dissertation was written. I would also like to thank my doctoral committee members Dr. Bamin Khomami, Dr. Paul Frymier, and Dr. Mi Li for their support and feedback.

I feel very fortunate for having talented and knowledgeable colleagues from the Ragauskas group, both past and present, who enabled much of this work: Dr. Yunqiao Pu, Dr. Chang Geun Yoo, Dr. Xianzhi Meng, Dr. Yun-Yan Wang, Dr. Naijia Hao, Dr. Luna Liang, Dr. Samarthya Bhagia, Shuyang Zhang, Rohit Kousika, Austin Conte, Peter Karoki, and Christian Kemefa. I would also like to thank my collaborators from the Center of Bioenergy Innovation and Center for Renewable Carbon were essential in my research efforts: Dr. Timothy J. Tschaplinski, Dr. Gerald A. Tuskan, Dr. Wellington Muchero, Dr. Udaya C. Kalluri, Dr. Hari Chhetri, Dr. Chunxiao Gong, Dr. Jin Zhang, Dr. Kai Feng, Dr. Mengjun Shu, Dr. Raphael Ployet, Dr. Jin-Gui Chen, Dr. Hari Chhetri, Dr. Tom Pendergast IV, and Katrien Devos. I would also like to thank my external collaborators Dr. Jaime Barros and Rachel Weber.

ABSTRACT

Fossil fuels currently serve as the primary energy source worldwide and are a major contributor to greenhouse gases which cause global warming. Lignocellulosic biofuels have shown the potential to be carbon neutral or even carbon negative. Therefore, lignocellulosic biomass has been investigated as a promising alternative to produce biofuels and bio-based products. However, several factors have hindered widespread adoption of biofuels. The inherent variability associated with biomass is often cited as a primary concern. Additionally, lignin, a biopolymer which constitutes ~25% of the plant cell wall, notoriously hindered biological conversion of biomass to biofuel. These factors will need to be addressed to implement the biorefinery concept.

To explore these identified issues, this dissertation focuses on the characterization of lignocellulosic biomass, especially by heteronuclear single quantum coherence (HSQC) nuclear magnetic resonance (NMR). Several compounds utilized in biomass pretreatment were evaluated for potential contamination of the resulting HSQC NMR spectra. Foliage tissue, which is seldom studied by NMR, was also examined which revealed unexpected lignin structures. Further, the lignin structure of field grown *Populus* genotypes infected with *Septoria musiva* were characterized. To reveal novel genetic associations with lignin structure through a genome wide association study (GWAS), a panel of 243 switchgrass genotypes were analyzed by HSQC NMR. Similarly, a population of 409 naturally variant *Populus* genotypes were characterized by HSQC NMR to reveal novel genetic associations with lignin structure. These findings contribute to the body of knowledge of lignin structural variability in poplar and switchgrass. Findings from these studies were further used to explore factors impacting the relationship between cell wall structure and recalcitrance and *p*-hydroxybenzoate lignin levels. Natural variant genotypes with differential molybdenum content were analyzed to examine the potential effects of a gene identified by GWAS. Further, the GWAS results provide an extensive list of targets to elucidate novel genetic mechanisms underlying lignin biosynthesis to improve poplar and switchgrass as biofuel feedstocks.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

4CL	4-Coumarate-CoA ligase
ATR	Attenuated total reflection
BLINK	Bayesian information and linkage disequilibrium iteratively nested keyway
CEL	Cellulolytic enzyme lignin
CO_2	Carbon dioxide
CP/MAS	Cross-polarization / magic angle spinning
CrI	Crystallinity index
Đ	Dispersity
DBDO	Dibenzodioxocin interunit linkage
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DOE	Department of Energy
DP	Degree of polymerization
EL	Enzyme lignin
FA	Ferulate
FARMCPU	Fixed and random model circulating probability unification
FDR	False discovery rate
FEM	Fixed effect model
FP	Functional partitioning
FTIR	Fourier-transform infrared spectroscopy

G	Guaiacyl lignin unit
GHG	Greenhouse gas
GPC	Gel permeation chromatography
GRM	Genetic relationship matrix
GWAS	Genome wide association study
H	*p*-Hydroxyphenyl lignin unit
H_2SO_4	Sulfuric acid
HPLC	High performance liquid chromatography
HPIC	High performance ion chromatography
HSQC	^{1}H-^{13}C Heteronuclear single quantum coherence
$I_{\alpha/\beta}$	Cinnamyl alcohol end unit
IEA	International Energy Agency
$J_{\alpha/\beta}$	Cinnamyl aldehyde end unit
LCC	Lignin-carbohydrate complex
LD	Linkage disequilibrium
MAD	Median absolute deviation
MAF	Minor allele frequency
MLM	Mixed linear model
MLMM	Multi-locus mixed model
M_n	Number average molecular weight
Mo	Molybdenum

MOT1	Molybdate transporter 1
M_w	Weight average molecular weight
NaOH	Sodium hydroxide
NMR	Nuclear magnetic resonance
NREL	National Renewable Energy Laboratory
PAL	Phenylalanine ammonium-lyase
PB	*p*-Hydroxybenzoate
*p*CA	*p*-Coumarate
PCA	Principal component analysis
PCC/CC	Pearson correlation coefficient
PTFE	Polytetrafluoroethylene (Teflon)
Q-Q	Quantile-quantile
QTN	Quantitative trait nucleotide
RNAseq	Ribonucleic acid sequence
RWR	Random walk with restart
S	Syringyl lignin unit
S/G	Ratio of syringyl (S) units to guaiacyl (G) units
SEC	Size exclusion chromatography
SNP	Single nucleotide polymorphism
T	Tricin
TF	Transcription factor

THF	Tetrahydrofuran
UV-VIS	Ultraviolet-visible spectroscopy
WT	Wild type

Chapter 1:
Introduction

The rapid growth in human population and improved standard of living has led to a dramatic increase in energy needs [1]. Today, fossil fuels serve as the primary global energy source and account for ~80% of total worldwide energy consumption [2]. The transportation sector alone accounts for 58% of this consumption while simultaneously producing 25% of all CO_2 emissions [3]. Additionally, greenhouse gas (GHG) emissions from the transportation sector have been increasing at a faster rate than any other sector [4]. The International Energy Agency (IEA) estimates that the worldwide energy demand from fossil fuels will increase from 13.4 billion tonnes of oil equivalent in 2012 to 18.3 billion tonnes by 2035. Fossil fuels are nonrenewable resources that produce GHG. Therefore, replacing fossil fuels with renewable alternatives has attracted significant attention to curb GHG emissions and limit the effects of global warming. Biofuels are one of the few resources with the potential to replace conventional fossil fuel energy sources [1].

Biofuels are a promising alternative to fossil fuels to meet transportation energy needs. Although biofuels already account for 40% of all renewable energy consumption, they currently account for only 3.4% of transportation fuels [3]. The production of bioethanol increased by 67% between 2008-2018, and biodiesel production increased threefold over the same timeframe [5]. Indeed, the IEA estimates that as much as one-third of all transportation fuel could come from biofuels by 2050. While electric vehicles are expected to dominate the consumer vehicle market, biofuels are poised to address transportation needs through other applications. One such application is sustainable aviation fuel [5]. The aviation industry generates 2.5% of all CO_2 emissions, and current fuel demand is expected to double by 2050 [6]. Another potential application of biofuels is in the maritime sector, which emits 2.6% of worldwide GHG emissions. In a recent promising endeavor, ExxonMobil completed a sea trial with marine biofuel which showed that CO_2 emissions could be reduced by up to 40% [6]. Perhaps more importantly, this biofuel can be utilized in relevant maritime applications without modification, acting as a nearly “drop in” replacement and allowing companies to make substantial progress toward their carbon reduction goals.

To facilitate the implementation of biofuels, the concept of the biorefinery has been described [7]. The biorefinery parallels the traditional oil refinery, with raw lignocellulosic biomass being fractionated and converted into an array of valuable products. A key issue in addressing the biorefinery vision is the need to effectively utilize low-cost, high-volume biomass feedstocks with variable levels of cellulose, hemicellulose, and lignin. Indeed, feedstock variability has been identified as a major barrier for the implementation of biorefineries [8]. Therefore, understanding biomass variation will be a key technology in this endeavor. This applies especially to lignin, a complex biopolymer which comprises ~25% of the secondary cell wall. While several factors influencing biomass conversion have been documented [9], studies have repeatedly shown that lignin is a major barrier in converting biomass to biofuels [10]. Along with the emergence of "lignin first biorefining", which considers the utilization of lignin at the design stage [11], understanding lignin structure and the factors that impact lignification are becoming increasingly important.

The overarching theme of this dissertation is the characterization of lignin. The U.S. Department Energy has identified biomass feedstock variability, including lignin, to have a significant impact on biofuel production. [12]. Therefore, several studies were performed to investigate the structural variability of lignin:

- As pretreatment is typically required in biorefineries to fractionate raw biomass, various chemicals used for pretreatments were evaluated for the potential to contaminate analytical spectra (Chapter 4).
- Field grown *Populus* trees infected with *Septoria* were thoroughly characterized to decipher changes to lignin structure induced by the fungal infection (Chapter 5).
- Leaf tissue, which is less well studied than wood tissue, was explored to reveal high levels of lignin structural variability and structures not typically observed in stem lignin (Chapter 6).
- Extensive deep phenotyping was applied to a population of natural variant *Populus*, and results were utilized in a genome wide association study (GWAS) to elucidate genetic associations with lignification (Chapter 7).

- Similarly, a panel of natural variant switchgrass was phenotyped to reveal novel candidate genes potentially involved in lignin and/or cell wall biosynthesis (Chapter 8).
- The results from the poplar GWAS lead to further investigation of select genotypes with specific variations in lignin structure (Chapter 9).

These results shed new light on lignin structures that can aid in the development of the fundamental understanding of lignification, disease resistance, and developing strategies to control lignin structure for the benefit of biorefineries.

Chapter 2:
Literature review

Sections of this chapter were originally published as:

Bryant, Pu, Y., Tschaplinski, T. J., Tuskan, G. A., Muchero, W., Kalluri, U. C., Yoo, C. G., & Ragauskas, A. J. (2020). Transgenic Poplar Designed for Biofuels. Trends in Plant Science, 25(9), 881–896.

Arthur J. Ragauskas conceived the topic of this review. Nathan Bryant, Wellington Muchero, Udaya Kalluri, and Chang Geun Yoo drafted the manuscript. All authors reviewed, revised, and approved the manuscript. Elsevier publishes under an open access license, meaning that authors retain rights and may be reprinted with citation of the original publication.

2.1 Lignocellulosic biomass

Early biofuels, referred to as first generation biofuels, were derived from crops such as corn. However, research estimated that feedstocks used for first generation biofuel production could be used to potentially feed upwards of 200 million people [13]. This has sparked a "food vs. fuel" debate and prompted the evaluation of other feedstocks to be used for biofuels. Second generation biofuels are therefore derived from lignocellulosic feedstocks. Lignocellulosic biomass is described as having three main components: lignin, cellulose, and hemicellulose, depicted in Figure 2-1. The composition of these components will vary based on factors such as plant species and environmental conditions.

2.2 Cellulose

Cellulose was originally discovered in 1839 by French chemist Anselme Payen, who showed that plants yielded a resistant fibrous substance when purified by an acid-ammonia treatment followed by extraction with water, alcohol, and ether [14]. Cellulose is the most abundant terrestrial biopolymer and is the main component of the secondary cell wall. Though cellulose can be converted to biofuels, it has excellent properties that allow it to be used for applications such as textiles, films, and paper. Wood typically contains 40-50 wt% cellulose [15]. Regardless of origin, cellulose is a linear polymer comprised of β-(1,4) linked β-D-glucopyranose units (Figure 2-2) with a degree of polymerization (DP) of 7,000-15,000 [16]. Groups of 36 cellulose

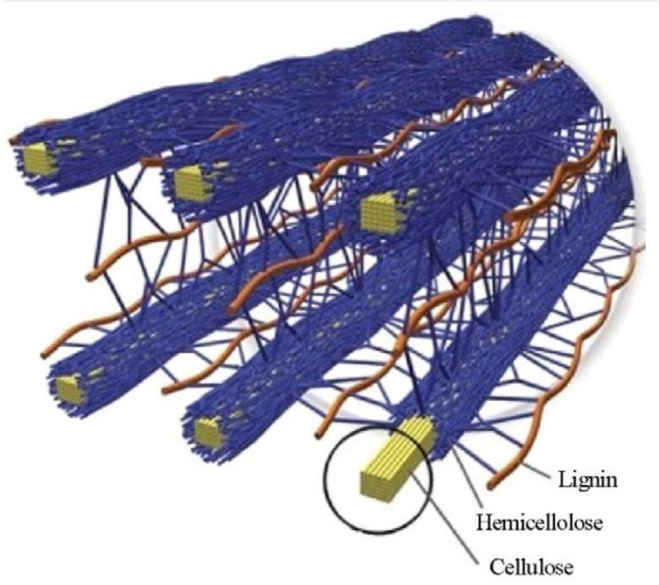


FIGURE 2-1 DEPICTION OF CELLULOSE, HEMICELLULOSE, AND LIGNIN ARRANGEMENT IN THE SECONDARY CELL WALL

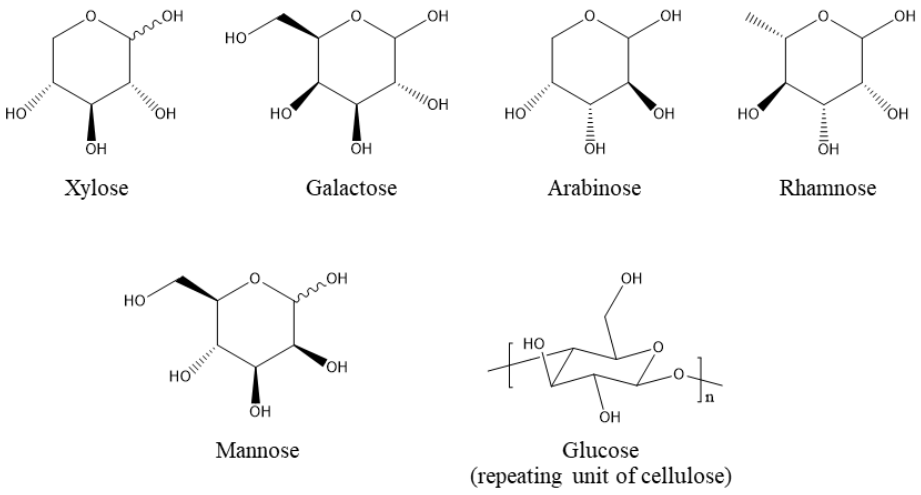


FIGURE 2-2 STRUCTURES OF THE PRIMARY MONOMERIC SUGARS THAT COMPRISE CELLULOSE AND HEMICELLULOSE

chains organize in parallel assemblies called microfibrils, and multiple microfibrils are further assembled into groups called fibrils [17]. Each β-D-glucopyranose unit contains three hydroxyl groups which participate in intra- and inter-molecular hydrogen bonding to give cellulose a partially crystalline structure, represented by the crystallinity index (CrI) [18]. The crystallinity of cellulose is considered to be a major factor impacting conversion of biomass to biofuels and bio-based products [19]. Generally, cellulose with higher CrI and DP tend to be more resistant to enzymatic hydrolysis, but demonstrate improved tensile properties [18].

2.3 Hemicellulose

Hemicellulose refers to a variety of five and six carbon sugars such as xylose, arabinose, mannose, glucose, and galactose, along with their associated acids. The structures of these primary monomeric sugars are displayed in Figure 2-2. Other sugars such as rhamnose and fucose, along with acetylation of the hemicellulose backbone, are also present [20]. Constituting 10-40% of the secondary cell wall, hemicellulose plays an essential role in plant growth and development [21, 22]. Hemicellulose is connected to cellulose by hydrogen bonds and to lignin through covalent bonds. The composition of hemicellulose can vary greatly between different biomass types. However, xylan is the most prominent hemicellulose in both monocots and dicots [23]. Xylans are heteropolysaccharides with a backbone of 1,4-linked β-D-xylopyranose units with a variety of side chains [24]. Dicots contain glucuronoxylans, which are xylans decorated with acids such as α(1→2)-linked glucuronic acid and 4-*O*-methyl-glucuronic acid [25]. Monocots contain additional glucuronoarabinoxylans such as α(1→3) and α(1→2)-linked arabinofuranosyl [25]. Other small groups such as acetyl, phenolic acids, ferulic acids, and coumaric acids can decorate either the backbone or side chains of hemicellolose [24]. Hemicellulose contributes to biomass recalcitrance due to its structural complexity, interaction with other cell wall polymers, and interference with enzyme accessibility [26]. However, since hemicellulose is highly branched, amorphous, and has a low degree of polymerization, it can be easier to utilize than cellulose and lignin [27].

2.4 Lignin

Lignin is the second most abundant terrestrial biopolymer (behind only cellulose) [28].

It is comprised of the three primary phenylpropane units: sinapyl alcohol, coniferyl alcohol, and *p*-coumaryl alcohol. These primary alcohols, often referred to as monolignols, are synthesized in the cytosol and then transported to the secondary cell wall [29]. There, laccases and/or peroxidases catalyze their incorporation into the lignin polymer where they are denoted as syringyl (S), guaiacyl (G), and *p*-hydroxyphenyl (H) units, respectively [29]. Monolignols are covalently connected by various interunit linkages. These include β-O-4 ether, β-β resinol, and β-5 phenylcoumaran linkages. These monolignol and interunit linkage structures are displayed in Figure 2-3.

2.5 Lignin biosynthesis

Through photosynthesis, plants fix carbon from atmospheric carbon dioxide and distribute it through a variety of metabolic pathways. One of these is the shikimate pathway which produces amino acids. This is considered the entryway to the production of secondary metabolites such as lignin [30]. Specially, the phenylpropanoid pathway begins with the amino acid phenylalanine [31]. From here, a variety of enzymes catalyze various reactions which produce three primary hydroxycinnamyl alcohol monomers, sometimes referred to as monolignols [31]. These are termed *p*-coumaryl alcohol, coniferyl alcohol, and sinapyl alcohol, and differ only in their degree of methoxylation of the aromatic ring [31]. Though it is known that these monolignols are synthesized in the cytosol, there is debate over the next steps in the lignification process [32]. Monolignols are translocated from the cytosol to the cell wall [33]. Once in the cell wall, monolignols are incorporated into the lignin polymer through dehydrogenative polymerization [28]. Specifically, enzymes such as peroxidases and laccases catalyze the oxidation of the monolignol, which is then incorporated into the lignin polymer by combinatorial radical coupling [33]. Once incorporated into the lignin polymer, the monolignols *p*-coumaryl, coniferyl, and sinapyl alcohol become *p*-hydroxyphenyl (H), guaiacyl (G), and syringyl (S) units, respectively. While monolignols can be coupled in several different ways, monolignol radicals tend to couple at the β carbon, giving rise to the presence of linkages such as β-O-4, β- β, and β-5 (though other linkage types certainly exist) [28].

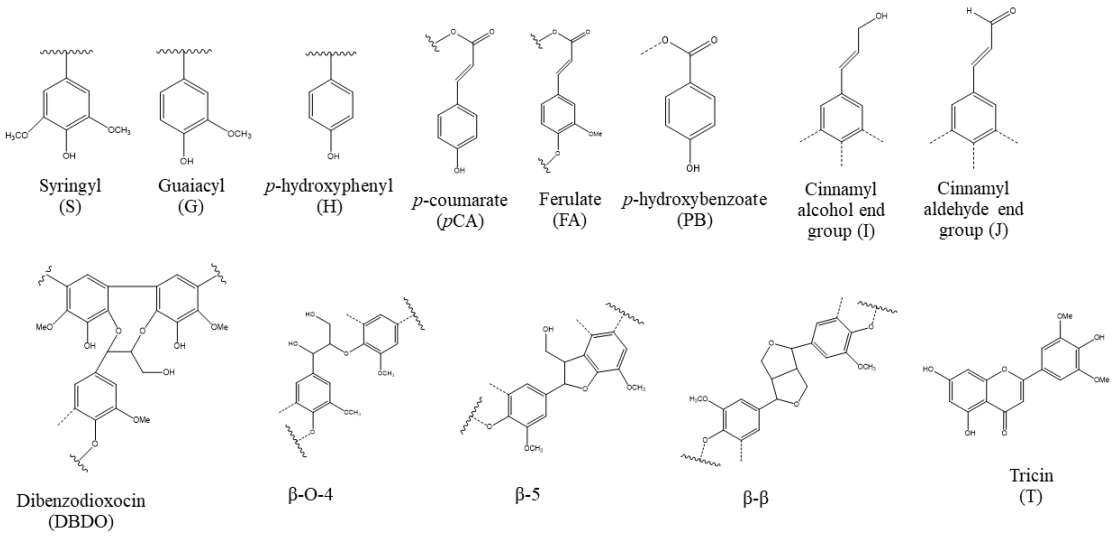


FIGURE 2-3 LIGNIN SUBUNITS AND INTERUNIT LINKAGES

The lignin biosynthesis pathway, displayed in Figure 2-4, can involve as any as 24 metabolites and enzymes from up to 11 different families [34]. The lignin structure produced is a direct result of the regulation/expression of these enzymes. However, these enzymes are far from independent. Indeed, the formation of the secondary cell wall requires the coordination of many metabolic pathways [35]. Certain genes even act as master switches, regulating the expression of multiple genes [35]. In recent years, many studies have made the effort to mis-regulate the expression of these enzymes to determine their biological function. This was spearheaded by Hu et al., when it was demonstrated that antisense inhibition of 4-coumarate:coenzyme A ligase (4CL) resulted in enhanced growth in *Populus tremuloides*. The transgenic plants exhibited a 45% reduction in lignin and 15% increase in cellulose [36]. Such secondary cell wall modifications have obvious potential benefits for conversion to biofuels. However, lignin is a vital component in plant growth and development, and shortcomings of low lignin plants have been well documented [37]. Therefore, manipulating the structural composition (rather than abundance) of lignin is often of great interest. A summary of lignin biosynthesis gene manipulation and their observed effects are summarized in Table 2-1.

2.6 Lignin variability

Lignin structure exhibits extensive variability, even within similar feedstocks. Lignin is often characterized by the ratio of S units to G units (S/G ratio). Softwood lignin is predominantly G units with few S or H units. Hardwood lignin is predominantly S and G units, present at a ~ 2:1 ratio, with minimal (<1%) H units. Herbaceous lignin typically has a ~2:1 ratio of G units to S units, and ~5% H units. An overview of monolignol composition among various lignocellulosic feedstocks is summarized in Table 2-2 [38]. The ratios of these monolignols can be influenced by factors such as biotic and abiotic stress since the phenylpropanoid pathway is involved in defense response through the production of other secondary metabolites such as flavonoids.

These monolignols are bound together through a variety of interunit linkages. The types of interunit linkages formed depend largely on the ratio of monolignols [32]. For instance, S units are methoxylated at the 3- and 5-positions, hindering the formation of bonds at these locations.

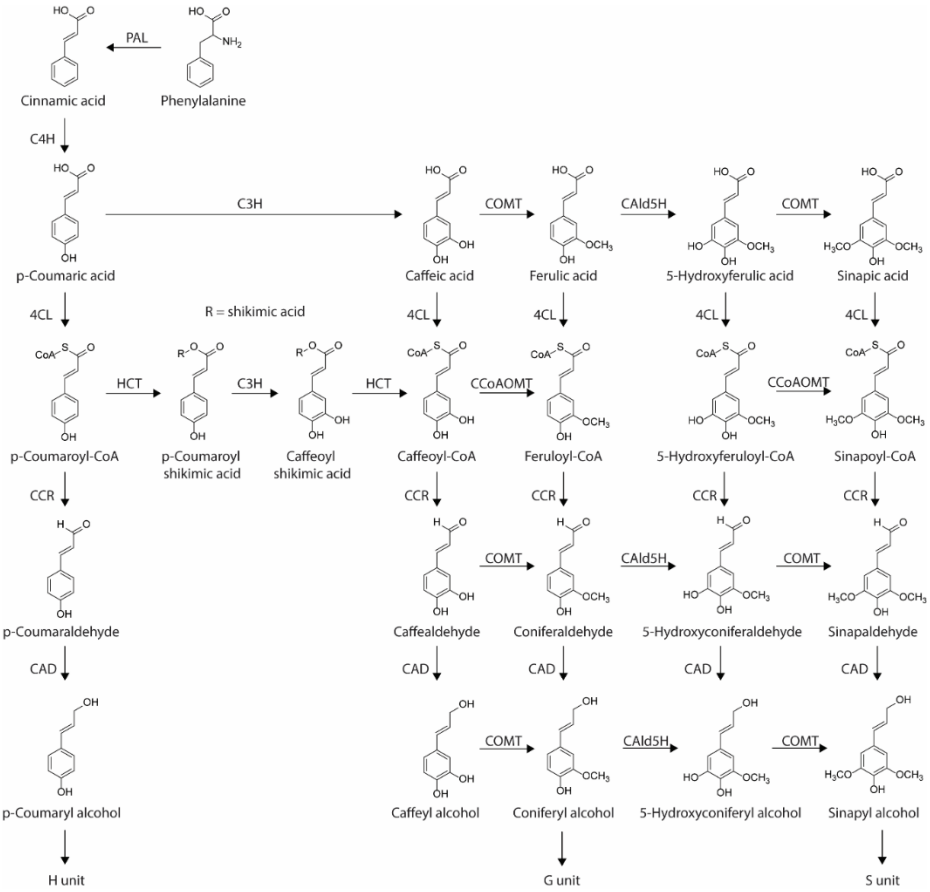


FIGURE 2-4 LIGNIN BIOSYNTHESIS PATHWAY

TABLE 2-1 EFFECTS OF GENE MANIPULATION ON LIGNIN STRUCTURE

Gene	Up/Down Regulated	Expression Level	Effect on Lignin	Effect on Cellulose/ Hemicellulose	Effect on Saccharification	Refs.
4CL	Down	-	Content reduced from 24.1% to 22.2%	Cellulose content increased from 47.9% to 51.5%	No significant change in glucose or xylose release after sodium hydroxide pretreatment	[39]
4CL	Down	22% - 64% of control for 4CL1-1 and 45% - 97% of control for 4CL1-2	Content reduced from ~22% to ~20% S/G ratio varied from 1.4 to 2.2 vs. 1.8 for control	Content varied from 40.4% to 49.3% compared to 42.2% for control	No significant improvement in sugar release after water pretreatment	[40]
4CL	Down	-	Greenhouse grown plants had reduced lignin content from 22.0% to 11.6% Field-grown plants had lignin content comparable to wild type	Most field-grown transgenic plants had glucan/xylan content within ±2% of wild type	Sugar yield of 2-year field grown transgenics increased from 39.8% to 49.3% (biomass weight basis) after green liquor pretreatment At comparable lignin content (~22%), sugar recovery was higher in transgenics (64.4-85.6%) than wild type (64.0%) after green liquor pretreatment	[41]
4CL	Down	-	Content reduced from 22% to 10.5% S/V ratio varied from 1.4-2.6 (vs. 1.7 in wild type)	Glucan content increased from 43.5% to 51.4% Xylan content increased from 15.2% to 19.0%	Saccharified sugar (mass %) increased from ~22% to ~30% after green liquor pretreatment	[42]
4CL	Down	-	Content reduced from 21.4% to 15%	Increased glucan content from 45.2% to 47.2%	Total sugar production increased from ~30% to ~47% (biomass weight basis) after green liquor pretreatment	[43]
4CL	Down	-	Greenhouse grown plants had reduced lignin content from 22.0% to 11.8%	-	2-year field grown transgenics had sugar yield of ~35.0-42.5% vs. ~37.5% (biomass weight basis) in wild type	[44]
4CL	Down	>90% reduction in xylem Pt4CL1	Content reduced from 21.6% to 11.8%	Increased cellulose content from 44.2% to 50.8%	-	[36]
4CL / CAld5H	Down/Up	-	Content ranged from 20.1%-32.7% (vs. 22.5%-26.7% in wild type)		2-year field grown transgenics had sugar yield of ~45.0% vs. ~37.5% (biomass weight basis) in wild type	[44]
4CL / CAld5H	Down/Up	-	Content reduced from 21.4% to 19.3%	Increased glucan content from 45.2% to 46.6%	Total sugar production increased from ~20% to ~26% (biomass weight basis) after ozone pretreatment	[43]

TABLE 2-1 CONTINUED

Gene	Up/Down Regulated	Expression Level	Effect on Lignin	Effect on Cellulose/ Hemicellulose	Effect on Saccharification	Refs.
4CL / CAld5H	Down/Up	-	Lignin content reduced from 22.0% to 11.6%	Glucan content increased from 43.5% to 49.5%	Mass % of sugar saccharified increased from ~22% to ~35% after green liquor pretreatment	[42]
			S/V ratio varied from 1.6-2.1 (vs. 1.7 in wild type)	Xylan content increased from 15.2% to 18.1%		
4CL & CAld5H1	Down	-	Lignin content reduced from 22.0% to 17.2%	Glucan content increased from 43.5% to 50.0%	Mass % of sugar saccharified increased from ~22% to ~26% after green liquor pretreatment	[42]
			S/V ratio unchanged	Glucan content increased from 43.5% to 45.4%		
CAld5H1	Down	-	Syringaldehyde/ Vanillan ratio decreased from 1.7 to 0.1	Less than 2% change in glucan and xylan content	Mass % of sugar saccharified comparable to wild type	[42]
	Up		Syringaldehyde/ Vanillan ratio increased from 1.7 to 2.4	Glucan content increased from 43.5% to 46.2%	Mass % of sugar saccharified increased from ~22% to ~26% after green liquor pretreatment	
CCoAOMT	Down	-	Content reduced from 24.1% to 21.6%	Cellulose content increased from 47.9% to 51.8%	Increased glucose + xylose release from ~35% in control to ~45% (biomass weight basis) after sodium hydroxide pretreatment	[39]
					Sugar release negatively correlated with soluble lignin content	
C3H	Down	Reduced as low as 5% of that in WT	Content reduced from 22.5% to 10%	-	-	[45]
			S/G ratio increased from 1.7 to 3.6	-	-	
			H units increased from 0.3% to 31%	-	-	
C3H	Down	54% expression of non-modified poplar	Content reduced from 21.1% to 14.8%	Cellulose content increased from 43.1% to 47.7%	-	[46]
				Cellulose crystallinity increased from 34.3% to 40.7%	-	
				Hemicellulose content decreased from 31.2% to 29.2%	-	

TABLE 2-1 CONTINUED

Gene	Up/Down Regulated	Expression Level	Effect on Lignin	Effect on Cellulose/ Hemicellulose	Effect on Saccharification	Refs.
PtrC3H3	Down	-	Decreased lignin content from 23.2% to 9.9%	Increased glucose content from 45.9% to 53.6%	-	[47]
			Increased S/G ratio from 2.67 to 9.9	Increased xylose content from 16.3% to 20.0%	-	
PtrC3H3/ C4H1&2	Down	-	Decreased lignin content from 23.2% to 11.7%	Increased glucose content from 45.9% to 52.1%	-	
			Increased S/G ratio from 2.67 to 2.77	Increased xylose content from 16.3% to 18.1%	-	
PtrCAD2	Down	-	Slightly increased lignin content from 23.2% to 23.5%	Decreased glucose content from 45.9% to 42.7%	-	
			Decreased S/G ratio from 2.67 to 1.86	Decreased xylose content from 16.3% to 15.6%	-	
CAD	Down	5% residual transcript level	Decreased content from 194.7 to 175.0 [mg/g CW]	No significant change in cellulose content	Increased cellulose conversion from ~35% to ~60% [% cellulose] after alkaline pretreatment	[48]
CAD	Down	-	S/G ratio increased by 17%	Unaltered structure	-	[49]
CSE	Down	Relative expression ~0.2-0.5	Increased H units from ~0.6% to ~1.2%	Increased cellulose content from ~37% to ~43%	Cellulose to glucose conversion improves from 30-35% to 45-50% after acid and alkaline pretreatment	[50]
			S/G ratio decreased from 2.06 to 1.9			
			Decreased lignin content from ~22% to ~17%			
CCR	Down	-	Decreased content from 21.7% to 16.6%	Increased cellulose content from 523.5 [µg/m CWR] to 560.5 [µg/mg CWR]	Ethanol yield increased from ~ 3 [g/L] to ~8 [g/L] during simultaneous saccharification and fermentation	[51]
HCT	Down	37% expression of non-modified poplar	Content reduced from 21.1% to 16.8%	Cellulose content unchanged	-	[46]
				Cellulose crystallinity increased from 34.3% to 39.9%	-	[46]
				Hemicellulose content increased from 31.2% to 33.8%	-	[46]

Therefore, S-rich lignin tends to form more β-O-4 ether bonds and have a more linear structure [52]. Lignin that contains more G or H units, with the 5-position and/or 3-position unimpeded, tends to form more condensed linkages (5-5, β-5, etc.). Unique interunit linkages can also arise in certain lignin types. For instance, *Populus* lignin exhibits the eight-member ring branching connection dibenzodioxocin (DBDO), whereas switchgrass exhibits a β-1/α-O-α linkage configuration referred to as spirodienone. Interunit linkages are of interest in engineered plants, as evidenced by the incorporation of chemically liable ester linkages into the lignin backbone in "zip lignin" to improve deconstruction under alkaline treatment conditions [53].

Certain types or species of plants can also incorporate non-canonical lignin structure(s). For instance, poplar produces a *p*-hydroxybenzoate (PB) unit, which is ester linked to the γ-position of S units. Herbaceous species, such as switchgrass, produce the analogous ester units *p*-coumarate (*p*CA) and ferulate (FA). These non-canonical units tend to be less well studied. For instance, new acyltransferase genes associated with PB production were recently discovered [54, 55]. These ester units, along with other hydroxycinnamates, can alter the oxidative potential of monolignols, thereby influencing lignification [56]. Switchgrass even incorporates tricin (T) into the lignin polymer, despite it not being produced by the lignin biosynthesis pathway.

2.7 Lignin variation by biomass feedstock

Lignin can vary drastically depending on which plant produces it. For classification purposes, feedstocks are typically described as hardwood, softwood, or herbaceous. This dissertation focuses on hardwood (poplar) and herbaceous switchgrass. Therefore, lignin variation from biomass feedstocks will focus primarily on these, though additional variability would be observed from a softwood feedstock such as spruce.

The mass fraction of lignin in the secondary cell wall will differ based on biomass feedstock. For instance, lignin content is reported in the range of 21-29% for poplar [57], but 16-25% for switchgrass [58]. Lignin will also vary in its monolignol composition, especially in the ratio of S units to G units (S/G ratio). The S/G ratio of poplar is typically reported to be 1.3-2.7 [57] implying that S units are present at twice the abundance of G units. By contrast, the S/G ratio of switchgrass is 0.4-0.8 [59], indicating a more G-rich lignin. Lignin from different feedstocks can

also be differentiated by the associated interunit linkages. While both poplar and switchgrass lignin exhibit the β-O-4 linkage as the predominant connection, the S-rich lignin of poplar gives rise to a higher abundance of β-O-4 linkages (~62%) [60] than the G-rich switchgrass lignin (~50%) [59]. This non-condensed ether linkage is the easiest to cleave during pretreatment, and therefore β-O-4 linkage abundance influences delignification. Poplar lignin also forms a β-1/α-O-α linkage that is not observed in switchgrass [61]. In the same manner, switchgrass forms an eight-member ring linkage (dibenzodioxocin) that is not observed in poplar [62]. However, perhaps the most distinguishing feature of lignin is how it is acylated. Poplar is uniquely acylated by *p*-hydroxybenzoate (PB) [63], an ester-linked moiety present in only a few other species. Switchgrass, as well as other herbaceous plants, is acylated by the ester-linked *p*-coumarate [59]. Another ester-linked unit, ferulate, also acylates switchgrass lignin and will be discussed further in the next section. Lignin is also incredibly plastic and can incorporate non-canonical monomers. This is especially demonstrated in switchgrass lignin, which readily incorporates the flavonoid tricin which acts as a major branching point [64]. Lignin can also vary in polymer-associated traits such as molecular weight. Molecular weight measurements are often difficult to compare due to the preparation steps involved which usually alter the lignin in some manner. Still, the weight average molecular weight of poplar is reported to be ~8,500 g/mol for milled wood lignin and ~3,200 for cellulolytic enzyme lignin [65]. By contrast, switchgrass milled wood lignin has a weight average molecular weight of 5,300-5,500 g/mol [65].

2.8 Lignin and the lignin carbohydrate complex

The characterization of lignin is important for numerous reasons. For instance, it has been well documented that the structure of lignin can have a great impact on recalcitrance (i.e., conversion of biomass to biofuel). However, it is important to recognize that lignin does not exist in a vacuum. One reason that biomass is recalcitrant to deconstruction is that lignin are carbohydrate moieties are chemically bound, forming a lignin-carbohydrate complex (LCC). The LCC greatly influences cell wall structure, as 47-66% of hardwood lignin moieties are bound to carbohydrates [66]. There are eight types of lignin-carbohydrate linkages: benzyl ether, benzyl ester, glycosidic, phenyl glycosidic, acetal, hemiacetal, and ferulate or diferulate esters [67]. Most of these arise through linkages generated from hydroxyl groups from either lignin or carbohydrates.

Herbaceous biomass such as switchgrass exhibits significantly higher amounts of LCC linkages than softwood or hardwood species [68]. A major constituent of LCC linkages in herbaceous biomass are the ferulate and diferulate esters, which link the carboxylic acid groups of lignin to hydroxyl groups of carbohydrates [69]. The ester bond formed by ferulates is easily cleaved by mild alkaline hydrolysis. The most prevalent LCC linkage in hardwoods is the phenyl glycoside (PhyGlc) bond. Unlike ferulate ester bonds, PhyGlc bonds are alkaline-stable. As a result, hardwoods with the highest amount of PhyGlc linkages exhibit the lowest kraft pulping performance [70]. However, PhyGlc linkages are exceptionally liable under acidic conditions, as are benzyl ether and ester bonds. Regardless of the specific biomass species, lignin, and its participation in the LCC adds another layer of complexity to the cell wall structure.

2.9 Lignin impact on recalcitrance

This work focuses on characterizing and understanding lignin structure. The goal is that these findings will lead to the ability to selectively control the lignin structure that plants produce. This can be utilized in two advantageous ways. First, the lignin structure could be manipulated to facilitate improved deconstruction of the cell wall. Though many factors impact recalcitrance, generally lignin that exhibits higher S/G ratio and β-O-4 linkage abundance is more amenable to deconstruction. A lignin structure more amenable to deconstruction would reduce the chemical and energy demands required for pretreatment, which has major cost implications that could positively impact the financial viability of the biorefinery. The second advantageous impact of manipulating lignin structure is to facilitate the valorization of lignin through its incorporation in bio-based products.

In the traditional bioprocessing paradigm, the primary focus is to maximize the extraction of polysaccharides, which will then be converted to biofuel (i.e., ethanol). This is typically accomplished through microbial and/or enzymatic deconstruction.

TABLE 2-2 PERCENTAGE OF S, G, AND H MONOLIGNOLS IN VARIOUS BIOMASS SPECIES[38]

Feedstock	S	G	H
Poplar (*Populus euramericana*)[a]	63	37	-
Birch (*Betual verrucosa*)[a]	78	22	-
Spruce (*Picea abies*)[a]	Trace	98	2
Miscanthus[b]	44	52	4
Wheat (*Tritcum aestivum*)[a]	56	49	5
Alfalfa (*Medicago sativa*)[c]	39	56	5

[a]Thioacidolysis

[b] ^{13}C NMR

[c]Thioacidolysis and acetyl bromide

In the traditional bioprocessing paradigm, the primary focus is to maximize the extraction of polysaccharides, which will then be converted to biofuel (i.e., ethanol). This is typically accomplished through microbial and/or enzymatic deconstruction. However, plant cell walls are inherently resistant to such deconstruction – in no small part due to lignin. Accounting for ~25% of the secondary cell wall, lignin acts as a physical barrier that prevents enzymes from accessing polysaccharides [71]. Additionally, lignin participates in non-productive binding of these enzymes [71]. To address these issues, pretreatment is often performed to remove lignin and provide better access to polysaccharides. Certain lignin interunit linkages are more susceptible to degradation. For instance, β-O-4 ether bonds tend to be more easily cleaved during certain types of pretreatment [72, 73]. The proportion of β-O-4 bonds tends to be positively correlated with syringyl units, and therefore biomass with a higher S/G ratio is generally considered to be easier to deconstruct [74]. Several studies observed negligible differences in saccharification efficiency on unpretreated biomass, but enhanced saccharification efficiency on pretreated biomass across varying S/G ratios [10, 75], suggesting that recalcitrance is a multifaceted problem. Other lignin features, such as ferulates, introduce cross linking between lignin and polysaccharides, which can also impact cell wall deconstruction and/or polysaccharide accessibility [56]. Indeed, ferulate content was shown to be a good predictor of cell wall digestibility [76]. The analogous ester linked moiety in poplar, *p*-hydroxybenzoate (PB), has also been implicated in biomass digestibility [77]. However, the effect of PB of sugar release is difficult to differentiate since PB is correlated with other key lignin features such as the S/G ratio [60, 78]. Considering the complexity of these findings, extensive characterization of lignin structure is of utmost importance to understand its impact on the conversion of biomass to biofuels.

2.10 Lignin valorization for bio-based products

While manipulating lignin structure can impact conversion efficiency, it can also be utilized to facilitate valorization of lignin for use in bio-based products. As previously stated, lignin is the second most abundant terrestrial biopolymer (behind only cellulose) with an estimated yearly biosynthesis of $2x10^{10}$ metric tons [79]. At the industrial level, after the separation from lignocellulosic biomass, lignin is considered a low value waste residue (~$150/tonne) [80]. In the pulp and paper industry 50 million tons of lignin is burned for electricity generation due to its

high heating value [81]. Some of the factors hindering lignin from being valorized are its strong smell, dark color, heterogenous chemical structure, resistance to depolymerization, and the complex mixture of products that arises after pretreatment [82]. However, a major goal of the biorefinery paradigm is the complete valorization of residual lignin [7]. Indeed, past and current research on lignin valorization has focused on applications such as bio-based materials, fuels, and chemicals [83]. The production of valuable products from lignin represents an opportunity to improve the economic feasibility of biorefineries.

As referenced previously, pretreatment is usually implemented to separate the secondary cell wall components. Different pretreatment methods are available and will produce lignin with specific molecular structures, and especially functional groups, which will dictate the suitability of lignin for various end uses. Lignin is being explored for potential use in applications such as food additives, carbon fibers, adhesives, 3D printing, fuels and platform chemicals, along with many other uses [84]. One application which has garnered much attention in recent years is the use of lignin for biocomposites, such as in 3D printing [85]. Lignin based thermoplastics are biodegradable, have polystyrene like properties, and offer a green solution to petroleum-based thermoplastics [86]. Lignin has several features that are favorable for extrusion for 3D printing purposes, such as aliphatic ether groups, β-O-4 linkages, and oxygenated aromatic bonds [87]. Another approach gaining popularity is the so-called biological funneling of lignin to renewable materials through the utilization of microbes. Enzymes such as *Pseudomonas putida* produce enzymes that are able to depolymerize lignin into a narrow range of products [88]. This narrow range of depolymerized lignin products are then more effectively converted to desirable end products. However, these microbes must be capable of digesting pretreated lignin products, or else such capabilities must be engineered into the microbe. In fact, utilization of lignin is considered so essential to the economic viability of biorefineries that “lignin first” biorefining strategies now consider lignin valorization in the design phase [89]. In all these facets, characterization and/or control of the lignin structure is instrumental.

Chapter 3:
Experimental

3.1 Materials and chemicals

Populus genotypes from the genome wide association study (GWAS) work were obtained from a field site in Corvallis, OR (44°34′14.81″N 123°16′33.59″W) owned and maintained by the Oak Ridge National Laboratory (ORNL) Plant Biology group. Switchgrass genotypes for the GWAS work were obtained from a population propagated and maintained by our collaborators at the University of Georgia at Iron Horse Farm (33.728430, − 83.303594) in Watkinsville, GA. Material for both GWAS studies was dried at 50°C for at least seven days and Wiley milled to at least 40 mesh prior to receipt. Material for the drought leaf study and molybdenum content study was obtained from a field site in Clatskanie, OR owned and maintained by the ORNL Plant Biology group. Standard switchgrass material was provided by National Renewable Energy Laboratory (NREL). Standard poplar material was provided by the Center for Bioenergy Innovation (formerly the Bioenergy Science Center). Reference poplar material (RM 8492) was purchased from the National Institute for Standardsd and Technology (NIST).

Unless otherwise specified, all chemicals and reagents used in this work were of analytical grade purchased from Thermo Fisher Scientific (Hampton, NM) and Sigma-Aldrich Corp (St. Louis, MO) and used as received. HMPA-d_{18} was purchased from Cambridge Isotope Laboratories (Andover, MA). *p*-Dioxane was distilled with sodium borohydride to remove the inhibitor and subsequently diluted to 96% with deionized water.

3.2 Lignin preparation

Biomass, and especially lignin, must be properly prepared to ensure precise and accurate measurements. Biomass contains extractives (nonstructural components such as fats, waxes, proteins, resins, etc.), which have been shown to bias subsequent analyses. Therefore, removing extractives is almost a universal first step in preparing biomass for analysis. This is accomplished by extracting biomass with water and/or organic solvent(s). For this dissertation, the standard procedure was to extract raw biomass with a toleuene/ethanol (2:1, v:v) mixture in a Soxhlet extraction apparatus at a rate of 4 cycles per hour for a period of at least eight hours. After Soxhlet extraction, the biomass is air dried at room temperature, as oven drying of wet biomass

can result in hornification (irreversible agglomeration). The resulting biomass is referred to as "extractives-free biomass".

Extractives are often unique to the specific biomass feedstock. To remove these extractives, various solvents are utilized in Soxhlet extraction, such as ethanol, toluene, acetone, dichloromethane, water. Multiple extractions utilizing two or more of these solvents may also be utilized. As discussed in NREL/TP-510-42619 and largely demonstrated in literature, sufficient removal of extractives from poplar is accomplished through exhaustive ethanol and toluene/ethanol extraction. However, a review of the literature revealed a variety of solvent and solvent combinations to remove extractives from switchgrass (or other herbaceous biomass). To examine the potential effects of these various Soxhlet extraction methods, seven extraction strategies (Table 3-1) identified from literature and performed on a standard switchgrass material. Subsequent analysis by HSQC NMR revealed that the choice of Soxhlet extraction solvent(s) had minimal impact on most measured lignin traits. As illustrated in Figure 3-2, no statistically significant differences were detected for the most abundant lignin structures, such as S units, G units, p-coumarates, and β-O-4 interunit linkages. However, some differences in minor lignin traits were identified. For instance, as presented in Figure 3-3, the relative abundance of cinnamyl alcohol end units (I) was measured to be approximately 0.7%. Though, when H_2O only was utilized for Soxhlet extraction, the cross peak for the cinnamyl alcohol end unit was not detectable. This could lead to a miscalculation of moiety abundance, and therefore consideration should be given to extraction solvent in the biomass preparation procedure.

3.2.1 Milling

Another common operation in biomass preparation is ball milling. Biomass is typically first "knife milled" through a screen (i.e., 40 or 80 mesh) in a Wiley type mill. However, this is sometimes too coarse for analyses, and a smaller particle size is required. The standard procedure is to add 500-1,000 mg of extractives free biomass to a 50 mL zirconium oxide lined jar, along with 8-10 zirconium oxide grinding balls (10 mm diameter). This is loaded into a Retsch planetary ball mill and grinding is conducted for 1 h at five-minute intervals to avoid

TABLE 3-1 SOLVENT AND SOLVENT COMBINATIONS TESTED FOR SOXHLET EXTRACTION

Extraction #	Solvent
1	H_2O
2	Toluene/ethanol
3	Acetone
4	Dichloromethane
6	H_2O & toluene/ethanol
7	H_2O & dichloromethane
9	Toluene/ethanol & acetone

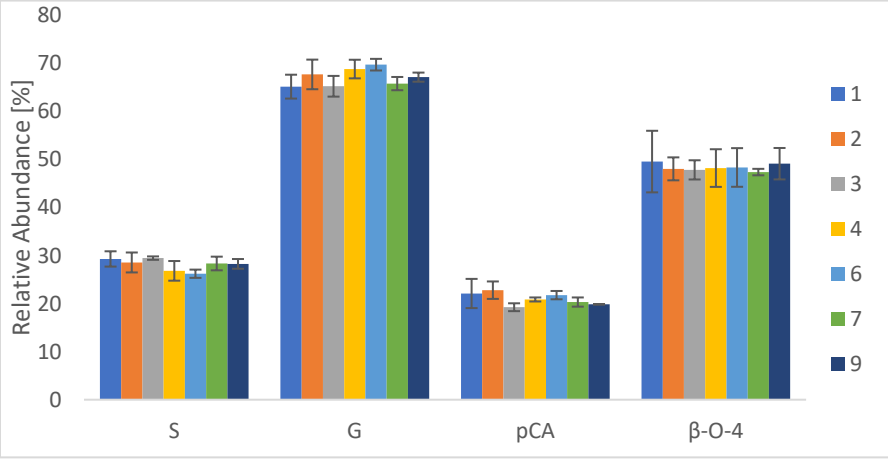


FIGURE 3-1 HSQC NMR RESULTS OF A REFERENCE SWITCHGRASS SAMPLE FOR EACH TEST SOLVENT FOR SOXHLET EXTRACTION

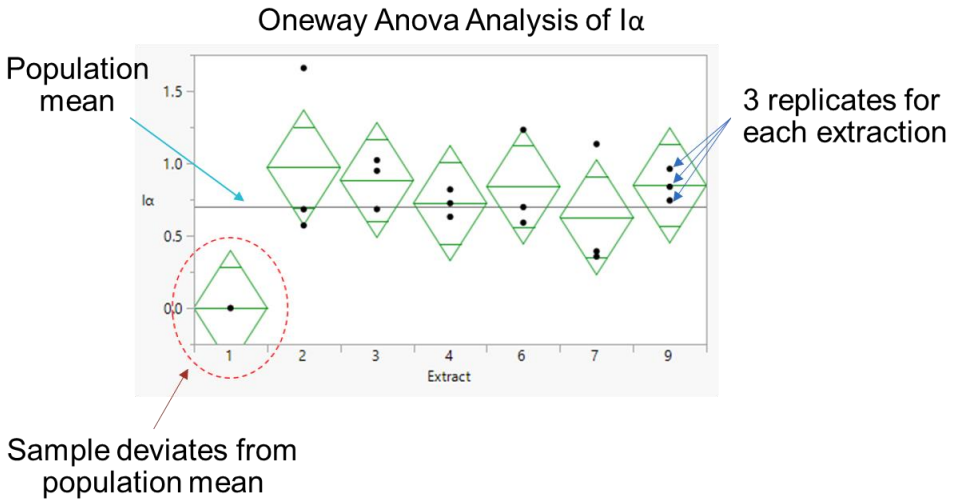


FIGURE 3-2 EXAMPLE OF DIFFERENCES IN HSQC NMR SPECTRA OF SWITCHGRASS DETECTED FROM DIFFERENT SOXHLET EXTRACTION SOLVENTS

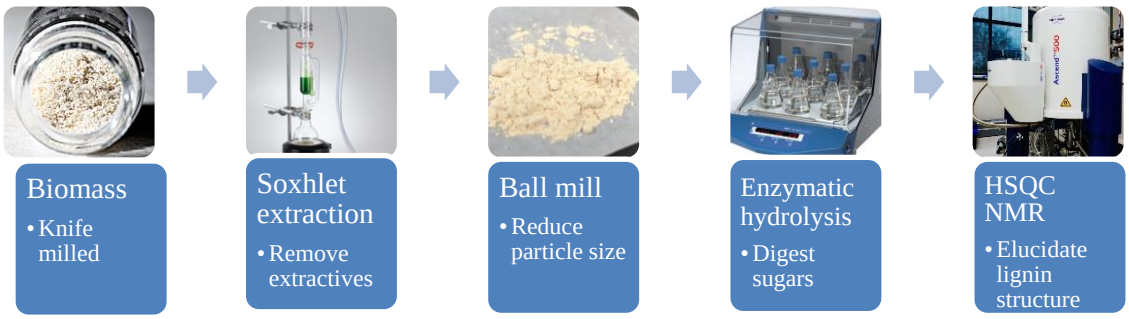


FIGURE 3-3 OVERVIEW OF KEY STEPS IN LIGNIN ISOLATION PROCESS

overheating. Ball milled biomass is a very fine powder. This facilitates more efficient enzymatic hydrolysis and is crucial for dissolution in NMR solvents.

3.2.2 Enzymatic hydrolysis

In some cases, it is desirable to isolate the lignin fraction of the secondary cell wall for analysis.

This is accomplished through enzymatic hydrolysis of polysaccharides by enzymes such as cellulases. For this work, cellulase from *Trichoderma* was added to extractives free, ball milled biomass at a loading of 50 mg enzyme per 1 g biomass in 50 mM sodium acetate buffer (pH=4.8). The mixture was incubated on an orbital shaker at 37^{o}C for 48 hours. Afterwards, the mixture was centrifuged and the supernatant was decanted. The lignin enriched solid residue was retained, washed with deionized water, centrifuged, and the supernatant was decanted. This washing was repeated one additional time. The resulting enzyme lignin (EL) was lyophilized for at least 48 h. This process is summarized in Figure 3-4.

3.3 HSQC NMR

For this work, the predominant NMR experiment utilized was heteronuclear single quantum coherence (HSQC). This is a two-dimensional technique which allows spectra to be resolved by J-coupled ^{1}H and ^{13}C nuclei. The resulting cross peaks can be integrated similar to one-dimensional NMR spectra, though the volume integrations of HSQC spectra are considered to be semi-quantitative. For NMR analysis, approximately 40 mg of enzyme lignin was dissolved in 0.6 mL DMSO-d_6 inside of a 5 mm NMR tube.

Alternatively, ball milled biomass can be directly dissolved into an appropriate solvent by HSQC NMR (i.e., without performing enzymatic hydrolysis). This is often termed whole cell wall (WCW) analysis, due to all primary cell wall components being present. WCW is utilized for several reasons, including (i) to eliminate the time intensive enzymatic hydrolysis and lyophilization steps, and (ii) to characterize cell wall polysaccharides. The solvent selection for WCW NMR is important, as both the lignin and polysaccharide fractions of the cell wall need to be solubilized. This is accomplished through the use of a co-solvent system. The Ragauskas group utilizes a DMSO-d_6/HMPA-d_{18} mixture (4:1, v:v). While HMPA-d_{18} is relatively

expensive, it produces a cross peak highly unlikely to interfere with any biomass cross peaks. Similar solvent systems such as DMSO-d_6/pyridine-d_5 mixture (4:1, v:v) are often employed, though pyridine-d_5 produces a cross peak in close proximity to biomass signals in the aromatic region of the HSQC NMR spectrum.

All solution state NMR experiments were conducted on a Bruker Avance III HD 500 MHz spectrometer (Bruker, Billerica, MA, USA) equipped with a N_2 cryoprobe. The Bruker pulse sequence hsqcetgpsip2.2 was used, which has the following parameters: spectra width of 12 ppm in the ^{1}H dimension with 1024 data points; spectra width of 220 ppm in the ^{13}C dimension with 256 increments, and 32 scans. An automatic phase correction and baseline was applied to all spectra, and the DMSO-d_6 reference peak of δ_C/δ_H 39.5/2.49 was used to calibrate the chemical shifts. An example of a resulting HSQC NMR spectra is offered in Figure 3-2.

3.4 FTIR

Fourier transform infrared spectroscopy (FTIR) is a highly utilized characterization technique due in part to its rapid analysis, nondestructive nature, and sensitivity to changes in molecular structure [90]. The FTIR instrument passes infrared (IR) radiation through the sample, and various structures (i.e., functional groups) will absorb some of this energy to produce characteristic absorption bands. Lignin mostly absorbs IR radiation in the mid-infrared region of 4,000-600 cm^{-1}. However, the 1,800-800 cm^{-1} range, referred to as the fingerprint region, typically provides the most information about lignin structural information. FTIR analyses in this dissertation were obtained with a PerkinElmer Spectrum 100 FTIR Spectrometer with Attenuated Total Reflection (ATR) accessory (PerkinElmer, Waltham, MA, USA). The standard scan settings are for the range of 4,000-600 cm^{-1} at a resolution of 2 cm^{-1} using 32 scans.

3.5 HPLC

High performance liquid chromatography (HPLC) is a chromatographic technique which is not dissimilar from GPC. Indeed, HPLC utilizes a mobile phase to move analytes through a stationary phase (column) which will elute analytes after some retention time in the column. The retention time of an analyte will depend on its interaction between the make-up of the mobile phase and the adsorption properties of the stationary phase. A compound will elute faster if it has

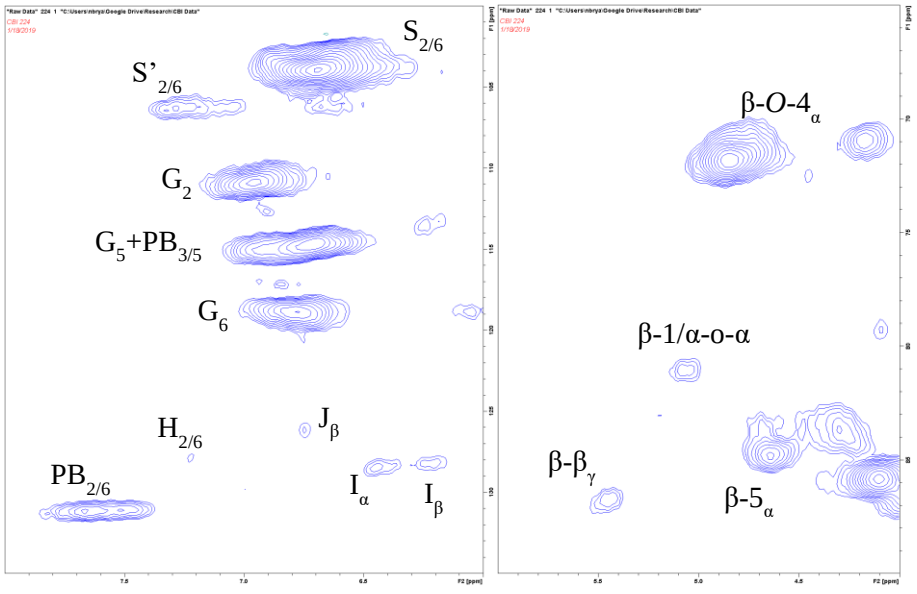


FIGURE 3-4 EXAMPLE HSQC NMR SPECTRA OF LIGNIN FROM *POPULUS TRICHOCARPA*

greater affinity with the mobile phase than the stationary phase. In normal phase chromatography, the stationary phase is hydrophilic and the mobile phase is hydrophobic. HPLC columns are typically packed with silica gel which contains highly polar silanol groups. The highly polar silanol groups will have higher interactions with polar analytes, retaining them in the column longer than non-polar analytes. Reversed phase HPLC can be also used, where the stationary phase is hydrophobic and the mobile phase is hydrophilic. In reversed-phase HPLC, long hydrocarbon chains are attached to the silica gel to make the stationary phase non-polar, allowing the polar analytes to elute from the column more quickly. Analytes eluted from the column are typically analyzed by an ultraviolet (UV) or refractive index (RI) detector.

HPLC is a widely used technique for quantitative sugar analysis. In fact, HPLC is the method recommended by the Association of Official Analytical Chemists as the standard technique for the routine analysis of carbohydrates [91]. For analyzing carbohydrates for biomass composition analysis, samples were analyzed by an Agilent 1200 series HPLC using the following conditions. For each sample 25 μL was injected into a Biorad Aminex HPX-87H column maintained at 65^{o}C. The mobile phase was 5 mM sulfuric acid prepared with degassed nanopure water and a flow rate of 0.6 mL/min was used. Adequate separation was achieved with the RI detector, with glucose eluted at approximately 9.5 min and xylose eluted at 10.25 min. An external five-point calibration curve was established using ≥99.5% purity monosaccharide standard material (Sigma Aldrich).

For determination of *p*-hydroxybenzoate content, samples were analyzed by an Agilent 1200 series HPLC. For each sample, 10 μL was injected onto a Symmetry C18 column (4.6 x 50 mm, 5 μm particle size) maintained at 35^{o}C. Adequate peak resolution was achieved using 75:25 (v:v) ratio of eluent B (0.1% trifluoroacetic acid in 70:30 acetonitrile:methanol) in eluent A (0.1% triflouroacetic acid in water) at a flow rate of 0.4 mL/min. Spectra were integrated at the UV maxima of 255 nm and an external five-point calibration curve of 4-hydroxybenzoic acid was used for quantification.

3.6 GWAS

Understanding the genetic architecture of complex traits, such as lignin, is fundamental to understanding the complex biology of plant cell wall formation. Most complex traits are influenced by many genetic loci, environmental conditions, and interactions between the two [92]. Indeed, the formation of the secondary cell wall requires the coordination of many metabolic pathways [93]. Genome wide association studies (GWAS) have become the method of choice for detecting causal single nucleotide polymorphisms (SNPs) for phenotypic traits. For clarification, a SNP refers to a nucleotide or base pair in DNA. An example would be a cytosine nucleotide being replaced by a thymine nucleotide. At the most fundamental level, GWAS aims to detect which SNP variations in the genome are associated with variations in the phenotype. This is accomplished through statistical testing. In principle, a statistical test such as ANOVA could be performed on each SNP individually across the entire population to assess the null hypothesis that there is no trait mean for any genotype group. In practice, GWAS statistical models are much more eloquent. One fundamental problem that GWAS models must overcome is that the genome is so large that patterns that appear to be associated with phenotypic variation are likely to arise by chance (i.e., false positives) [94]. Methods can be implemented to reduce the rate of false positives identified by GWAS. On method is implementing a false discovery rate (FDR). This method, typically based on *p*-values, is a way to control the expected proportion of Type I errors. Another common method is the Bonferroni correction, which sets the significance threshold based on the number of statistical tests to establish a corrected significance threshold. Another common issue which can bias GWAS results is relatedness among individuals. The most widely used GWAS method, the mixed linear model (MLM) [95], accounts for the population structure and kinship among related individuals in the population. Recently, multi-locus models have been gaining more attention. The first multi-locus model reported was the multi-locus mixed model (MLMM) [96]. In this approach, genetic and error variance components are calculated to assign each SNP a *p*-value. The most significant SNP is added to the model as a co-factor, and the process is repeated until the genetic variance unaccounted for by covariate SNPs falls below a specified threshold. Finally, a backward stepwise regression is performed (removing the least significant co-factor SNP at each iteration) to assign effect sizes

and *p*-values to all SNPs. Other multi-locus methods which are being more commonly utilized are fixed and random model circulating probability unification (FarmCPU) [97] and Bayesian information and LD iteratively nested keyway (BLINK) [98]. Both models are similar in that they bin the genome to calculate and account for kinship.

For the *Populus* GWAS, whole genome resequencing, single nucleotide polymorphism (SNPs)/nucleotide insertions and deletions (indels) calling and SnpEff analysis for the 917 individuals of this *Populus* GWAS population was previously described by Zhang et al. [99]. The *P. trichocarpa* Nisqually-1 reference genome v3.1 was used for read alignment and variant calling. The resulting SNP and indel dataset is available at http://bioenergycenter.org/besc/gwas/. This study utilized genotypic data for a subset of 409 genotypes from this dataset. To assess genetic control, we used the GEMMA software to calculate kinship for the *Populus* GWAS population as the correction factor for genetic background effects [99]. Genotype-to-phenotype associations were performed using 8,301,860 SNP and indel variants with minor allele frequencies > 0.05. The HSQC spectra from 2D HSQC NMR were used as phenotypes. Statistical significance of associations was evaluated using the Storey's Q-value threshold. Deviation of *p*-values from normality was assessed using quantile-quantile (Q-Q) plots. Candidate gene identification and RNAseq mapping for co-expression analysis were performed using the *P. trichocarpa* v3.1 reference genome. RNAseq of xylem tissue of 378 *Populus trichocarpa* transgenics plants knockdown or overexpressing monolignol genes and transcription factors involved in the regulation of cell wall biosynthesis were downloaded from the Sequence Read Achieve (SRA; accession: PRJNA314500) [100]. Library quality was assessed using FastQC (v0.11.9; https://www.bioinformatics.babraham.ac.uk/projects/fastqc/), residual adapters and low-quality reads were trimmed using Trimmomatic v0.39 [101] reads were mapped to the reference genome using STAR v 2.7.6a (default parameters and --outFilterMultimapNmax 100 [102] and transcript per million (TPM) values were extracted for all annotated genes using Stringtie [103]. 18 samples with low mapping rates (<80% of mapped reads) were excluded for the subsequent analysis. Co-expression of candidate genes with 86 phenylpropanoids and lignin-related genes was estimated by calculating pairwise Pearson correlation coefficient (PCC) across 360 samples using the function rcorr() from the Hmisc R package [104, 105]. For each potential

candidate gene, multiple individual scores were calculated: (i) significance threshold of $-\log_{10}(p\text{-value})$=6, 7, and 8 were assigned an individual score of 1, 2, and 3, respectively; (ii) connectivity with SNPs was scored according to $\log_{10}$(number of connected SNPs); (iii) connectivity of the SNPs with phenotypes; number and average value of significant co-expression associations ($|PCC| \geq 0.5$, FDR < 0.001) with lignin-related genes. All individual scores were scaled to obtain values ranging from 0 to 1. These individual scores were summed to obtain a final overall score, which was utilized to prioritize candidate genes for consideration (higher scoring genes were considered better candidates).

3.7 Error Analysis

Errors in chemical analysis are defined as the difference between the measured value and the true value. It is inevitable that each measurement has an associated degree of uncertainty. As a standard practice to quantify this uncertainty, samples were analyzed in duplicate when possible. In the case where it was not possible to run samples in duplicate (i.e., material availability), a standard biomass of similar type was analyzed in duplicate to quantify the measurement error.

3.7.1 NMR

The most utilized analytical technique in this work is NMR, and specifically the HSQC experiment. The HSQC pulse sequence is relatively long, requiring approximately 6 hours per sample. Additionally, GWAS requires a large sample size for sufficient power. A large sample size and long experiment time necessitates that each sample be measured only once. This naturally leads to questions regarding experimental error and reproducibility. Therefore, efforts have been made to evaluate address these inevitable questions.

First, a standard *Populus* – referred to as BESC Standard (BESC STD) – was analyzed in triplicate regularly (i.e., at least annually) to verify the reproducibility of the HSQC NMR measurement. This BESC STD material has previously been well characterized [106]. The results from these year-to-year verification experiments are presented in Table 3-2. To highlight two key phenotypes, the S/G ratio has an associated standard deviation of 0.02-0.05 (compared to an average of 1.4), and the β-O-4 content has an associated standard deviation of 0.91-1.42

TABLE 3-2 RESULTS FROM ANNUAL HSQC NMR ANALYSIS OF STANDARD POPULUS LIGNIN REPLICATES

	2018		2019		2020/2021		2022		Year-to-year	
	Mean	**StDev**	**Mean**	**StDev**	**Mean**	**StDev**	**Mean**	**StDev**	**Mean**	**StDev**
S	58.5	0.27	57.9	0.78	58.0	0.82	59.0	0.54	58.3	0.49
G	41.3	0.27	41.5	0.82	41.4	0.85	40.7	0.46	41.2	0.36
H	0.2	0.06	0.6	0.05	0.6	0.05	0.3	0.19	0.4	0.18
PB	14.1	0.06	16.3	0.36	16.3	0.37	16.5	0.42	15.8	1.15
S/G	1.4	0.02	1.4	0.05	1.4	0.05	1.4	0.03	1.4	0.02
β-O-4	61.5	1.42	60.5	1.36	60.5	1.34	61.6	0.91	61.0	0.63
β-5	4.0	0.25	3.7	0.70	3.7	0.71	4.0	0.12	3.9	0.19
β-β	6.1	0.36	5.1	0.59	5.1	0.59	3.9	0.74	5.1	0.90

(compared to an average of 61.0). These correspond to coefficient of variation (CV) values (ratio of the standard deviation to by the mean) of <5%, which is generally considered acceptable.

Additional work was performed on the NMR analysis of switchgrass. The standard switchgrass genotype "AP13" was processed and shipped to us in two separate batches by our collaborators at the University of Georgia. We prepared and analyzed these samples as independent technical replicates in June 2020 and March 2021. The results are presented in Table 3-3.

Between these two samples, the standard deviation of the S/G ratio was 0.03, and the standard deviation of the β-O-4 content was 0.99, which are both in line with the BESC STD *Populus* triplicate analysis. These numbers exemplify the robustness of repeatability of both the lignin isolation process and the NMR analysis.

To ensure that various lignin and cell wall properties were not impacting the HSQC NMR measurements, additional samples from across the switchgrass GWAS population were selected for replicate testing. Since the lignin S/G ratio is often correlated with many other phenotypes, samples with S/G ratios designated as high (#104), medium (#129), and low (#226) were selected. Retained material from these genotypes was re-analyzed in triplicate and compared to the original analysis, with results plotted in Figure 3-5. For all three genotypes, the S/G ratio between the original and replicates were similar within one standard deviation. For β-O-4 content the results were more variable, with the original and replicates samples similar within 2.5 standard deviations. Based on these results, it can be concluded that there is not a bias in the variability of the HSQC NMR measurements across the population structure.

While HSQC NMR is a powerful tool for evaluating lignin structure, other methods are also commonly used. A method that has been widely used for decades is thioacidolysis. This method is mechanistically different than NMR, as thioacidolysis destructively cleaves the β-O-4 bond and releases thioethylated H, G, and S monomers that are subsequently identified by GC/MS.

The S/G ratio of thioacidolysis and HSQC NMR are generally well correlated, with a recently study reporting a Pearson correlation coefficient (PCC) of 0.84 from a set of *Populus* genotypes [107]. To confirm the HSQC NMR results against another reputable analytical method, a set of

TABLE 3-3 HSQC NMR RESULTS OF SWITCHGRASS LIGNIN TECHNICAL REPLICATES ANALYZED 9 MONTHS APART

	Jun-20	Mar-21	Average	StDev
S/G Ratio	0.48	0.53	0.51	0.03
β-O-4	51.54	50.14	50.84	0.99

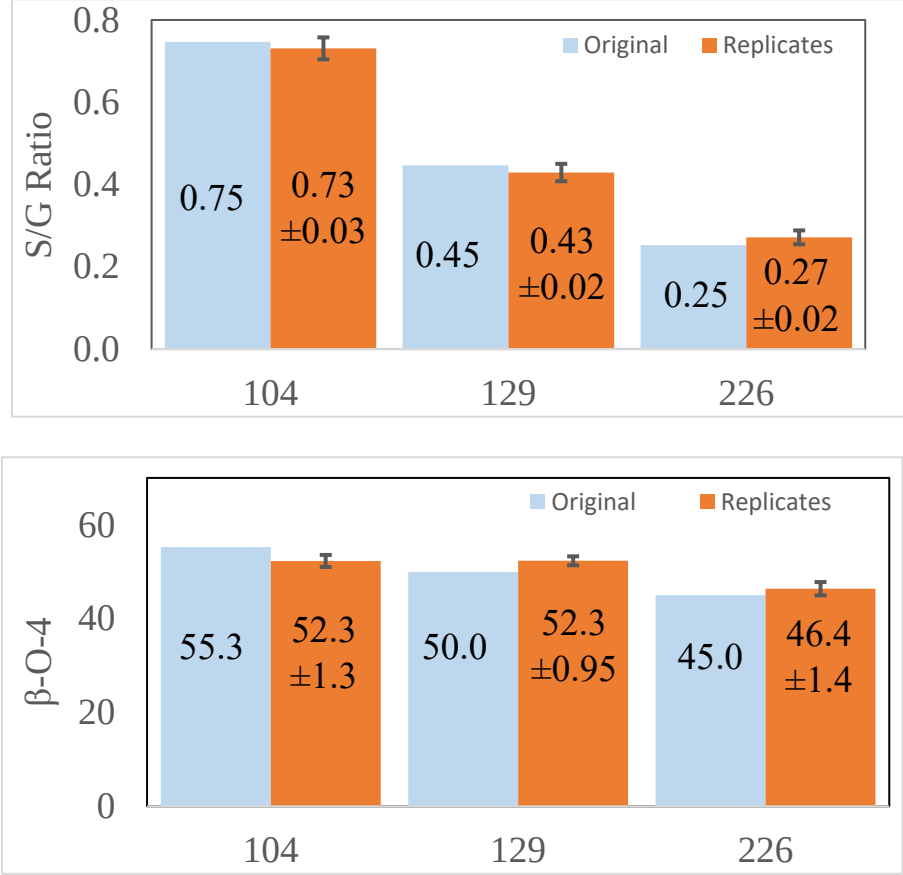


FIGURE 3-5 COMPARISON OF ORIGINAL SINGLE RUN HSQC NMR RESULTS TO ADDITIONAL REPLICATE RUNS

45 samples which spanned the range of the population S/G ratio was provided to our collaborators at the University of North Texas (now at University of Missouri) for independent analysis by thioacidolysis. A comparison between HSQC NMR and thioacidolysis results, plotted in Figure 3-6, returns 0.76 Pearson correlation coefficient. This slightly lower PCC (compared to Happs et al. [107]) is likely because switchgrass contains fewer β-O-4 bonds than poplar, which results in thioacidolysis releasing more dimers and trimers (and other higher order oligomers) which are difficult (or impossible) to elucidate by GC/MS, therefore biasing theresults. This observed strong correlation with independent thioacidolysis analysis provides additional confidence that HSQC NMR is elucidating authentic results.

3.7.2 HPLC

HPLC is the recommended analytical method for several biomass related measurements, though the most common is to quantify polysaccharides present in biomass. A natural variant Populus genotype referred to as BESC-131 is a well characterized biomass [108] that was used as a standard for each run. Figure 3-7 displays an example of the glucose and xylose peaks for this standard biomass. The measured glucose and xylose content of BESC-131 is reported in Table 3-4. This BESC-131 standard was measured to be approximately 47% (by weight) in the last two runs, which is well within the reported 48.4% glucose content [108]. The xylose content of BESC-131 was measured to be 8.2-12.5% for the last two runs, which is slightly lower than the ca. 16% reported xylose content. However, Meng et al. utilized high performance anionic exchange chromatography with pulsed amperometric detection, which is more sensitive than the refractive index detector (RID) used here for detecting minor sugars.

3.7.3 FTIR

FTIR has the benefits of being both fast and non-destructible. Therefore, it has good potential for the screening of whole biomass. Indeed, FTIR has been used to evaluate cellulose crystallinity[109], lignin S/G ratio [110], and many other cell wall components such as hemicellulose and pectin [111]. FTIR spectra are also commonly utilized in principal component analysis – an unsupervised learning technique – to quantify differences in cell wall composition. Therefore, a standard switchgrass material was analyzed in triplicate to quantify the

reproducibility of the FTIR measurement, and the resulting spectra are displayed in Figure 3-8. As FTIR can be utilized for many different measurements, and associated interpretations are also highly dependent on baselining and normalization, no specific calculations are offered here.

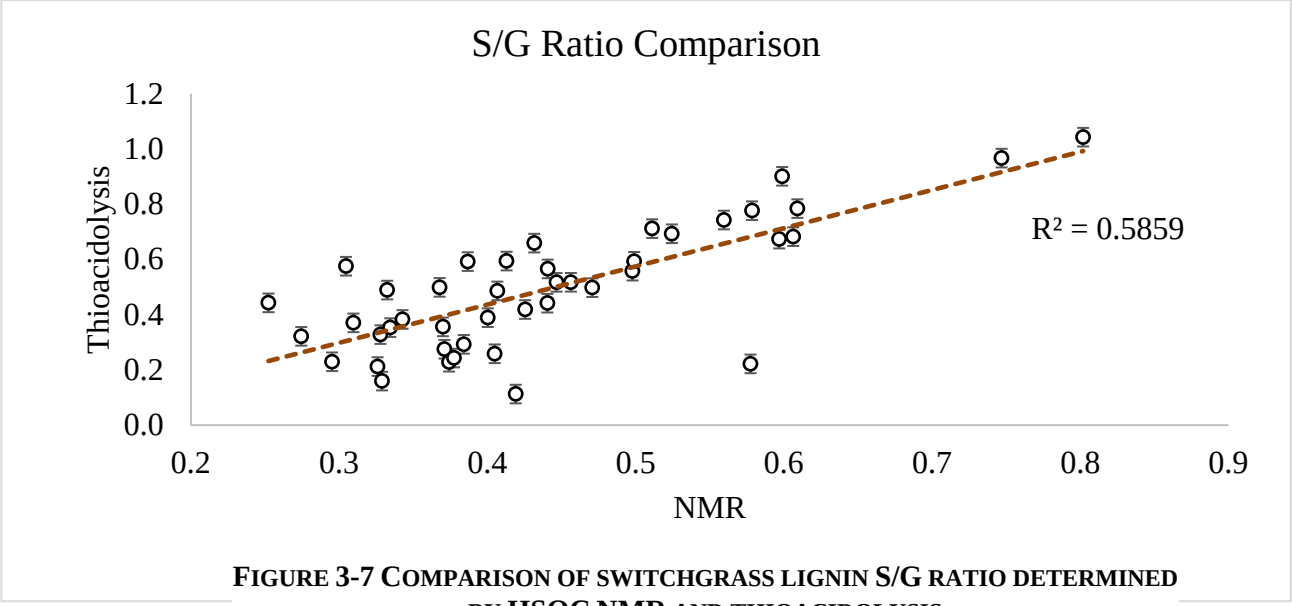


FIGURE 3-7 COMPARISON OF SWITCHGRASS LIGNIN S/G RATIO DETERMINED BY HSQC NMR AND THIOACIDOLYSIS

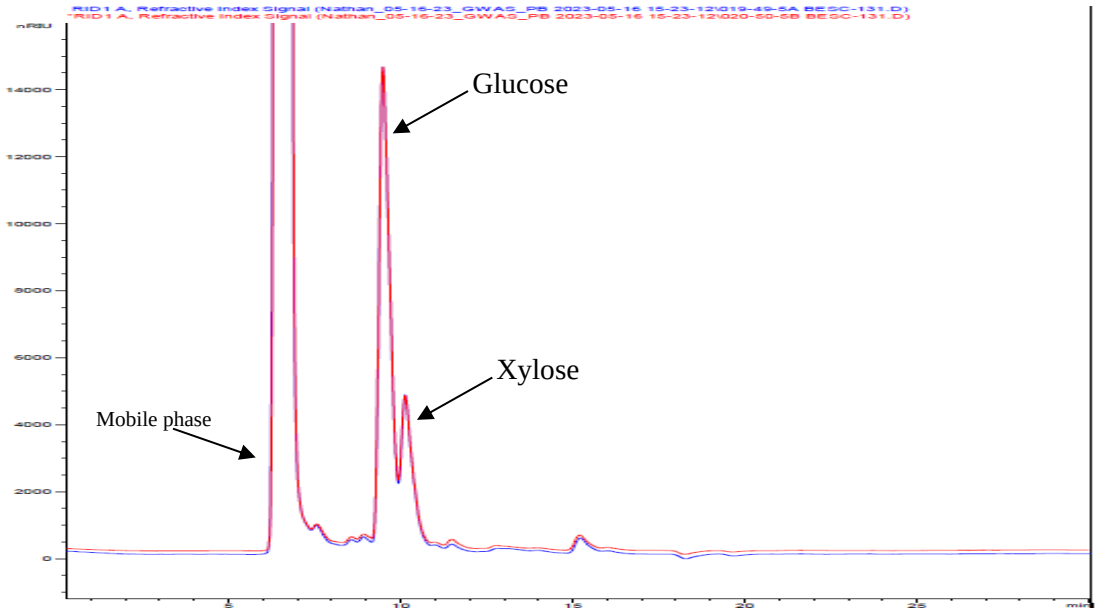


FIGURE 3-6 HPLC SPECTRA OF CELL WALL POLYSACCHARIDES FROM COMPOSITIONAL ANALYSIS

TABLE 3-4 MASS FRACTION OF CELL WALL POLYSACCHARIDES FROM STANDARD POPULUS DETERMINED BY HPLC

	Glucose	Xylose
April 2023	47.0% ± 2.3%	12.5% ± 1.0%
January 2023	47.1% ± 2.1%	8.2% ± 2.0%

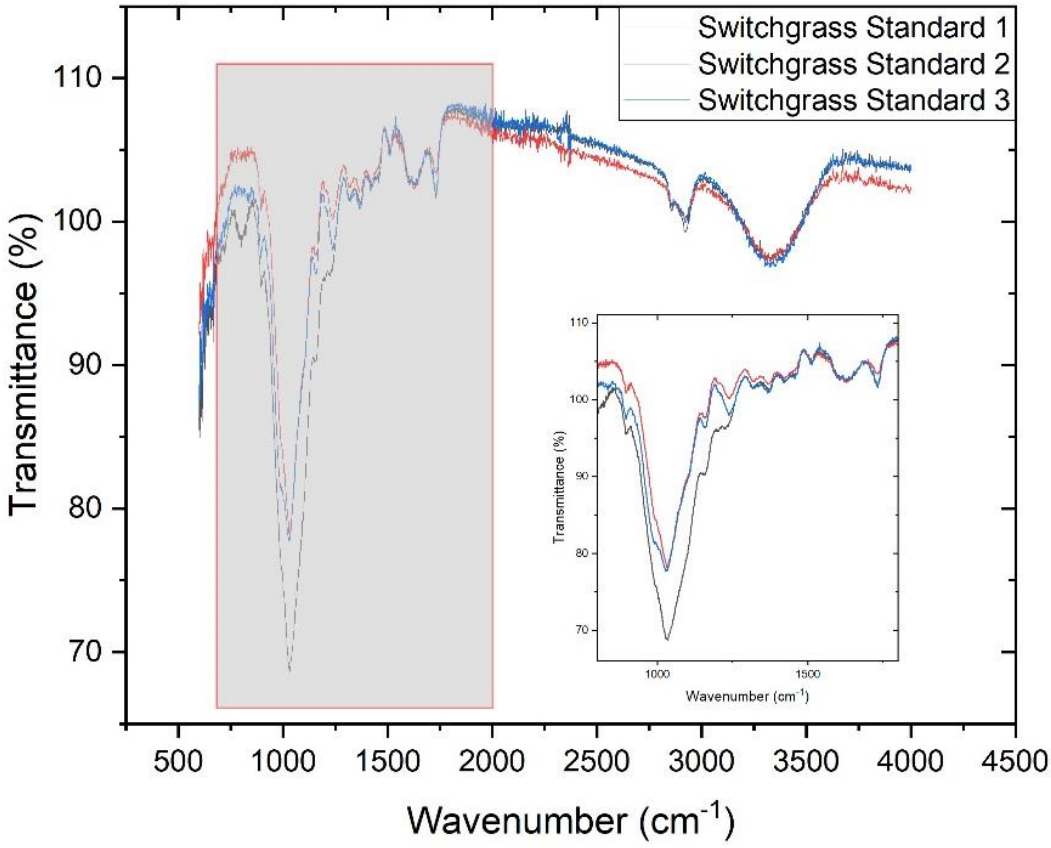


FIGURE 3-8 THREE REPLICATE FTIR ANALYSES OF NATIVE

Chapter 4:

2D HSQC Chemical shifts of Impurities from biomass pretreatment

Sections of this chapter was originally published as:

Bryant, G. Yoo, C., Pu, Y., & Ragauskas, A. J. (2020). 2D HSQC Chemical Shifts of Impurities from Biomass Pretreatment. *ChemistrySelect (Weinheim)*, 5(11), 3359–3364.

Arthur J. Ragauskas conceived the study. Nathan Bryant prepared samples, analyzed the data, and drafted the manuscript. Nathan Bryant and Yunqiao Pu performed the experimental work. All authors reviewed, revised, and approved the manuscript. Wiley does not require a release for an author to reprint an article in their dissertation.

4.1 Abstract

Two dimensional (2D) heteronuclear single quantum coherence (HSQC) nuclear magnetic resonance (NMR) is a powerful analytical method which can be used to elucidate the structure of biomass. During processing, biomass is typically subjected to some form of chemical treatment which can be performed with a variety of compounds. The presence of these compounds, even in trace amounts, has the potential to contaminate the sample and lead to misinterpretation of the HSQC spectra. Here we report the chemical shifts of 29 compounds commonly used in biomass processing which have the potential to contaminate the biomass samples and lead to the misinterpretation of peaks associated with biomass (*Populus trichocarpa*) pretreated via autohydrolysis. The identification of these chemical shifts could serve as a valuable tool in preventing errors in characterizing biomass via HSQC.

4.2 Introduction

Two dimensional (2D) heteronuclear single quantum coherence (HSQC) nuclear magnetic resonance (NMR) is an effective analytical method of elucidating the structural features of lignin [112]. 2D HSQC NMR offers improved resolution over corresponding one-dimensional NMR techniques (i.e., ^{1}H or ^{13}C) and allows for the determination of structural information such as composition and linkages of both polysaccharides and lignin simultaneously [112].

Though desirable in order to reduce dependence on fossil fuels, the production of ethanol from renewable resources such as lignocellulosic biomass is hindered by the natural recalcitrance of the

plant cell wall [113]. In order to overcome cell wall recalcitrance and achieve high saccharification efficiency, lignocellulosic biomass is usually subjected to a variety of thermochemical pretreatments [9, 113, 114]. Pretreatment methods such as dilute acid, alkali, and organosolv have been well explored for decades [115]. More recently, new pretreatment technologies utilizing ionic liquids (IL) and deep eutectic solvents (DES) have emerged [115]. While these chemical pretreatment methods help to overcome the cell wall recalcitrance, there will inevitably be some level of these chemical impurities in the pretreated biomass [114]. Jiang et al. specifically identified the presence of several degradation products such as carboxylic acids, alcohols, esters, and aromatics in dilute acid pretreated corn stover [116].

Although not the only factors, it is known that lignin content, structure, and composition contribute to recalcitrance [9]. For instance, it has been shown that the ratio of syringyl (S) to guaiacyl (G) lignin may play a significant role in pretreatment and subsequent hydrolysis [9]. Studer et al. observed improved sugar release from hot water pretreated *Populus trichocarpa*, even in samples with higher lignin content [10]. Analysis of biomass composition both before and after treatment is necessary to evaluate both the effect and efficiency of the pretreatment method [113].

The presence of solvents, decomposed compounds, and other contaminants in biomass could potentially affect structural characterization after pretreatments. Therefore, it is beneficial to identify the HSQC peaks of these impurities. In this study, we evaluated several chemicals which could be introduced to biomass during commonly used pretreatment methods.

4.3 Results and Discussion

Many pretreatments of lignocellulosic biomass are conducted with solvents, catalysts, and other chemicals. While biomass is typically washed after pretreatment to remove these chemicals, trace levels of chemicals may still be present in biomass and potentially contaminate an HSQC spectrum. Our previous study evaluated potential biomass contaminants for ^{31}P NMR spectra [114]. Another study clearly identified that a reagent used in lignin isolation could be misinterpreted as the *p*-hydroxyphenyl peak in an HSQC spectrum [117]. In order to evaluate the effects of trace contaminants on HSQC spectra of pretreated biomass, 0.001 mol of each contaminant was added to approximately 50 mg of poplar subjected to an autohydrolysis pretreatment. The peaks

associated with contaminants identified within the vicinity of biomass peaks in the HSQC spectra are tabulated in Tables 4-1 through 4-8.

A relatively recent innovation in biomass pretreatment is the use of IL and DES. Several ILs and DESs were selected for analysis via HSQC based on previous studies. ILs selected include 1-Ethyl-3-methylimidazolium acetate [118], 1-Butyl-3-methylimidazolium acetate [119], 1-ethyl-3-methylimidazolium diethyl phosphate [120], 1,3-dimethylimadazolium dimethyl phosphate [121], and 1-benzyl-3-methylimidazolium chloride [122]. DESs selected for this study include a mixture of the hydrogen bond acceptor choline chloride (ChCl) with various hydrogen bond donors. These mixtures include ChCl:lactic acid (1:10 mole ratio) [123, 124], ChCl:malic acid (1:1 mole ratio) [123, 124], ChCl:citric acid (1:2 mole ratio) [123][16], ChCl:oxalic acid (1:1 mole ratio) [124], ChCl:acetic acid (1:2 mole ratio) [123], ChCl:succinic acid (1:2 mole ratio) [123], ChCl:urea (1:2 mole ratio) [124, 125], ChCl:malonic acid (1:1 mole ratio) [124], ChCl:1,4-butanediol (1:2 mole ratio) [126], and ChCl:glycerol (1:1 mole ratio) [125].

HSQC is often utilized to examine the structure of lignin. Specifically, the peaks for monolignol aromatic units exist primarily in the 6.0-8.0 δ_H / 100-140 δ_C region of the spectra, commonly referred to as the aromatic region. Similarly, peaks associated with lignin interunit linkages are found in the 4.0-6.0 δ_H / 60-100 δ_C region of the spectra, also referred to as the aliphatic region. Peaks associated with polysaccharides are found in a wide range in the spectrum, but typically reside in the 2.5-6.0 δ_H / 50-105 δ_C region. These regions were closely examined for the presence of peaks associated with contaminants. Contaminant peaks, especially those of high intensity, overlapping, bordering, or in proximity of peaks associated with biomass, were observed, documented, and recorded in Tables 4-7 and 4-8.

In this study, for each compound, two separate sets of samples were analyzed. The first sample contained compound and the NMR solvent. The second sample contained the compound, autohydrolysis pretreated biomass, and the NMR solvent. The same concentration of compound and NMR solvent was used in each sample. This method allows for the identification of the pure compound peaks that may not be clear or obvious when combined with additional peaks from the biomass sample.

4.4 Materials and Methods

All reagents utilized for this study were obtained from commercial sources and used as received. DESs were synthesized by adding the hydrogen bond acceptor (in this case, choline chloride) and hydrogen bond donor in the specified mole ratios and stirred in a sand bath at 60-80^{o}C until a clear uniform liquid was formed. To promote consistency of HSQC signals among the contaminants, equal amounts of each chemical were used. Specifically, 0.001 mol of chemical contaminant was added to its respective sample. For this study, two sets of samples were prepared. The first set of samples contained the chemical contaminant and 0.75 mL of the NMR solvent DMSO-d_6/HMPA-d_{18} (4:1, v:v) in order to identify the chemical shift of the contaminant clearly. The second set of samples contained the chemical contaminant, 0.75 mL of the NMR solvent DMSO-d_6/HMPA-d_{18} (4:1, v:v), and 40 mg of biomass. The biomass used for these samples was *Populus trichocarpa* which had been pretreated via autohydrolysis and subsequently washed with deionized water. The pretreated biomass was ball-milled prior to NMR analysis. One sample containing only biomass and NMR solvent was also prepared and used as a control. The HSQC spectra of the samples were acquired with a Bruker Avance III HD 500 MHz spectrometer. The Bruker pulse sequence hsqcetgpsip2.2 was utilized on an N_2 cryoprobe with the following parameters: spectra width of 12 ppm in the ^{1}H dimension with 1024 data points; spectra width of 220 ppm in the ^{13}C dimension with 256 increments, and 32 scans. The HSQC spectra were analyzed with Bruker TopSpin 3.5pl6 software. The DMSO-d_6 solvent peak (δ_C/δ_H at 39.5/2.49) was used to calibrate the chemical shifts.

By running samples in this manner, one observation that can be made is the differences in chemical shifts between the biomass and non-biomass samples. In general, the chemical shifts were comparable. For organic solvents and enzymes, the differences in chemical shifts were largely limited to <0.05 PPM δ_H and <0.1 PPM δ_C. Similarly, the deep eutectic solvents also showed minimal differences between samples. However, more notable differences in chemical shifts were noted when analyzing the spectra of the ionic liquid compounds. For instance, 1-ethyl-3-methylimidazolium acetate alone was observed to have chemical shifts at 10.51/137.9, 7.99/123.5, and 8.11/122.0 (δ_H/δ_C), corresponding to C2, C4, and C5 carbons in the heterocyclic ring. When combined with the pretreated biomass, the same 1-ethyl-3-methylimidazolium acetate chemical shifts were observed at 9.85/137.1, 7.90/121.7, and 7.81/123.4 (δ_H/δ_C),

TABLE 4-1 NMR CHEMICAL SHIFTS OF ORGANIC SOLVENTS

Chemical	Chemical + Biomass	
	δ_H	δ_C
1,4-butanediol	1.42	29.0
	3.37	60.6
1,4-dioxane	3.59	66.3
Acetic Acid	1.91	8.3
	1.91	20.7
Acetone	2.07	30.3
	3.55	66.1
Ethanol	1.02	18.1
	3.16	48.3
	3.43	55.8
Glycerol	3.32	63.1
	3.43	72.4
Methanol	3.15	48.3
Tetrahydrofuran	1.81	25.0
	3.64	66.8
Toluene	2.26	20.5
	7.11	124.9
	7.14	128.3
	7.21	127.9

TABLE 4-2 NMR CHEMICAL SHIFTS OF IONIC LIQUIDS

Chemical	Chemical + Biomass	
	δ_H	δ_C
1,3-dimethylimidazolium phosphate	2.49	39.4
	2.52	23.7
	3.10	48.2
	3.29	51.1
	3.88	35.1
	7.86	123.3
	7.96	161.5
	9.63	137.7
1-benzyl-3-methylimidazolium chloride	3.88	35.5
	5.60	51.2
	7.29	128.4
	7.55	128.3
	7.94	123.6
	8.09	122.0
	9.96	136.6

TABLE 4-2 CONTINUED

Chemical	Chemical + Biomass	
	δ_H	δ_C
1-Butyl-3-methylimidazolium acetate	0.82	18.6
	0.82	12.9
	1.02	18.4
	1.14	18.5
	1.18	31.3
	1.19	12.8
	1.24	18.6
	1.36	18.5
	1.60	25.3
	1.73	31.3
	1.74	18.4
	3.89	35.1
	4.20	47.9
	4.21	31.3
	7.90	123.4
	7.99	122.1
	10.21	137.8
1-Ethyl-3-methylimidazolium acetate	1.38	14.8
	1.62	24.7
	3.88	35.1
	4.22	43.6
	7.81	123.4
	7.90	121.7
	9.85	137.1

TABLE 4-3 NMR CHEMICAL SHIFTS OF IONIC LIQUIDS

Chemical	Chemical + Biomass	
	δ_H	δ_C
1-ethyl-3-methylimidazolium diethyl phosphate	1.06	59.0
	1.06	16.3
	1.41	43.7
	1.42	15.0
	3.65	59.0
	3.69	16.3
	3.89	35.2
	4.24	43.7
	4.30	15.1
	7.95	123.5
	8.05	122.0
	9.93	137.3

TABLE 4-4 NMR CHEMICAL SHIFTS OF DEEP EUTECTIC SOLVENTS

Chemical	Chemical + Biomass	
	δ_H	δ_C
Choline chloride:1,4-butanediol	1.39	60.6
	1.39	28.9
	3.16	52.9
	3.34	60.6
	3.34	29.0
	3.46	66.8
	3.80	54.9
Choline chloride:Acetic Acid	1.89	20.8
	3.18	52.8
	3.47	66.8
	3.81	54.9
Choline chloride:Citric Acid	2.65	42.8
	2.70	42.8
	3.14	53.2
	3.42	67.0
	3.72	63.7
	3.80	55.3
	4.45	58.6
Choline chloride:Glycerol	3.16	52.9
	3.41	72.6
	3.46	66.7
	3.80	54.9
	3.82	66.7
	4.62	62.9

TABLE 4-5 NMR CHEMICAL SHIFTS OF DEEP EUTECTIC SOLVENTS

Chemical	Chemical + Biomass	
	δ_H	δ_C
Choline chloride:Lactic Acid	1.20	65.7
	1.20	20.3
	1.36	68.2
	1.37	16.7
	3.12	53.2
	3.41	66.9
	3.82	55.4
	4.01	65.7
	4.14	65.7
	4.47	57.9
	4.89	68.3
	4.92	68.9
	5.04	68.0
Choline chloride:Malic Acid	2.60	39.5
	3.21	53.1
	3.50	67.0
	3.82	63.7
	3.84	55.2
	4.28	67.1
	4.51	58.2

TABLE 4-5 CONTINUED

Chemical	Chemical + Biomass	
	δ_H	δ_C
Choline chloride:Malonic Acid	3.16	53.0
	3.20	42.0
	3.45	66.9
	3.80	55.0
Choline chloride:Oxalic Acid	3.16	52.9
	3.22	52.8
	3.45	66.8
	3.80	55.0
	3.84	63.0
	4.58	59.2
Choline chloride:Succinic Acid	2.39	28.8
	2.50	28.7
	3.16	52.9
	3.45	66.9
	3.73	63.6
	3.80	55.1
	4.44	57.7

TABLE 4-6 NMR CHEMICAL SHIFTS OF DEEP EUTECTIC SOLVENTS

Chemical	Chemical + Biomass	
	δ_H	δ_C
Choline chloride:Urea	1.18	28.6
	2.05	30.5
	2.71	52.8
	3.16	52.9
	3.45	54.9
	3.46	66.8
	3.61	53.0
	3.80	66.6
	3.80	54.9

TABLE 4-7 NMR CHEMICAL SHIFTS OF ENZYMES

Chemical	Chemical + Biomass	
	δ_H	δ_C
β-glucosidase	1.21	28.7
	3.97	55.0
	Overlaps guaiacyl C5 signal at 6.61/114.6	
	7.01	129.7
	7.20	127.9
	7.21	128.9
	7.23	129.1
Xylanase	2.07	30.3
	3.55	66.1

TABLE 4-7 CONTINUED

Chemical	Chemical + Biomass	
	δ_H	δ_C
Cellulase	0.78	21.0
	0.79	15.1
	0.81	18.8
	0.85	22.6
	0.86	11.2
	1.21	17.5
	1.22	28.6
	3.06	69.9
	3.28	71.9
	3.31	79.8
	Overlaps Cγ-Hγ of β-*O*-4 substructure signal at 3.40/60.7	
	3.40	73.1
	Overlaps Cγ-Hγ of β-*O*-4 substructuresignal at 3.45/60.8	
	Overlaps Cγ-Hγ of β-*O*-4 substructure signal at 3.45/62.8	
	3.48	73.3
	3.57	71.4
	3.63	73.0

TABLE 4-7 CONTINUED

Chemical	Chemical + Biomass	
	δ_H	**δ_C**
Cellulase	Overlaps Cγ-Hγ of β-*O*-4 substructure signal at 3.67/60.4	
	4.29	96.8
	4.70	98.7
	4.88	91.9
	4.97	100.6
	6.65	114.7
	6.90	129.4
	7.02	129.9
	7.19	127.7
	7.21	128.9

TABLE 4-8 NMR CHEMICAL SHIFTS OF PRETREATMENT DEGRADATION PRODUCTS

Chemical	Chemical + Biomass	
	δ_H	δ_C
Furfural	6.71	112.8
	7.49	122.9
	8.04	148.9
	9.40	178.1
	9.54	178.1
	9.59	152.5
	9.66	178.1
	9.81	178.0
Hydroxymethylfurfural	4.04	55.8
	4.49	55.9
	4.94	55.7
	5.86	55.9
	6.57	109.4
	7.44	124.2
	9.52	151.7
	9.57	117.5

respectively. This illustrates a difference of 0.18-0.66 PPM δ_H and 0.1-0.8 PPM δ_C and can be easily observed when the two spectra are compared adjacently as in Figure 4-1. Additionally, the C2 of 1-butyl-3-methylimidazolium acetate had a chemical shift at 10.43/138 (δ_H/δ_C), but when combined with the pretreated biomass the chemical shift was located at 10.21/137.8 (δ_H/δ_C), a difference of 0.22/0.2 (δ_H/δ_C). Differences in shifts were not consistent for all peaks within the spectra. For instance, for 1-butyl-3-methylimidazolium acetate the chemical shift of the C6 carbon occurred at 1.73/31.3 (δ_H/δ_C) in both samples.

Perhaps the greatest benefit with identifying potential biomass contaminant peaks is the potential to prevent errant interpretation of the HSQC spectrum. As described previously, it has been shown that certain enzymes may interfere with interpretation of the *p*-hydroxyphenyl peak in the HSQC spectra [8]. It therefore stands to reason that the chemical shifts of other compounds could potentially contaminate biomass signals in the HSQC spectra. This was confirmed after examining the spectra of the compounds in this experiment. An example is illustrated below with the ethanol spectrum. The spectra on the left, from the pretreated biomass in NMR solvent, shows a methoxy signal at 3.69/55.4 (δ_H/δ_C). The spectra on the right shows that the C1 carbon of ethanol in NMR solvent generates a signal in approximately the same location at 3.45/55.8 (δ_H/δ_C). When ethanol is combined with pretreated biomass in the NMR solvent in the middle spectra, overlap of these two signals can be observed. This overlap may not be immediately obvious, especially at certain intensities. However, with the knowledge of the location of this ethanol signal and careful examination of the spectra, two signals in this area can be distinguished. Figure 4-2 clearly illustrates the individual location and overlap of these two signals. From this example, it is clear how compounds such as organic solvents can cause misinterpretation of the biomass signal. If this ethanol signal is mistakenly integrated with the methoxy peak, it could result in the methoxy content being overestimated.

In summary, the identification of signals in the HSQC spectra is crucial to characterize biomass structure accurately. Trace chemical contaminants can introduce chemical shifts that have the potential to interfere with biomass signals and lead to incorrect analysis of the spectra.

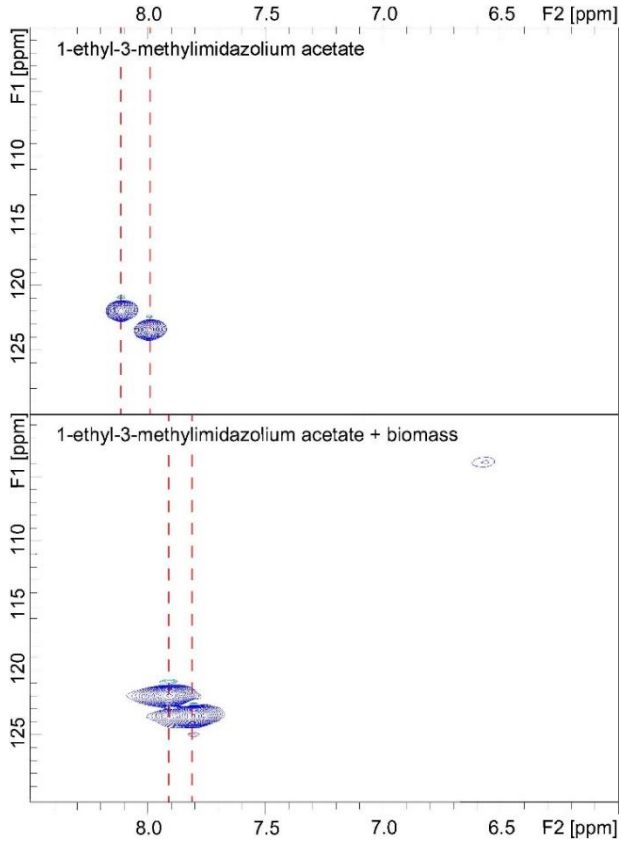


FIGURE 4-1 EXAMPLE OF CHANGE OF CHEMICAL SHIFT IN THE PRESENCE OF PRETREATED BIOMASS

In this study, 29 chemicals used in biomass pretreatment were combined with pretreated biomass and analyzed by HSQC in a DMSO/HMPA solvent system in order to identify the impact of these trace contaminants on the resulting spectra. Observed chemical shifts of these compounds, especially those with the potential to interfere with the interpretation of pretreated biomass HSQC spectra, were documented and reported in Tables 4-1 through 4-6. Several contaminant signals which overlapped with biomass signals were identified. Additionally, it was identified several instances in which there was a difference in the chemical shifts between the sample that contained only the chemical and the sample that contained the chemical and biomass. Observed chemical shifts of these compounds, especially those with the potential to interfere with the interpretation of pretreated biomass HSQC spectra, were documented and reported in Tables 4-7 and 4-8.

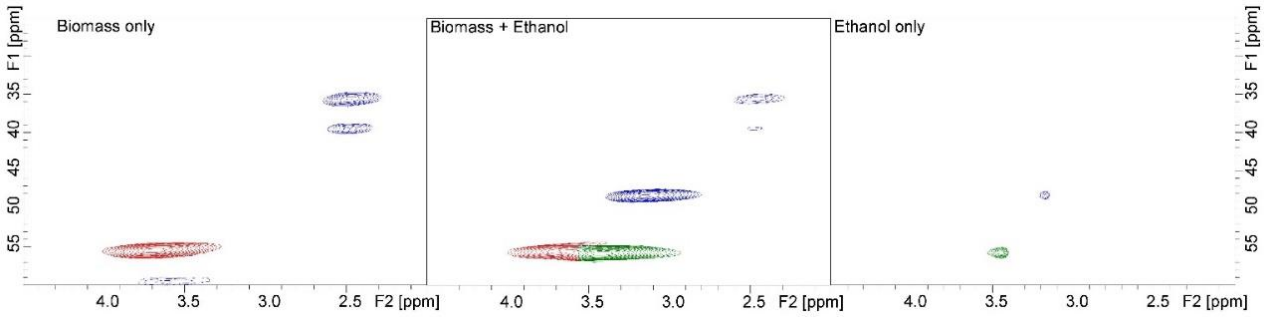


FIGURE 4-2 EXAMPLE OF CHEMICAL SHIFT OVERLAP BETWEEN BIOMASS CONTAMINANT

Chapter 5:

Cell Wall Response of Field Grown *Populus* to *Septoria* Infection

Sections of this chapter was originally published as:

Nathan Bryant, Wellington Muchero, Rachel A. Weber, Jaime Barros, Jin-Gui Chen, Timothy J. Tschaplinski, Yunqiao Pu, & Arthur J. Ragauskas. (2023). Cell wall response of field grown Populus to Septoria infection. Frontiers in Plant Science, 14. https://doi.org/10.3389/fpls.2023.1089011

Arthur J. Ragauskas conceived the study. Wellington Muchero and Jin-Gui Chen harvested the material and identified the fungus. Nathan Bryant, Yunqiao Pu, Rachel Weber, and Jaime Barros performed experiments and analyzed the data. Nathan Bryant drafted the manuscript. All authors reviewed, revised, and approved the manuscript. *Frontiers* publishes under a creative common license, meaning authors retain copyright and material can be reprinted with citation of the original publication.

5.1 Abstract

Due to its ability to spread quickly and result in tree mortality, *Sphaerulina musiva* (*Septoria*) is one of the most severe diseases impacting *Populus*. To resist *Septoria* infection, the tree will form a lignified periderm to contain the pathogen. Previous studies have identified that Septoria infection induces differential expression of phenylpropanoid biosynthesis genes. However, more extensive characterization of changes to the cell wall in response to *Septoria* infection is lacking. This study investigated the long-term impacts of *Septoria* infection on field-grown poplar, with a particular focus on lignin. Fourier-transform infrared spectroscopy and subsequent principal component analysis revealed that infected and border (along the periderm) tissue were similar and could be distinguished from the non-infected (healthy) tissue. Nuclear magnetic resonance results indicated the general trend that infected tissue had a higher syringyl:guaiacyl ratio and lower *p*-hydroxybenzoate content than the healthy tissues from the same genotype. Finally, Klason lignin content in the infected and/or border tissue was shown to be higher than healthy tissue, which is consistent with previous observations of periderm development and metabolite

profiling. These results provide insights on the response of *Populus* wood characteristics to *Septoria* infection, especially between healthy and infected tissue within the same genotype.

5.2 Introduction

Populus has garnered interest as an economically important species for applications such as biofuel production due to its rapid growth, genetic diversity, and other advantageous attributes [127]. One of the most critical drivers of biofuels' economic feasibility is biomass yield [128]. Challenges to poplar growth productivity include susceptibility to fungal infections from pathogens such as *Sphaerulina musiva,* also known as *Septoria* [129]. *Septoria* is particularly devastating due to its potential to be fatal to the tree and its ability to spread quickly among entire populations. Indeed, *Septoria* is considered one of the most severe diseases impacting hybrid poplar [130]. The two symptoms typically associated with *Septoria* infection are leaf spots and stem cankers [131]. Leaf spots have a significant negative effect on photosynthesis and can lead to defoliation. Stem canker can cause broken tops, leading to severe growth penalties or death [132, 133].

The secondary cell wall plays a role in many processes, including response to biotic and abiotic stress. It has been shown that lignin, a complex biopolymer that typically constitutes between 16 to 29% of *Populus* secondary cell walls [134], is often synthesized and deposited at the site of fungal infections to form a periderm that prevents the spread of the pathogen. This result is consistent with previous studies of *Septoria* infection of *Populus*, with increased lignin deposition observed at the infection site. Previous transcriptome analyses of *Septoria* and other fungal infections have also identified differentially expressed genes in the phenylpropanoid and lignin biosynthesis pathway [135, 136] . However, lignin content and/or structure are rarely reported in these types of studies. Additionally, these studies have evaluated response shortly after inoculation, and there is no information on the long-term effects of *Septoria* canker response in field-grown poplar. For instance, Bucciarelli et al. documented the differences in Klason lignin content and S/G ratio between susceptible and resistant *P. tremuloides* genotypes in response to *Entoleuca mammata* infection, but within 96 hrs of inoculation [137]. To this effect, an exploratory study of four distinct field-grown *Populus* genotypes exhibiting signs of

canker growth was conducted to determine if changes in lignin content and/or composition induced by *Septoria* infection could be elucidated. Healthy and infected tissue from each genotype was analyzed for Klason lignin content and composition by HSQC NMR. Additionally, whole-cell biomass was analyzed by FTIR. Two healthy genotypes were analyzed in parallel as a control. Wood discs utilized in this study are displayed in Figure 5-1.

5.3 Materials and Methods

5.3.1 Sampling and preparation

Three-year-old *Populus trichocarpa* genotypes were sampled from a field site in Boardman, OR. The study was established in July 2016 using naturally varying genotypes from the *P. trichocarpa* genome-wide association mapping panel (Muchero et al., 2018). Samples of tree stems were taken at 20-30 cm above the soil line the form of approximately 2.5cm thick disks after 3 growing seasons in November 2018. Two hundred fifty-four out of 1,054 trees from this field site were observed to have stem and branch cankers in late summer 2018 – several months before the November 2018 harvest [138]. Wood disks were oven-dried at 70°C for 14 days to remove moisture and prevent dry matter loss due to microbial activity. Five wood disks with noticeable signs of *Septoria* infection were selected for analysis. Two healthy disks were also selected for comparison. *Septoria* infection was previously verified at this field site using visual characterization as well as by isolation and sequencing of *S. musiva* isolates as described by Søndreli et al. [138]. Briefly, wood from the margin between healthy and necrotic tissue was obtained from samples visually observed exhibiting stem canker. Tissue was plated on KV8 medium amended with streptomycin sulfate and chloramphenicol at 100 mg/L and 240 mg/L, respectively. The presence of *S. musiva* was confirmed by comparing the sequence of the ITS region (accession MN275180 to MN275187) to JX901814 with 99% identify. *S. musiva* infection results in a characteristic sunken stem canker that eventually leads to stem breakage. Diseased samples were selected on the basis of the severity of the infection which evident by discolorized wood.

Woody biomass for analysis was obtained by drilling the disks with a 13mm spade drill bit and collecting the wood shavings. Each disk was drilled at three locations: (1) a healthy section with

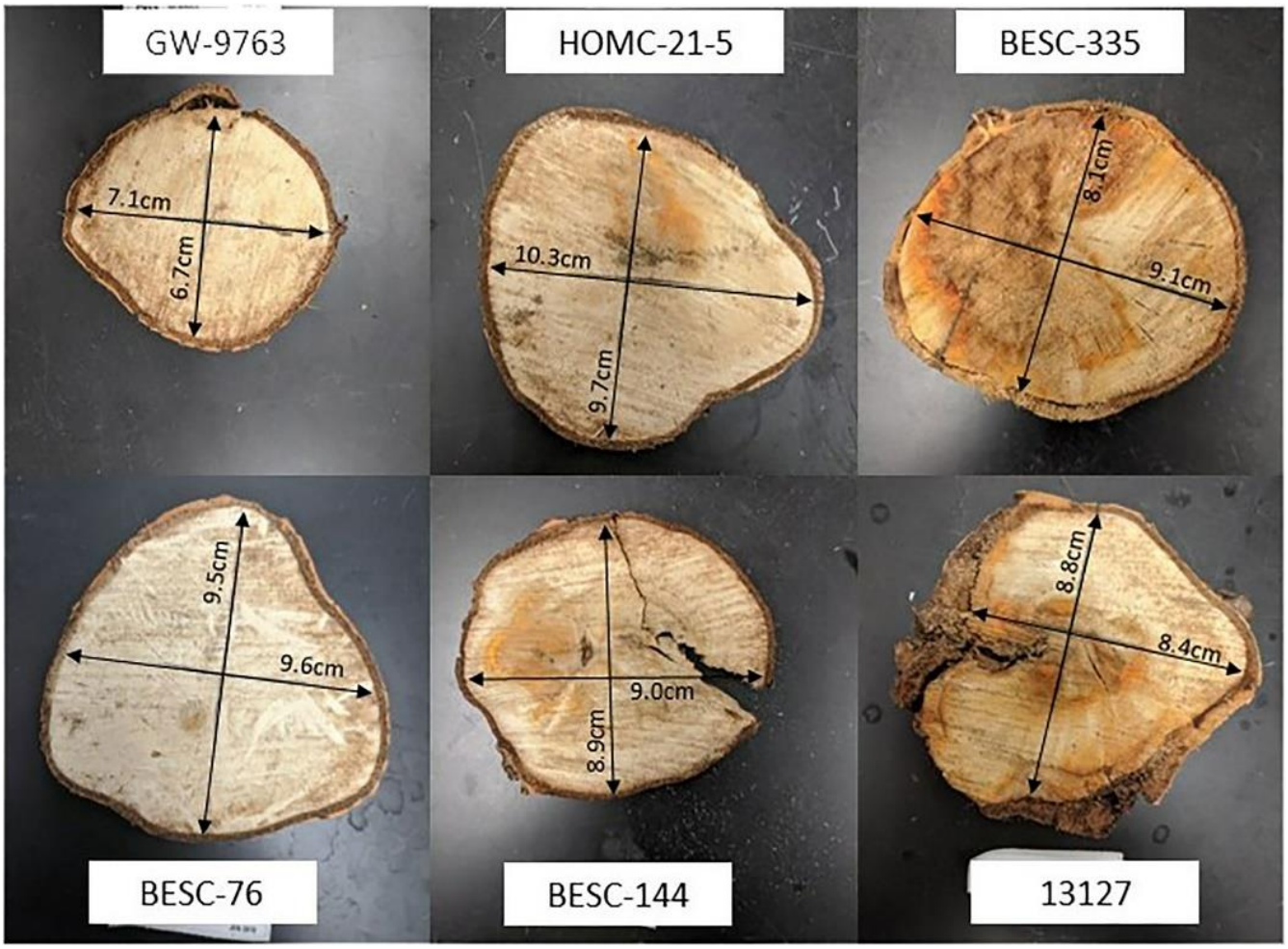


FIGURE 5-1 PICTURES OF WOOD DISCS USED TO EVALUATE EFFECTS OF *SEPTORIA* INFECTION

no signs of *Septoria* infection; (2) a section within the *Septoria* infected area; and (3) a section on the border of the *Septoria* infected area. Each drilling location was selected to be a consistent distance from the pith to avoid the effects of radial variation. Drilling for sample material was done between the pith and the bark. The drill was cleaned between each use to mitigate cross-contamination. Each sample was Wiley-milled using a 40-mesh screen, and the mill was cleaned between each use. Milled wood samples were Soxhlet extracted with toluene/ethanol (2:1, v:v) overnight to remove extractives and then air-dried in a fume hood for at least 48 h. Extractives-free biomass was used in all further analyses.

5.3.2 FTIR

Extracted biomass was analyzed by FTIR via a Perkin Elmer Spectrum 100 FTIR spectrometer and collected via Spectrum software. A background scan was performed prior to sample analysis. Extracted samples were analyzed from 4,000-600 cm-1 with a resolution of 2 cm-1 and 32-scan accumulation. Each sample was analyzed in triplicate, and the three spectra were averaged for further analysis. The resulting spectra were subsequently processed and analyzed with Origin software. All spectra were baseline corrected and normalized at the highest peak (set to 1). PCA modeling was performed within the OriginPro software environment using the spectral data in the fingerprint region of 1800-900 cm-1.

5.3.3 HSQC NMR

To prepare samples for HSQC analysis, extracted biomass was ball milled at 600 RPM for 2 hours to obtain a fine powder. Each sample was then combined with cellulase (Cellulysin) in an acetate buffer (pH 5.0) and shaken on an incubator shake at 35^{o}C for 48h. After centrifugation, the supernatant was discarded, and the solid residue was recovered and washed with DI water three times. The solid residue was then lyophilized for at least 48h to produce enzyme lignin (EL). EL was dissolved in DMSO-d_6 in a 5mm NMR tube, sonicated for 1h, and allowed to swell overnight before analysis. For whole cell wall (WCW) analysis of BESC-335, powdered material was transferred to a 5mm NMR tube for direct dissolution with DMSO-d_6/HMPA-d_{18} (4:1). All NMR spectral data were recorded using a Bruker Avance III HD 500 MHz spectrometer. A standard hsqcetgpsip2.2 Bruker pulse sequence was used with an N_2 cryoprobe with the

following specifications: ^{1}H spectra width of 12 ppm and 1024 data points; ^{13}C spectra width of 220 ppm with 256 increments and 32 scans. All HSQC spectra were analyzed with Bruker TopSpin 3.5pl6 software. The DMSO-d_6 solvent peak at δ_C/δ_H 39.5/2.49 was used to calibrate the spectra.

5.3.4 Klason lignin analysis

The Klason lignin content of each sample was determined based on modified NREL established procedures [139]. Briefly, extracted biomass was first dried overnight at 45C in a vacuum oven. The first hydrolysis was performed with 78% sulfuric in a 30^{o}C block heater for 1 hour with mixing every 5-10 minutes. For the second hydrolysis, the mixture was diluted to 4% sulfuric acid by adding DI water and then autoclaved at 121C for 1 hour. The acid insoluble residue (AIR) and filtrate were separated by vacuum filtration. The AIR was dried at 105C overnight and weighed for Klason lignin determination.

5.3.5 Lignin Composition and analysis by GC-MS

Thioacidolysis was performed on EL residues that were recovered from HSQC NMR analysis. The EL residues recovered from the NMR tube by precipitation through the addition of DI water, centrifugation, and decanting the DMSO-d_6/DIW supernatant. The recovered EL was then lyophilized for 48h. Thioacidolysis was then performed to determine the lignin monomer composition using gas chromatography mass spectrometry (GC/MS), as previously described [140, 141]. Briefly, 12 ml of thioacidolysis reagent was made containing 10.5 ml of 1,4-dioxane, 0.3 ml of boron trifluoride diethyl etherate, and 1.2 ml of ethanethiol. Bisphenol E in 1,4-dioxane was added as internal standard with final concentration of 1.42 mg/ml. The reagent was vortexed, and 500 µl were added to 3 mg of lyophilized EL residues. Samples were then incubated at 100°C for 2 hours and vortexed every 45 minutes. After incubation, samples were cooled to room temperature. A total of 250 µl of the supernatants were transferred to 4 ml vials and 95 µl of saturated $NaHCO_3$ was added. Samples were dried down at 40 °C under a slow flow of nitrogen gas. For derivatization, a pyridine:BSTFA (1:1) solution was made, and 100 µl were added to each sample. Samples were then incubated in an orbital shaker for 30 minutes at 37°C with gentle agitation. Finally, 50 µl of each sample were transferred to GC vials and sent for GC-

MS analysis. GC/MS was performed on an Agilent 6890N GC with a 5973N series MS detector with a DB-5 ms capillary column (30 m × 0.25 mm × 0.25 μm film thickness). Mass spectra were recorded in electron impact mode (70 eV) with 60–650 m/z scanning range. The ions extracted were 239, 269, 299, and 347 m/z for the thioethylated coumaryl (H), coniferyl (G), and syringyl (S) monomers, and internal standard bisphenol E, respectively.

5.3.6 Chemical analysis of PB

The amount of PB was determined by an established method [142] with slight modifications. Briefly, 1mL of 2M sodium hydroxide and 100μL of 1mg/mL *o*-coumaric acid (internal standard) was added to 20mg of extractive-free powder. Samples were incubated at 30ºC for 24h and the reaction was subsequently terminated by the addition of 100μL of 72% sulfuric acid. Samples were then incubated on ice for 5 minutes. The supernatant was collected by centrifugation and filtered through a 0.45μm nylon syringe filter prior to high-performance liquid chromatography (HPLC) analysis.

Samples were analyzed by an Agilent 1200 series HPLC. For each sample, 10μL was injected onto a Symmetry C18 column (4.6 x 50mm, 5μm particle size) maintained at 35ºC. Adequate peak resolution was achieved using 75:25 (v:v) ratio of eluent B (0.1% trifluoroacetic acid in 70:30 acetonitrile:methanol) in eluent A (0.1% triflouroacetic acid in water) at a flow rate of 0.4mL/min. Spectra were integrated at the UV maxima of 255nm and a five-point calibration curve was used for quantification. The PB measurement obtained by HPLC was normalized by lignin content as determined by the Klason lignin method for comparison to HSQC NMR results.

5.4 Results and discussion

5.4.1 FTIR and PCA

FTIR analysis of extracted biomass was measured in the range of 600-4000 cm^{-1}. All samples were run in triplicate, and the three spectra were averaged to account for variability. The averaged spectra were baseline corrected and normalized to the highest peak (set to 1). The individual spectra are plotted in Figure 5-2 for the fingerprint region of 1800-600 cm^{-}. Several peaks typically associated with biomass and lignin are observed in this region. Notably, the large band at 1028 cm^{-1} is associated with C=O stretching vibrations from cellulose and

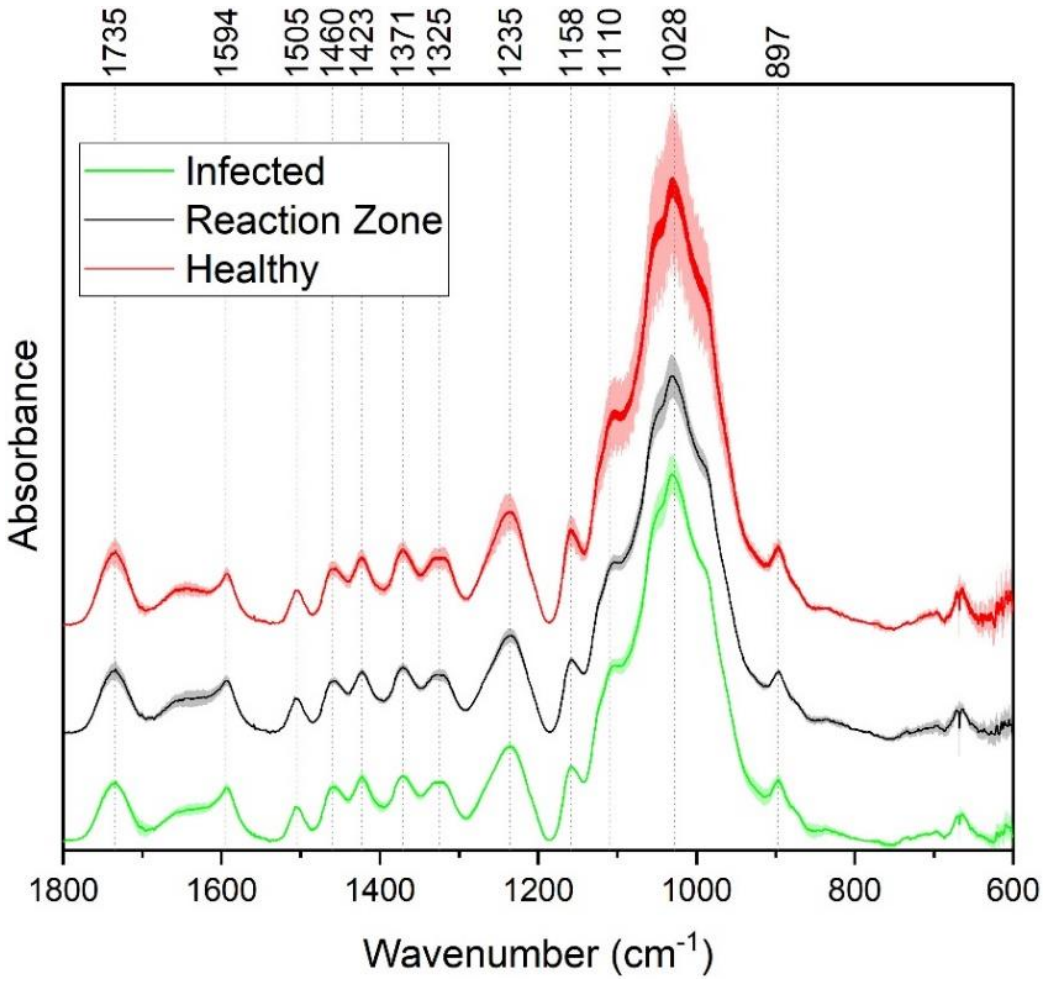


FIGURE 5-2 FTIR SPECTRA OF INFECTED (GREEN), REACTION ZONE (BLACK) AND HEALTHY (RED) REGIONS OF WOOD DISCS

hemicellulose [143]. Another band associated with the C=O stretching in hemicellulose (xylan) was also observed at 1735 cm^{-1} [143].. Peaks around 1158 cm^{-1} and 897 cm^{-1} were observed, corresponding to the C-O-C vibrations of carbohydrates [144] and C-H deformations of G units [145], respectively. Bands characteristic of the aromatic skeletal vibrations of lignin at 1505 cm^{-1} and 1594 cm^{-1} were also observed [143]. Additionally, a peak at 1328 cm^{-1}, associated with syringyl ring breathing with a CO stretching [146], was observed in all samples. These and other peaks of interest are summarized in Table 5-1.

Changes in the peak intensity between healthy, border, and infected tissue can provide insight to the effects of fungal degradation. As summarized in Table 5-2, changes were observed in the intensity of the 1505cm^{-1} peak, which is typically designated as the reference peak for lignin. Sample HOMC-21-5 was observed to have a 1505cm^{-1} ratio of 1.12 and 1.39 between border and infected tissue, respectively, compared to the healthy tissue. Samples BESC-144 and BESC-335 exhibited increased 1505cm^{-1} ratios of 1.70 and 1.50 (respectively) in the border tissue. However, BESC-144 had only a modest 1505cm^{-1} ratio increase of 1.24 in the infected region, and BESC-335 exhibited an almost negligible difference.[147]. The intensity ratio(s) of the 1505cm^{-1} lignin peak to the 1735cm^{-1}, 1369cm^{-1}, 1156cm^{-1}, and 897cm^{-1} polysaccharide peaks were also observed to vary according to tissue type. The intensity ratios of the 1505/1735cm^{-1}, 1505/1369cm^{-1}, 1505/1156cm^{-1}, and 1505/897cm^{-1} peaks were generally higher for the border and infected tissue in samples HOMC-21-5, BESC-144, and BESC-335 (Table 5-3).

The spectral information from the fingerprint region of 1800-600 cm^{-1} was considered for the PCA. The resulting PC1 and PC2 explained 92.6% and 3.4% of the variation, respectively. Subsequent PCs accounted for less than 2% of the variation each. PC loadings were examined to determine which bands contributed to each principal component. For PC1, the band around 1030cm^{-1} was determined to be most significant contributor. This band has been associated with various properties of biomass and lignin, such as the C-O-C bonds in β-O-4 aryl ether linkages and other C-O bonds linked to primary and secondary alcohols. The peak around 1158 cm^{-1}, associated with the C-O-C vibrations in cellulose and hemicellulose, also contributed to PC1. Loading analysis for PC2 revealed that bands around 1640 cm^{-1} and 1735 cm^{-1} contributed most

TABLE 5-1 FTIR BAND ASSIGNMENTS

Observed Peak (cm^{-1})	**Peak Assignment**	**Reference**
1735	C=O stretching in lignin and hemicellulose	[148]
1594	Aromatic skeletal vibrations and C=O stretching in lignin	[148-150]
1505	Aromatic C=C skeletal vibrations in lignin	[148-151]
1460	C-H bending of methyl and methylene groups	[148, 150-152]
1423	C-H deformation, CH_2 bending vibration, carboxyl group stretching	[148-150, 152]
1371	C-H bending, stretching	[151]
1325	C=O stretching of syringyl units	[148, 150-153]
1235	C-C, C-O, and C=O stretching of guaiacyl unit	[148, 150, 153]
1158	C-O stretching of ester group	[149, 151, 152]
1110	Aromatic C-H deformation of syringyl units	[148, 152]
1028	C-O stretching of primary alcohols	[148, 151, 152]
897	C-H deformation vibration of cellulose	[152]

TABLE 5-2 FTIR ABSORBANCE VALUES AT WAVENUMBERS OF INTEREST

Absorbance at Reference Wavenumbers									
Genotype	**Tissue**	**1735 cm^{-1}**	**1594 cm^{-1}**	**1370 cm^{-1}**	**1328 cm^{-1}**	**1235 cm^{-1}**	**1157 cm^{-1}**	**1120 cm^{-1}**	**897 cm^{-1}**
13127	Healthy	1.69	1.25	1.78	1.52	2.65	2.11	3.78	1.91
	Border	1.61	1.33	1.74	1.53	2.59	1.96	3.65	1.85
	Septoria	1.77	1.40	1.85	1.60	2.82	2.06	4.03	1.92
HOMC-21-5	Healthy	2.00	1.46	2.17	1.95	3.08	2.68	4.68	2.16
	Border	2.04	1.46	1.88	1.61	3.00	2.13	4.10	1.70
	Septoria	1.78	1.44	1.83	1.65	2.64	2.09	3.79	1.82
BESC-144	Healthy	2.15	1.41	2.34	2.12	3.36	3.03	5.37	2.57
	Border	1.55	1.52	1.80	1.62	2.59	2.16	3.94	1.76
	Septoria	1.59	1.64	1.79	1.74	2.66	2.18	4.34	1.50
BESC-335	Healthy	2.55	1.62	2.25	2.12	3.70	2.93	5.47	2.32
	Border	2.10	1.67	2.03	1.88	2.82	2.06	4.06	1.72
	Septoria	1.67	1.63	1.96	1.76	2.61	2.12	4.06	1.77

TABLE 5-3 RATIOS OF ABSORBANCE VALUES OF REACTION ZONE AND INFECTED TISSUE RELATIVE TO HEALTHY TISSUE

Wavelength Ratio Compared to Healthy									
Genotype	**Tissue**	**1735 cm^{-1}**	**1594 cm^{-1}**	**1370 cm^{-1}**	**1328 cm^{-1}**	**1235 cm^{-1}**	**1157 cm^{-1}**	**1120 cm^{-1}**	**897 cm^{-1}**
13127	Healthy	-	-	-	-	-	-	-	-
	Border	0.95	1.06	0.98	1.00	0.98	0.93	0.96	0.97
	Septoria	1.05	1.12	1.04	1.05	1.06	0.98	1.07	1.01
HOMC-21-5	Healthy	-	-	-	-	-	-	-	-
	Border	1.02	1.00	0.87	0.83	0.97	0.80	0.88	0.79
	Septoria	0.89	0.98	0.84	0.85	0.86	0.78	0.81	0.84
BESC-144	Healthy	-	-	-	-	-	-	-	-
	Border	0.72	1.08	0.77	0.77	0.77	0.71	0.73	0.69
	Septoria	0.74	1.17	0.76	0.82	0.79	0.72	0.81	0.58
BESC-335	Healthy	-	-	-	-	-	-	-	-
	Border	0.83	1.03	0.90	0.89	0.76	0.70	0.74	0.74
	Septoria	0.66	1.00	0.87	0.83	0.71	0.73	0.74	0.76

significantly to this principal component. These bands are associated with C=O stretching in lignin and xylan, respectively. The PCA output including eigenvalues, loadings, and scores are provided in the supplementary material.

FTIR has been widely utilized as a non-destructive method of rapidly analyzing biomass properties, and the assignments of spectra peaks are well documented. In particular, the intensity of the 1505cm^{-1} lignin reference peak can provide insights to changes in cell wall structure [147]. The 1594cm^{-1} peak is also often assigned as a lignin reference peak, and indeed across all samples the 1594cm^{-1} and the 1505cm^{-1} peaks were highly correlated (R-squared=0.90). The most significant changes in the 1505cm^{-1} peak in relation to tissue type and polysaccharide peaks were observed in samples BESC-144 and BESC-335. Specifically, in the tissue of the border region, the 1505cm^{-1} peaks exhibit the highest intensity, especially in relation to polysaccharide bands. This is an indicator of increased lignin content and has previously been observed in fungal decay [147]. Indeed, a reasonable agreement between the 1505/1735cm^{-1} ratio and Klason lignin content was observed here (Pearson correlation coefficient = 0.45; *p*-value = 0.058).

While FTIR is a convenient analytical technique, trends and differences between samples can often be difficult to elucidate. Therefore, principal component analysis (PCA) was utilized to extract additional information from the spectra. PCA is a valuable dimension reduction technique that emphasizes differences in samples based on spectral variation and has been used extensively for analyzing biomass and lignin (Xu, Yu, Tesso, Dowell, & Wang, 2013). The resulting score plot in Figure 5-3 indicates that border (black) and infected (green) tissue are quite similar. Additionally, PCA does a reasonable job at distinguishing the healthy tissue (red) from the border and infected tissue, even though the healthy tissue of HOMC-21-5 falls just inside the 95% confidence interval (CI) of the border and infected tissue. The exception is sample 13127, where the PCA identifies the healthy tissue to be more characteristically similar to border and infected tissue samples. This aligns with the data from Table 5-3, where there 1505cm^{-1} intensity ratios do not change appreciably with tissue type. These results suggest that sample 13127 exhibits symptoms of *Septoria* infection that are not immediately visible. This is supported by

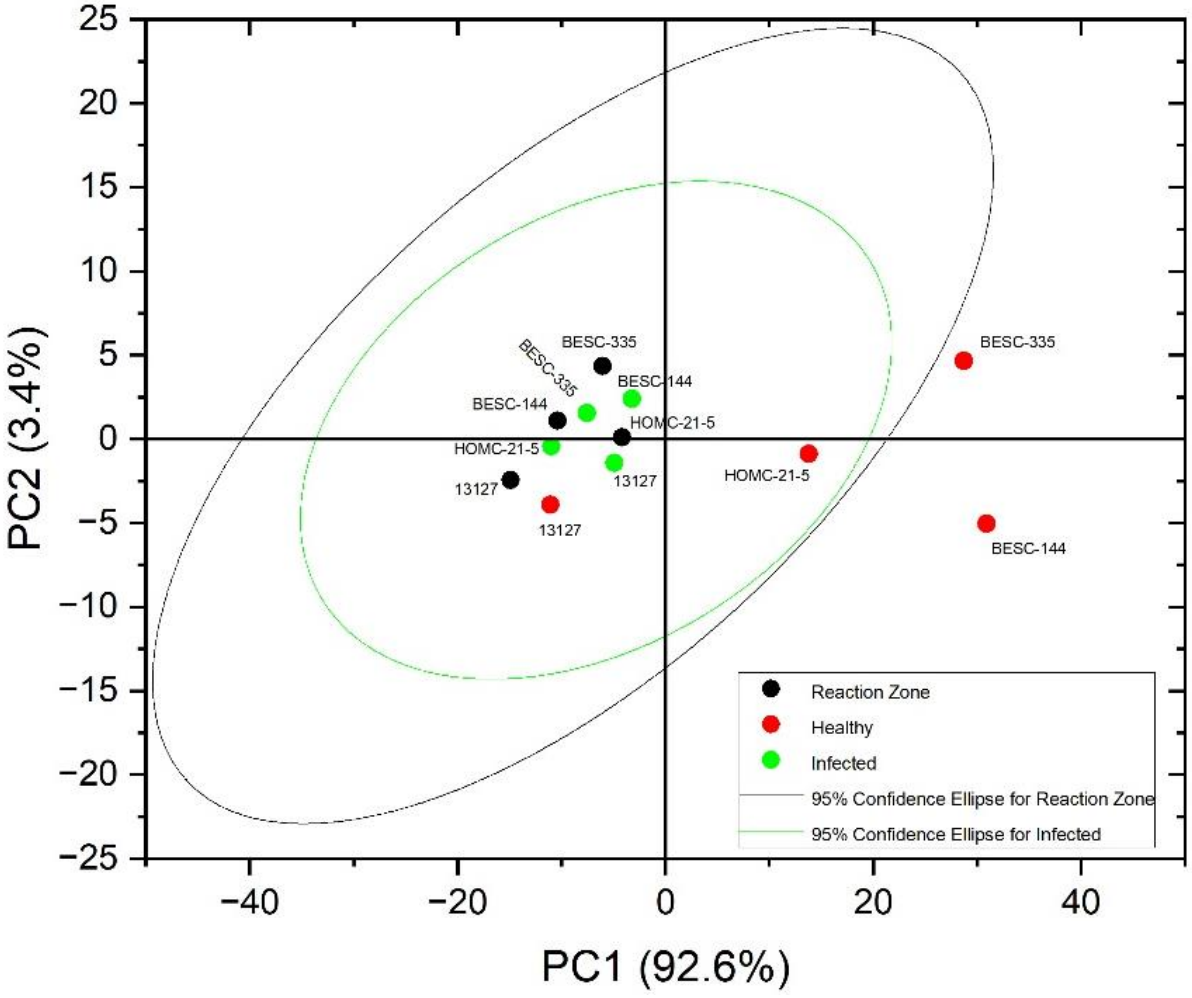


FIGURE 5-3 PRINCIPAL COMPONENT ANALYSIS SCORE PLOT FROM FTIR SPECTRAL DATA

the Klason lignin measurements, as the healthy tissue of 13127 has significantly higher lignin content than the healthy tissue from HOMC-21-5, BESC-144, and BESC-335 (Figure 5-4).

5.4.2 HSQC NMR

Each *Populus* sample was subjected to enzymatic hydrolysis, and 2D HSQC NMR was employed to analyze the resulting enzyme lignin fraction to examine lignin structure and identify differences between healthy and infected tissue. The aromatic region of the HSQC spectra for the healthy and infected tissue for each line is shown in Figure 5-5. All spectra indicated the presence of the expected aromatic syringyl (S), guaiacyl (G), and p-hydroxyphenyl (H) units, as observed in Figure 5-5. The S units exhibited a strong $S_{2,6}$ correlation at δ_C/δ_H 103.8/6.70 ppm. Correlations associated with G units were observed at δ_C/δ_H 111.0/6.98 ppm (G_2), 115.1/6.72, 6.98 ppm (G_5), and 119.1/6.80 ppm (G_6). The $H_{2/6}$ peak was also observed at 128.2/7.17 ppm. Other correlations observed in the aromatic region include *p*-hydroxybenzoate ($PB_{2/6}$) at δ_C/δ_H 130.4/7.62 ppm and cinnamyl alcohol (I_α) at δ_C/δ_H 128.3/6.45 ppm. All spectra also exhibited peaks in the aliphatic region associated with β-O-4, β-5, β-β, and spirodienone interunit linkages.

Relative abundance of each lignin structural feature is summarized in Table 5-4. Samples were independently analyzed by thioacidolysis for lignin composition. The ratio of S units to G units (S/G ratio) determined by thioacidolysis and HSQC NMR are highly correlated (Pearson correlated coefficient = 0.63; *p*-value < 0.01), thereby validating the S/G ratio measurements. The relative content of H units as determined by thioacidolysis and NMR were also compared. The H unit content exhibited poor correlation between the two methods (Pearson correlation coefficient = -0.1; *p*-value = 0.7). It has previously been shown that the HSQC NMR signal for $H_{2/6}$ can be overlapped by various proteins, leading to overestimation of H unit content [154]. We expect this to be the case here. Therefore, careful consideration was given to results depending on the interpretation of H unit content. A comparison of NMR, FTIR, and thioacidolysis results are summarized in Table 5-5.

Similarly, NMR may overestimate PB since it exists as a terminal unit. Therefore, the PB content of a subset of samples was determined via alkaline hydrolysis and HPLC. Due to material availability, not all samples were analyzed. However, the two samples with the highest PB

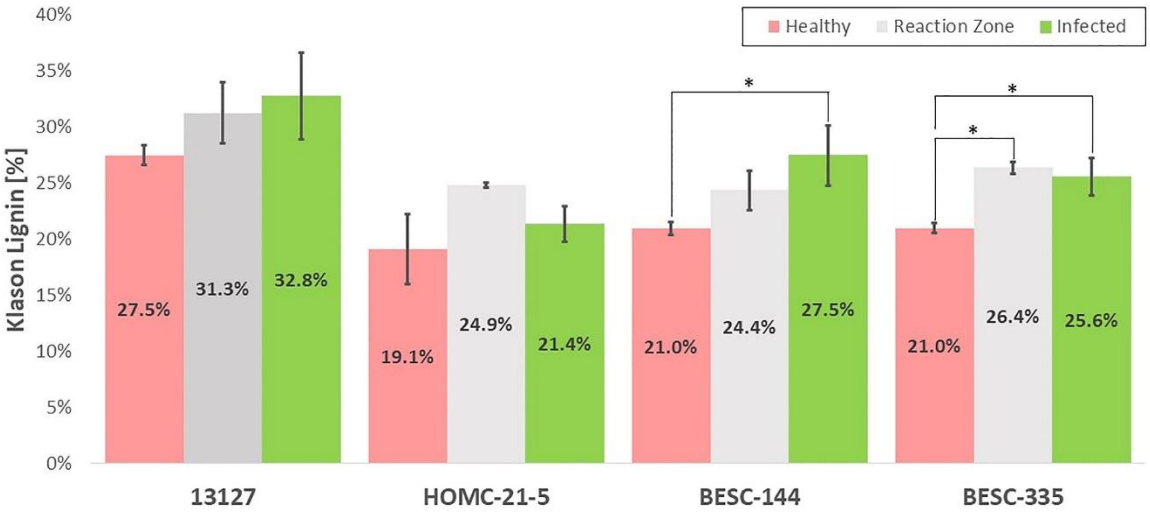


FIGURE 5-4 KLASON LIGNIN CONTENTS OF EACH REGION FROM ALL WOOD DISCS

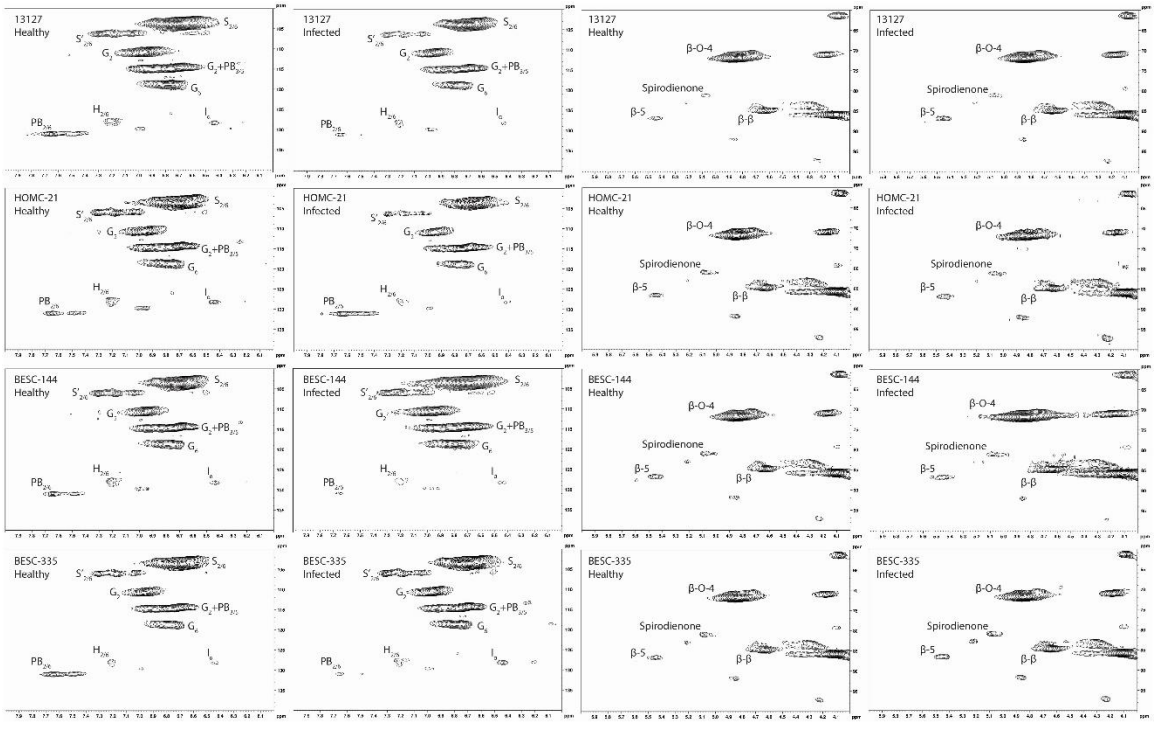


FIGURE 5-5 AROMATIC AND ALIPHATIC HSQC NMR REGIONS OF ALL WOOD DISCS

TABLE 5-4 HSQC NMR RESULTS OF LIGNIN STRUCTURE FROM HEALTHY, REACTION ZONE, AND HEALTHY REGIONS

Sample	Status	S	G	H	PB	S/G	β-O-4	β-5	β-β	Spirodienone
13127	Healthy	67.4	31.5	1.11	4.80	2.14	58.8	3.56	7.57	1.30
	Border	67.7	31.3	1.03	2.01	2.16	60.1	3.19	8.51	1.34
	Infected	68.2	29.9	1.95	2.96	2.28	57.9	3.18	8.14	1.37
HOMC-21-5	Healthy	70.4	28.1	1.52	3.69	2.50	60.4	2.73	7.46	1.64
	Border	69.7	29.6	1.72	2.35	2.35	62.7	2.15	7.78	0.81
	Infected	69.1	29.1	1.81	8.22	2.38	57.2	2.53	5.70	0.85
BESC-144	Healthy	68.5	29.0	2.55	4.26	2.37	56.9	3.02	7.52	1.66
	Border	67.7	30.0	2.36	3.62	2.26	64.7	3.18	8.11	1.56
	Infected	69.4	29.4	1.18	1.98	2.36	61.3	2.87	8.08	1.28
BESC-335	Healthy	70.5	27.6	1.84	6.42	2.55	58.5	3.02	8.20	1.57
	Border	71.1	27.6	1.35	3.94	2.58	58.0	2.48	7.81	1.47
	Infected	71.4	26.6	1.93	1.74	2.68	58.0	2.75	8.49	1.63

TABLE 5-5 COMPARISON OF S/G RATIO AS DETERMINED BY FTIR, NMR, AND THIOACIDOLYSIS

S/G ratio				
Genotype	**Region**	**FTIR**	**NMR**	**Thioacidolysis**
13127	Healthy	0.58	2.14	1.91
	Reaction Zone	0.59	2.16	1.99
	Septoria	0.57	2.28	1.82
HOMC-21-5	Healthy	0.63	2.50	1.97
	Reaction Zone	0.53	2.35	2.03
	Septoria	0.62	2.38	1.93
BESC-144	Healthy	0.63	2.37	2.01
	Reaction Zone	0.63	2.26	1.84
	Septoria	0.66	2.36	1.83
BESC-335	Healthy	0.58	2.55	-
	Reaction Zone	0.67	2.58	1.99
	Septoria	0.68	2.68	2.15

content as measured by NMR (HOMC-21-5 infected, 8.22%; BESC-335 healthy, 6.42%) were included for this analysis. The PB content of each sample was normalized by its associated Klason lignin content. While HPLC measured a lower absolute value of PB (0.73%-1.44%) than NMR (1.61%-8.22%), the two measurements were highly correlated (R^2=0.71; *p*-value=0.002). The alkaline hydrolysis method also identified samples HOMC-21-5 infected and BESC-335 healthy as having elevated PB content (1.44% and 1.22%, respectively). To further validate this measurement, a sample that has previously been utilized as an internal reference (denoted as BESC standard poplar [155]), was also included in this analysis. This sample was run in duplicate via HPLC and in triplicate via HSQC NMR. The PB content of this sample was measured to be 2.13% ± 0.20% by HPLC and 16.27% ± 0.36% by HSQC NMR. This point fits the trendline established by the ten samples from this study that were analyzed, providing greater confidence of the correlation between the two measurements.

To explore potential changes in cell wall polysaccharides, the healthy, border, and infected tissue of sample BESC-335 were subjected to whole cell wall (WCW) HSQC NMR by directly dissolving extractive free, ball milled biomass in a DMSO-d_6/HMPA-d_{18} (4:1) solvent system. Changes in the S/G were observed between the healthy and infected tissue of the more severely infected samples. Genotype 13127 exhibited differences in S/G ratios, measuring 2.14 in the healthy tissue but increased to 2.43 in the infected tissue. The S/G ratio also increased across tissue condition in genotype BESC-335, measuring 2.55, 2.58, and 2.68 in the healthy, border, and infected tissue, respectively. This would be consistent with the observed upregulation caffeate O-methyltransferase (COMT) due to *Septoria* infection **[136]**. COMT is involved in the methylation of monolignols, and its upregulation would increase the ratio of S lignin **[156]**. However, the samples exhibiting lower severity of infection did not exhibit this trend. For the HOMC-21-5 genotype, the S/G ratio was higher in the border tissue (2.83) than the healthy (2.50) or infected (2.38) tissue. The second low severity genotype BESC-144, exhibited a similar S/G ratio in the healthy (2.37) and infected (2.36) tissue, but a decreased S/G ratio (2.26) in the border tissue. All samples exhibited relatively low and consistent amounts of spirodienone (1.38±0.26) and β-5 (2.89±0.38) linkages. As expected, the most abundant interunit linkage in all samples was the β-O-4 aryl ether linkage. The variation of β-O-4 abundance was slightly more

variable (σ=3.01) in the low severity genotypes (HOMC-21-5, BESC-144) than in the high severity (σ=0.86) genotypes (13127, BESC-335). The β-β and spirodienone linkages exhibited a similar trend of higher variability in the low severity genotypes. It has historically been demonstrated that β-O-4 content exhibits a strong positive correlation to the S/G ratio, due in part to the propensity of S units to form this linkage. However, the four *Septoria* infected genotypes samples did not conform to this trend. Additionally, the S/G ratio was not correlated with either the β-β or spirodienone linkages. This is surprising, as both of these linkages have been shown to correlate with syringl unit content **[157]**. The one trend that was observed across these twelve samples was the negative correlation between the S/G ratio on the β-5 content (p-value < 0.001). This correlation is expected, as G lignin can undergo coupling at the 5-position. A lack of correlation between the S/G ratio and all linkage types was observed in the control samples. This could be contributed to relatively low abundance and limited variation in some of the linkage types. For instance, across the twelve samples taken from the infected wood disks, spirodienone linkages ranged from 0.85-1.66% and β-5 linkages ranged from 2.19-3.56% abundance. Additionally, there was generally a neutral or negative correlation between the β-5 and β-β linkages, suggesting there may be a trade-off between the abundance of condensed C-C linkages.

The PB content varied widely among the four genotypes and the tissue type, though the PB content was generally lower in border and/or infected tissue than in healthy tissue. In genotype BESC-144, the PB content of the infected tissue was approximately 50% lower than that of the healthy tissue. Similarly, BESC-335 showed a reduction in PB content from 6.42% in the healthy tissue to 1.74% in the infected tissue. The PB content of genotype 13127 was slightly higher in the infected tissue (2.96%) than the border tissue (2.01%), though both were lower than the healthy tissue (4.80%). However, HOMC-21-5 exhibited the opposite trend. While HOMC-21-5 measured a PB content of 3.69% in the healthy tissue and 1.76% in the border tissue, the infected tissue exhibited a PB content of 8.22%. Many aspects regarding the biosynthesis and function of PB remain a mystery. PB primarily acylates the γ-position of S units, and it has long been hypothesized that PB promotes the formation of S-rich lignin. However, there was no correlation between S and PB abundance (p-value = 0.56). A recent study has also called this hypothesis into question [158]. There is some evidence that PB production is associated with fungal infections

[159], though that contradicts the trend observed here. Frankenstein et al. reported that wounding lead to decreased *p*-hydroxybenzoic acid content [160], which is more consistent with these observations. While there appears to be appreciable PB variability among the twelve samples, based on the current knowledge (or lack thereof) of PB biosynthesis, it is difficult to determine if variation in levels is causative or merely correlative.

In the non-anomeric region (δ_C/δ_H 50-90/2-6) several cellulose signals were observed, including internal cellulose units (CI_4, CI_5, CI_6). Additionally, the non-reducing ends (CNR_3, CNR_5) of cellulose were identified. The non-anomeric region also contained many xylan related signals, such as xylan internal units (XI_2, XI_4, XI_5), reducing ends ($XR\alpha_4$, $XR\beta_4$), and non-reducing ends (XNR_2). Hardwoods contain acylated xylan, and associated acetylated xylan signals (2-*O*-Ac-β-D-Xyl, 3-*O*-Ac-β-D-Xyl) were also observed. In the anomeric (δ_C/δ_H 90-105/3.5-6) region, the signal for internal cellulose [(1→4)-β-D-Glc*p*] was observed, though the signal for the cellulose non-reducing end [(1→4)-β-D-Glc*p* (NR)] was generally better resolved. The signal associated with internal xylan units [(1→4)-β-D-Xyl*p*] was also prominent. 4-O-methyl-α-D-glucuronic acid (4-O-MeGlcA) was readily observed in the border and infected tissue. This signal was present in the healthy tissue, though at levels very close to background. In summary, while some differences between the different regions were observed, the healthy, border, and infected tissue appear to largely contain similar cellulose and hemicellulose structures.

5.4.3 Klason lignin analysis

The Klason lignin content the health, border, and infected tissue of each sample was determined gravimetrically after two-step acid hydrolysis [139]. As depicted in Figure 5-4, the lignin content was generally elevated in the border and/or infected tissue compared to the healthy tissue. Both genotypes 13127 and BESC-144 exhibited consistent increases in lignin content from the healthy to infected regions. The lignin content of genotype 13127 increased from 27.48% in the healthy region to 31.27% and 32.81% in the border and infected regions, respectively. Similarly, the lignin content of genotype BESC-144 was measured to be 20.96% in the healthy region, 24.38% in the border region, and 27.51% in the infected region. BESC-335 also exhibited higher lignin content around the infection site (25.55%) and border region (26.39%) than the healthy region, which measured 21.00% lignin. While the infected tissue HOMC-21-5 exhibited lignin content

(21.35%) similar to the healthy tissue (19.12), the associated border tissue was measured to have a much higher lignin content of 24.87%.

The expected general trend of increased lignification was observed in each of the four infected *Populus* genotypes in this study by analyzing the Klason lignin content of the healthy, border, and infected region of each genotype. It has been previously shown that *Populus* forms a lignin rich periderm around the site of *Septoria* infection for containment. This is consistent with the trend observed in Figure 5-4, with the border and/or infected tissue generally having a higher Klason lignin content than the healthy tissue. The most pronounced difference in Klason lignin content was observed in BESC-335, which exhibited the most severe and consistent degree of infection (Figure 5-1). In this sample, both the border and infected tissue exhibited significantly higher lignin content (p-value < 0.05) than the healthy tissue. Similarly, sample BESC-144 exhibited higher lignin content in the infected region. These results signify the recruitment of lignin (or possibly lignin-like phenolics) toward a periderm to contain the fungal infection. Genotype 13127 exhibited a general increase in lignin content from healthy to border to infected tissue. However, this increase was determined to not be statistically significant. This is primarily driven by the high variability of the infected tissue (σ=3.9%) compared to the healthy tissue (σ=0.9%). Genotype HOMC-21-5 displayed a trend slightly different from the other three lines. The border region had the highest lignin content, while the healthy and infected regions exhibited lower and similar lignin content. From Figure 5-1, it appears that HOMC-21-5 had the lowest severity of infection. Poplar species exhibit various degrees of susceptibility toward *Septoria* infection. For instance, *P. deltoides* tends to less susceptible to infection, whereas *P. maximowiczii* tends to be more susceptible to infection [161, 162]. HOMC-21-5 may be a genotype less susceptible to Septoria infection by efficient recruitment of lignin to form a periderm to contain an infection, resulting at higher lignin content at the border between the healthy and infection tissue. Still, this result contrasts the trends of the other three lines, where the infected regions exhibited notably higher lignin content than the healthy regions. However, these results align with the NMR results, as the border sample exhibited both the highest S/G ratio (2.83) and the highest lignin content (24.87%) of the HOMC-21-5 line. Likewise, the healthy and infected regions had lower S/G ratios (2.50, 2.38), corresponding to their lower

lignin content (19.12%, 21.36%). The area of the infected site could explain this outcome. HOMC-21-5 exhibited a less severe degree of infection which may have impacted the ability to collect the proper tissue. Additionally, the infection may not have been consistent throughout the thickness of the wood disks. It is therefore possible that healthy tissue may have been present in the infected material of HOMC-21-5 which would have an impact on the measurement. It is known that *Populus* will develop necrophylactic periderm (NP) layers around the sites of fungal infections to prevent the pathogen's spread (Qin & LeBoldus, 2014). These NP layers are created by depositing lignin and lignin-like phenolic compounds such as suberin (Mullick, 1977). A study of poplar confirmed upregulation of several lignin biosynthesis genes after being infected with *Sphaerulina*, including CoA 3-O-methyltransferase (CCoAMT), cinnamoyl CoA reductase, (CCR), and cinnamyl alcohol dehydrogenase (CAD) [136]. As such, the lignin content of infected regions is expected to be higher than in the healthy regions. However, to our knowledge, increased lignification due to *Septoria* infection in poplar has not been analyzed by quantitative methods. These FTIR and PCA results confirm that there are clear changes to cell wall structure induced by *Septoria* infection.

5.5 Conclusion

The fungus *Septoria musiva* poses a serious threat to the productivity of economically important poplar. While various studies have explored short-term changes in the days following inoculation, the long-term phenotypic changes of lignin in field-grown, naturally infected poplar stems from *Septoria* canker have not been previously reported. FTIR spectroscopy indicated there were there were likely changes to lignin content and/or structure between healthy and border/infected tissue, though trends were not consistent across all samples. Subsequent PCA of whole cell wall biomass samples identified that border and infected tissue are similar, and can be distinguished from healthy samples. Likewise, two samples were found to have significantly higher lignin content in the infected tissue, and one of these also had significantly higher lignin content in the border tissue. HSQC NMR examined differences in lignin structure. While differences in lignin structure were observed, there did not appear to be a uniform trend between tissue types. Two samples with elevated PB content were detected by NMR, which was validated by alkaline hydrolysis and HPLC. Whole cell wall HSQC NMR analysis of one sample suggest

that polysaccharide structures remain similar across tissue types Additionally, the Klason lignin content trended well with the S/G ratio. These results corroborate the observations of previous studies, which noted differentially expressed lignin biosynthesis or increased lignification via staining after *Septoria* inoculation. A better understanding of lignin response to fungal infections can lead to improved resilience and higher biomass yield of *Populus*, which is critical for the successful implementation of biofuel production.

Chapter 6:

Variable lignin structure revealed in *Populus* leaves

Sections of this chapter was originally published as:

Nathan Bryant[a], Nancy Engle[b], Timothy Tschaplinski[b], Yunqiao Pu[b], Arthur J. Ragauskas[a,b,c,*](2023). Variable lignin structure revealed in *Populus* leaves *RSC Advances*, *13*(29), 20187–20197. https://doi.org/10.1039/d3ra03142j

Arthur J. Ragauskas conceived the study. Timothy Tschaplinski and Nancy Engle harvested the leaves. Nathan Bryant, Yunqiao Pu, Timothy Tschaplinski, and Nancy Engle performed experiments and analyzed the data. Nathan Bryant drafted the manuscript. All authors reviewed, revised, and approved the manuscript.

6.1 Abstract

Lignin has long been a trait of interest, especially in bioenergy feedstocks such as *Populus*. While the stem lignin of *Populus* is well studied, foliar lignin has received significantly less consideration. To this end, leaves from 11 field grown, natural variant *Populus trichocarpa* genotypes were investigated by NMR, FTIR, and GC-MS. Five of these genotypes were sufficiently irrigated, and the other six genotypes were irrigated at a reduced rate (59% of the potential evapotranspiration for the site) to induce drought treatment. Analysis by HSQC NMR revealed highly variable lignin structure among the samples, especially for the S/G ratio, which ranged from 0.52-11.9. Appreciable levels of a condensed syringyl lignin structure were observed in most samples. The same genotype subjected to different treatments exhibited similar levels of condensed syringyl lignin, suggesting this was not a response to stress. Strong amino acid signals were also observed in the aromatic region of the NMR spectra. Principle component analysis revealed that FTIR absorbances associated with syringyl units (830 cm^{-1}, 1317 cm^{-1}) greatly contributed to variability between samples. Additionally, the ratio of 830/1230 cm^{-1} peak intensities were reasonably correlated (*p*-value=0.1) with the S/G ratio determined by NMR. Analysis by GC-MS revealed significant variability of secondary metabolites such as tremuloidin, trichocarpin, and salicortin. Additionally, salicin derivatives were found to be well correlated with NMR results, which has been previously hypothesized. These results highlight previously unexplored nuance and variability associated with foliage tissue of poplar.

6.2 Introduction

Lignin is vital for many biological functions such as tissue development, water transport, and stress tolerance [163]. The lignin polymer is typically described as having three primary monolignols: sinapyl alcohol, coniferyl alcohol, and *p*-coumaryl alcohol [164]. These phenylpropane units, are produced through the general phenylpropanoid and monolignol specific pathways, and are subsequently incorporated into the lignin polymer as syringyl (S), guaiacyl (G), and p-hydroxyphenyl (H) units, respectively [165].

Populus is poised to be a promising resource for renewable biofuels and bioproducts due to favorable characteristic such as fast growth and high cellulose content [166]. The cell wall biopolymer lignin from *Populus* stem tissue has been extensively characterized due its importance in applications such as bioenergy [57]. However, less is known about lignin associated with leaf tissue. Most of the studies on this topic describe the lignin content [167-169] but not its structure.

Leaves produce a host of secondary metabolites, including lignin. However, from our literature search, there are few, if any, studies which offer detailed examination of the lignin structure from natural variant *Populus* foliage. This represents a knowledge gap, as there is potentially interesting lignin chemistry occurring in leaves. Indeed, lignification in leaves has recently become a topic of interest [170, 171]. It is well documented that lignin traits vary between biomass type (i.e., hardwood vs. softwood vs. herbaceous) [172]. Additionally, differences in lignin content, structure, and enzyme activity have been observed between tissue type, such as between stem and leaf [173, 174]. Abiotic stress has also been demonstrated to alter lignin structure. For instance, poplar leaves exposed to ozone were observed to produce condensed lignin structure [175]. Foliage lignin is also important to ecosystems, as it has been shown to play a significant role in leaf litter decomposition [176, 177]. We have therefore characterized foliage samples harvested from 11 *Populus trichocarpa* natural variant genotypes from a Genome-wide Association Study population at Boardman, OR (3-year-old trees established in 2016) by HSQC NMR, FTIR, and GC-MS, with an emphasis on examining

6.3 Materials and Methods

6.3.1 Plant growth and sampling

Foliage samples were harvested from three-year-old, natural variant *Populus trichocarpa* in July 2018. Samples were named based on the following convention. The first part (i.e., "1121") denotes the genotype. The second part denotes whether the samples was grown in the north end ("-N") or south end ("-S") of the field. The field site near Boardman, OR was established in 2016. During growth, this area experienced a natural environmental drought. The field was segregated into north and south regions. The north end was irrigated and therefore samples from the north end (denoted "-N") of the field are designated as non-drought (control) samples. The south end (denoted "-S") was irrigated at 59% of potential evapotranspiration (PET) for Boardman, OR in 2017 and this treatment regime was maintained through 2020, with this treatment designated as the drought samples. Irrigation delivered to the control treatment targeted 100% PET. The south end had been subjected to drought for 2 growing seasons (2017, 2018) before sampling in the middle of the third growing season (July 2018).

6.3.2 Metabolite sampling and characterization

One mature leaf was collected from each tree. Leaves were collected from the same side of each tree to minimize differences in sun exposure. Leaf samples were immediately fast-frozen in the field prior to shipment on dry ice and stored at -80 ^{o}C until they were lyophilized and ground to a powder using a SPEX Sample Prep Geno/Grinder. Metabolites were extracted from ~25 mg of powdered tissue twice overnight with 2.5 mL of 80% ethanol. Sorbitol (75 µl of 1 mg/ml aq.) was added as an internal standard to the first day extract as an internal standard. The extracts were combined and a 500 µl aliquot was dried under nitrogen and silylated to generate trimethylsilyl (TMS) derivatives, as described previously [178, 179]. After 2d, 0.1 µl aliquots were injected into an Agilent Technologies Inc. (Santa Clara, CA, USA) 7890A/5975C inert XL gas chromatograph-mass spectrometer fitted with a Restek Rtx-5MS with Integra-guard (5 % diphenyl/95 % dimethyl polysiloxane) 30 m × 250 µm × 0.25 µm film thickness capillary column and operated using conditions previously described [178]. A large user-created database of mass spectral electron impact ionization (EI) fragmentation patterns of TMS-derivatized

compounds as well as the Wiley Registry 10th Edition was used to identify metabolites of interest to be quantified. Metabolite peaks were extracted using a key selected ion, characteristic mass-to-charge (m/z) ratio, scaled back to total ion current using pre-determined scaling factors and normalized to the quantity of internal standard recovered, mass extracted, volume analyzed, and injection volume.

6.3.3 Lignin isolation and analysis

After lyophilization and grinding to a powder using a SPEX Sample Prep Geno/Grinder, samples were Soxhlet extracted with toluene:ethanol (2:1, v:v) overnight to remove extractives. We make the distinction that this toluene:ethanol extraction differs from the 80% ethanol extraction described in the previous section, and aliquots were only subjected to one extraction unique to the intented analysis. After Soxhlet extraction, samples were air dried for at least 24 h. Air dried samples were then ball milled for two hours at 600 RPM, at five-minute intervals to avoid overheating. The ball milled samples were subjected to enzymatic hydrolysis with a cellulase enzyme for 48 h in a sodium acetate buffer to digest polysaccharides. After enzymatic hydrolysis, samples were centrifuged and washed with DI water two times to remove the cellulase enzyme. The resulting enzyme lignin was lyophilized for at least 48 h to remove residual moisture.

Enzyme lignin was analyzed via 2D HSQC NMR with a Bruker Avance III 500-MHz spectrometer. Approximately 40mg of lignin enriched residue was dissolved in DMSO-*d6* in a 5 mm NMR tube and subsequently sonicated for 1 h. The standard hsqcetgpsip2.2 Bruker pulse sequence was used with an N2 cryoprobe with the following specifications: ^{1}H spectra width of 12 ppm and 1024 data points; ^{13}C spectra width of 220 ppm with 256 increments and 32 scans. All HSQC spectra were analyzed with with Bruker TopSpin 3.5pl6 software. The DMSO-*d6* solvent peak at δ_C/δ_H 39.5/2.49 was used to calibrate the spectra, and the $S_{2/6}$ signal was used at an internal standard. The signals used for volume integration are as follows: δ_C/δ_H 103.8/6.70 ppm for $S_{2/6}$, δ_C/δ_H 111.0/6.98 ppm for G_2, δ_C/δ_H 128.2/7.17 ppm for $H_{2/6}$, δ_C/δ_H 130.4/7.62 ppm for $PB_{2/6}$, δ_C/δ_H 128.3/6.45 ppm for I_α, δ_C/δ_H 71.8/4.86 for β-O-4, δ_C/δ_H 86.8/5.46 for β-5, and δ_C/δ_H 84.8/4.65 for β-β. All results are presented on an S+G basis and should be interpreted as

abundance per 100 aromatic units. Sample 856-S was selected for a whole cell wall NMR analysis. Approximately 40mg of toluene:ethanol extracted and ball milled material was dissolved directly in a co-solvent system of DMSO-d*6*/HMPA-d*18* (4:1, v:v) in an NMR tube and analyzed by the same method discussed above.

6.3.4 FTIR and PCA

Leaf tissue that had been Soxhlet extracted with toluene:ethanol was analyzed via FTIR with a Perkin Elmer Spectrum 100 FTIR spectrometer. After running a background scan, each sample was analyzed from 600 cm^{-1} to 4000 cm^{-1} with a 32-scan accumulation and a 2 cm^{-1} resolution. Data was collected with Spectrum software. The resulting spectral and principal component analysis was performed with Origin Pro software.

6.4 Results and Discussion

6.4.1 HSQC NMR

Heteronuclear single quantum coherence (HSQC) nuclear magnetic resonance (NMR) was utilized to characterize the isolated lignin structure from the leaf tissue. Specifically, the aromatic region (δ_C/δ_H 100-140/6.0-8.0) of the spectra was examined for monolignol content, and the aliphatic region (δ_C/δ_H 100-60/6.0-4.0) was examined for interunit linkage content. Samples were observed to have appreciable S units as indicated by the $S_{2/6}$ peak (δ_C/δ_H 103.7/6.71). All samples exhibited a signal at (δ_C/δ_H 110.7/6.98), indicating the presence of G units. The S/G ratio ranged from 0.41 to 11.9, which is a surprisingly high degree of variability (Figure 6-1). For all samples, the β-O-4 aryl ether linkage was the most abundant, as indicated by the signal at (δ_C/δ_H 71.8/4.86). The β-O-4 linkage is usually positively correlated with the S/G ratio.However, this relationship was not observed here. β-5 and β-β linkages were observed at (δ_C/δ_H 86.8/5.47) and (δ_C/δ_H 84.8/4.64), respectively, but at levels significantly lower than β-O-4 linkages.

Most samples also contained condense S units as indicated by the signal around δ_C/δ_H 6.32/106.4. However, condensed G units were not observed, suggesting that S units are selectively condensed. In a previous study, the presence of condensed monolignols was observed through decreased thioacidolysis yields [175]. However, thioacidolysis works by cleaving the liable β-O-4 ether bond, and therefore highly condensed lignin structures may bias results.

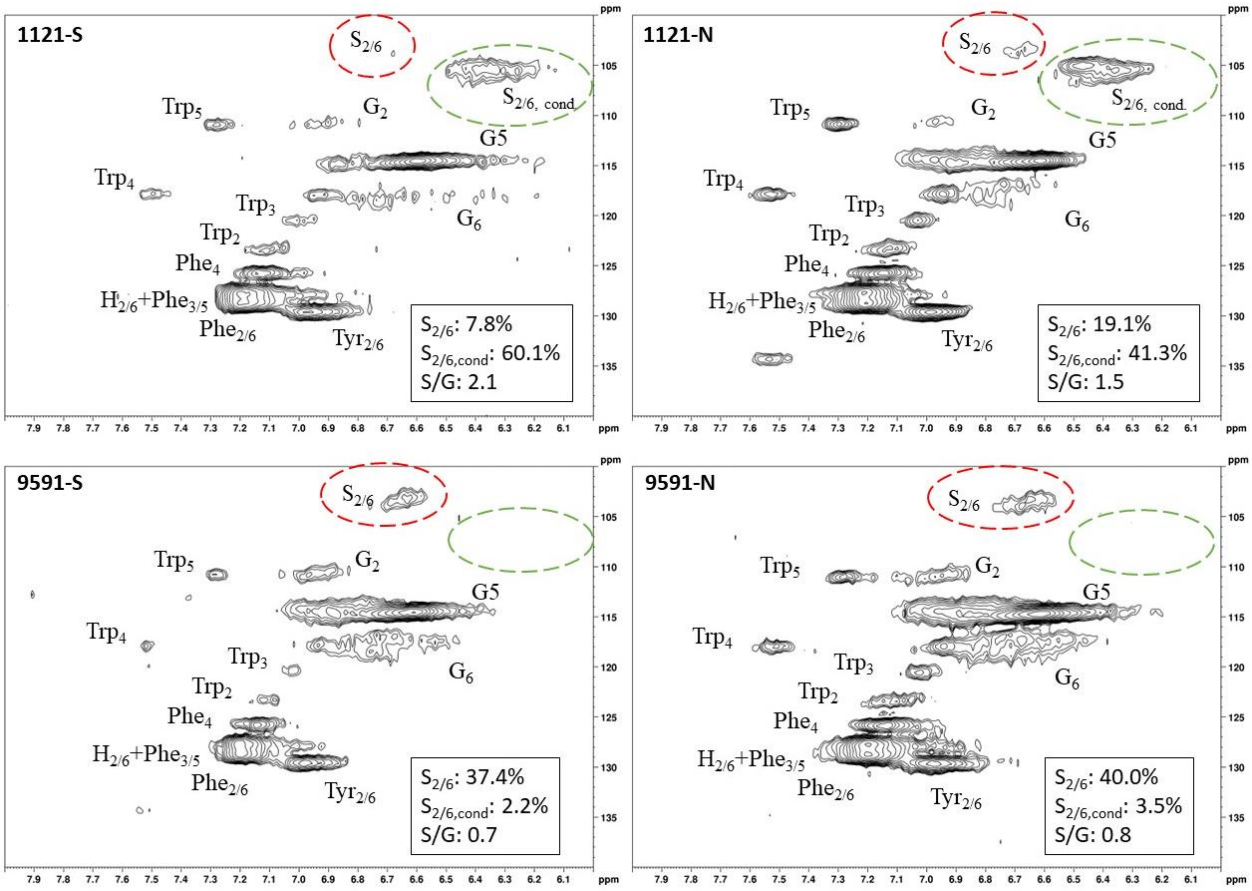


FIGURE 6-1 ALIPHATIC AND AROMATIC REGIONS OF TWO GENOTYPES REPLICATED IN THE DROUGHT AND IRRIGATED (CONTROL) REGIONS

Indeed, Cabané et al. reported that the proportion of S units decreased as condensed units increased (i.e., thioacidolysis yield decreased). While the referenced study introduced ozone stress which induced condensed lignin, it would be interesting to evaluate whether only condensed S units were present, or if all monolignols exhibited condensed structures. This observation highlights a unique beneficial utility offered by HSQC NMR.

Initially, there appears to be an abundance of H units as indicated by the signal at (δ_C/δ_H 127.9/7.19). However, upon closer inspection, this instead resembles characteristic amino acid contamination that has been described in HSQC NMR spectra [180]. To avoid erroneous quantification of H units due to contamination, this signal was not included in the semi-quantitative evaluation of lignin structure. One amino acid that appears prominently in the aromatic region is phenylalanine (Phe). Phenylalanine is a precursor to many secondary metabolites and has been identified as a response to drought stress [181, 182]. Though it did not interfere with lignin signals, tryptophan (Trp) was also observed in the spectra. Increased levels of tryptophan helps regulate osmotic balance in response to drought stress [183]. Tyrosine (Tyr) has also been shown to be associated with drought stress tolerance [184]. Strong phenylalanine, tryptophan, and tyrosine signals were present in all samples. These amino acids also play a variety of other roles in plant growth and development. The prominence of amino acid signals in these leaf spectra also illustrate the difference in tissue type, as stem tissue typically registers very low or zero amino acid levels. The H unit levels, which suffer most prominently from amino acid contamination, are not included in the reported HSQC NMR results in Table 6-1.

HSQC NMR revealed tremendous variability among the 11 samples. Perhaps the most striking observation is the degree of variability exhibited in the S/G ratio. While the actual S/G level may vary by analytical method, the S/G ratio of *Populus* stem lignin is generally reported in the range of 1.5-2.5 [57]. These 11 foliage tissue samples exhibited a significantly wider range of S/G ratios, ranging from 0.33 to 9.79. From Figure 6-2, the $S_{2/6}$ signal of sample 9860 (S/G=0.73) is distinguishable at only slightly higher than background levels. However, in sample 1031 (S/G=9.79), the $S_{2/6}$ signal is very prominent and clearly more abundant than the G_5 signal.

TABLE 6-1 RESULTS OF LIGNIN STRUCTURAL TRAITS

Sample	S	S, cond	S, total	G	S/G	β-O-4	β-5	β-β
9589-S	1.11	59.8	60.9	39.1	1.56	22.1	4.24	1.97
1181-S	2.33	71.5	73.9	26.1	2.83	50.3	2.75	1.19
1121-S	7.82	60.1	67.9	32.1	2.12	17.9	2.47	1.73
856-S	4.01	85.8	89.8	10.2	8.83	83.2	13.1	4.18
LILC-26-4-S	31.7	5.61	37.3	62.7	0.59	63.3	5.53	2.92
9591-S	37.4	2.2	39.6	60.4	0.70	63.6	2.9	3.5
9860-N	23.1	11.5	34.5	65.5	0.53	26.6	5.17	1.88
1031-N	0.38	91.9	92.3	7.74	11.9	17.8	1.93	0.62
9953-N	33.8	4.08	37.8	62.2	0.61	43.5	6.15	2.87
1025-N	26.5	20.2	46.7	53.3	0.88	74.5	10.3	2.96
425-N	33.0	1.22	34.2	65.8	0.52	43.5	4.28	3.67

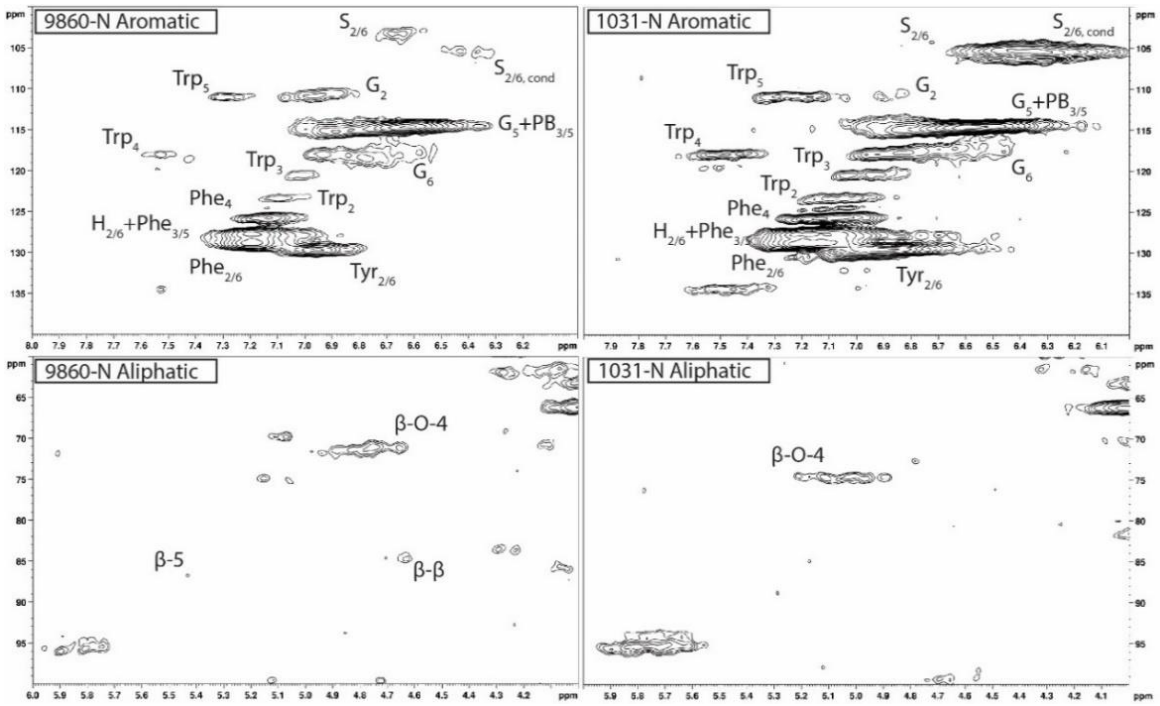


FIGURE 6-2 ALIPHATIC AND AROMATIC HSQC NMR REGIONS OF TWO GENOTYPES

As HSQC NMR is a semi-quantitative technique, these absolute differences are difficult to distinguish here. However, Cabane et al. quantified foliar lignin S/G ratio of *Populus tremula x alba* to be approximately 0.6 by thioacidolysis [175], which compares well with approximately half of the samples analyzed in this study. The β-O-4 linkages also exhibited striking variability across these foliage samples. In stem lignin the β-O-4 bond is the most abundant linkage, with reported values typically in the range of 60%-65%. The foliage samples here exhibited β-O-4 content as low as 26.1% and as high as 73.9%, with only two samples (LILC-26-4-S and 9591-S) exceeding 60%. This is consistent with the findings of Cabane et al. [175], who hypothesized the decreased β-O-4 linkage content was due to the lower S/G ratio. Typically, the β-O-4 linkage content is observed to be positively corrected with the S/G ratio [52]. However, a negative trend between these two phenotypes is observed here. It is unclear what factors may be contributing to this observed variability of these natural variant, field-grown samples. Similar variability was observed in both the south (drought) and north (control) sample sets, so treatment is likely not a contributing factor. One potential explanation may be that lignin structure varies by leaf anatomy. It is well documented that lignin can vary by tissue type (i.e., leaf vs. stem). In this case, petiole and/or midrib of the leaf may have a lignin structure different from the lamina (i.e., higher S/G ratio), and varying ratios of petiole:midrib:lamina could explain some of the observed variability. While these samples consist of enzyme lignin, another commonly utilized strategy for HSQC NMR analysis is whole cell wall (WCW) analysis, wherein the whole cell wall (after extraction and ball milling) is directly dissolved in the NMR solvent, therefore bypassing the enzymatic hydrolysis step. Sample 856-S was selected for whole cell wall WCW NMR analysis as a comparison to the isolated enzyme lignin. The typical $S_{2/6}$ and G_2 signals were noticeably absent from the WCW spectra but were observed to be well resolved in the enzyme lignin spectra. It is expected that these results are influenced by the low lignin content associated with leaf tissue, which has been measured to be approximately 10% [170]. The enzymatic hydrolysis procedure allows a more lignin rich residue to be analyzed which improves the corresponding signals. Samples were also subjected to two step acid hydrolysis to determine Klason lignin content. However, this procedure produced suspect results, and similar difficulties with Klason lignin measurement of foliage tissue has been previously documented [185].

6.4.2 FTIR

Fourier-transform infrared (FTIR) spectroscopy is an analytical tool widely used to identify the functional groups of many compounds, including lignin. Care must be taken when interpreting FTIR spectra of biomass, as bands can be attributed to multiple biomass components. Figure 6-3 displays the FTIR spectra of the toleuene:ethanol extracted samples, which will contain cell wall components besides lignin (i.e., sugars). Prominent bands in the FTIR spectra are summarized in Table 6-2 and were assigned based on comparison with existing published literature [186-191]. All samples exhibited a wide band centered around 3290 cm^{-1}, which is associated with O-H stretching. Peaks were also observed at 2920 and 2850 cm^{-1}, corresponding to C-H vibration of CH_2 and CH_3 functional groups. Figure 6-3 displays the absorbance spectra of the fingerprint region of 1800-800 cm^{-1}, as this region typically provides the most information regarding cell wall structure. The peak at approximately 1735 cm^{-1} represent is typically associated with xylan and has also been associated with stretching vibration of nonconjugated and conjugated ketones.. The peak at 1440 cm^{-1} corresponds to asymmetric C-H deformations. The 1515 cm^{-1} peak is attributed to lignin aromatic skeletal vibrations.

The FTIR spectra were further analyzed by principal component analysis (PCA). PCA is a useful mathematical procedure for analyzing data. The goal of applying PCA to FTIR is to transform a large data set (i.e., thousands of data points from FTIR spectra) into a few key parameters called principal components (PCs). The resulting PCs are typically characterized by the amount of variation they represent, with PC1 accounting for the most variation, PC2 accounting for the second most variation, and so forth. Additionally, PC scores are assigned to each sample, and samples that have similar spectra will be scored similarly. The final result is that PCA reduces the FTIR spectra of many samples down to two dimensions (typically PC1 and PC2), which can then be conveniently represented on a familiar X-Y plot. Samples with similar PC scores will be plotted in close proximity, revealing cluster patterns and allowing differentiation between different groups. The resulting PC1 and PC2 scores from each sample are presented in the score plot in Figure 6-4. All spectra were baseline corrected, normalized from [0, 1], and the second derivative of the spectra were taken. Data in the fingerprint region of 1800-800 cm^{-1} were considered for PCA.

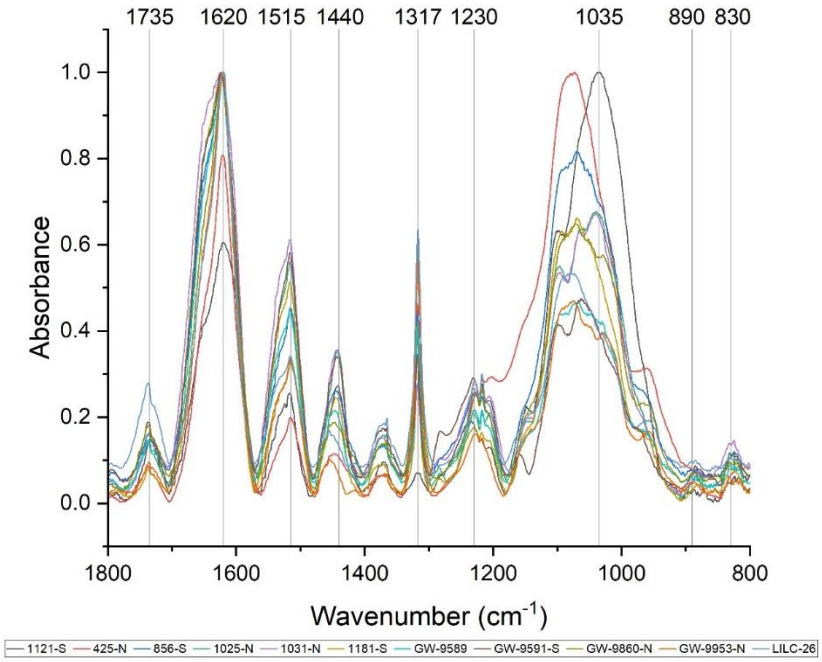


FIGURE 6-3 FTIR SPECTRA OF WHOLE LEAF BIOMASS

TABLE 6-2 FTIR PEAK ASSIGNMENTS

Observed Peak (cm^{-1})	**Peak Assignment**	**Reference**
1735	C=O stretching in lignin and hemicellulose	[148]
1620	C=O stretching	[192]
1515	Aromatic C=C skeletal vibrations in lignin	[148-151]
1440	O-H in-place deformation in cellulose	[192]
1318	C=O stretching of syringyl units	[148, 150-153]
1230	C-C, C-O, and C=O stretching of guaiacyl unit	[148, 150, 153]
1035	C-O stretching	[148, 151, 152]
890	C-H deformation vibration of cellulose	[152]
830	C-H Bending of syringyl units	[193]

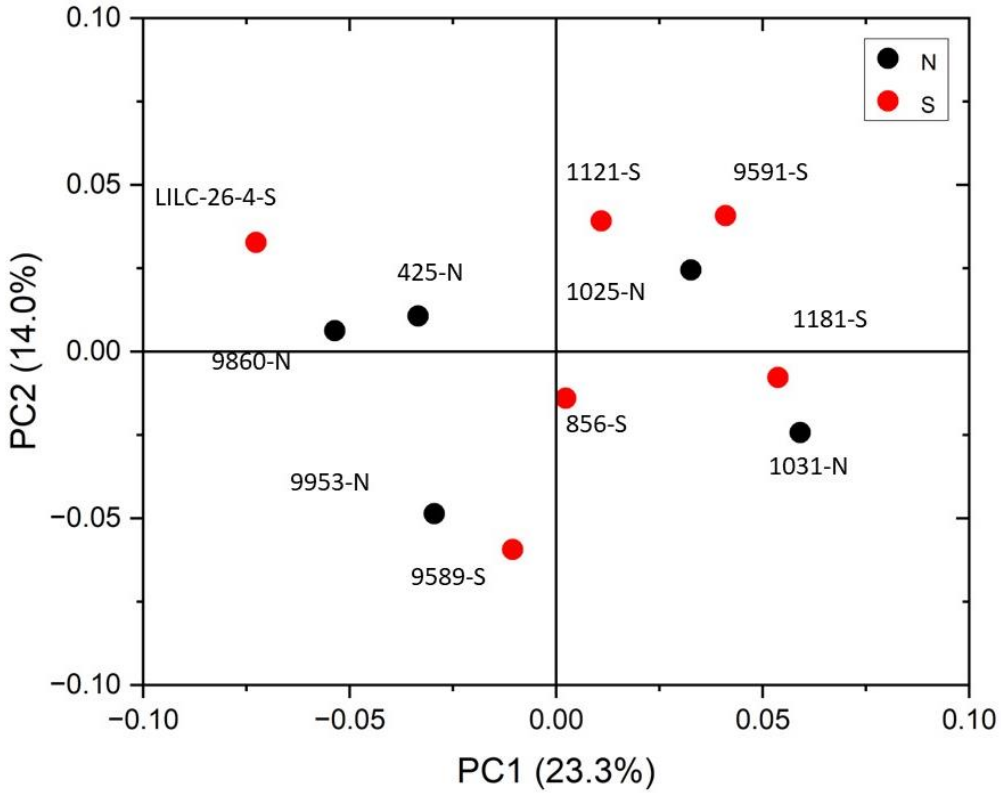


FIGURE 6-4 PRINCIPAL COMPONENT ANALYSIS OF FTIR SPECTRAL DATA

PC1 and PC2 accounted for 23.3% and 14.0% of the variability, respectively. The remaining PCs accounted for less than 12% (each) of the variability. No obvious spatial clustering patterns were observed. Specifically, samples from the south (S; drought treatment) and north (N; irrigated treatment) sets are not readily distinguished by PCA, indicating that treatment was not a significant contributor to the observed variability. Examining the loadings of each PC provides insight as to which structural features (as determined by wavenumbers) contribute most significantly to the observed variation. PC1 is driven primarily by the band around 1317 cm^{-1}, which is typically associated with the C-O stretching of the S unit ring. This aligns with the (semi-quantitative) observation from the NMR spectra, where the S unit signal is more variable than the G unit signal. Wavenumbers that contributed to PC2 include the peak between 1620-1630 cm^{-1}, which is associated with C=C aromatic skeletal vibrations. The peak around 830 cm^{-1}, associated with C-H bending of syringyl units, also contributed to PC2. Another contributor to PC2 includes wavenumber 1081 cm^{-1}, which is a shoulder of the peak centered around 1035 cm^{-1} peak associated with C-O stretching of primary alcohols in lignin and polysaccharides. The peak around 1685 cm^{-1}, associated with conjugate carbonate of carboxylic acid and ketone groups, was common to both PC1 and PC2. Various peaks in the FTIR spectrum have been associated with lignin structure, such as syringyl and guaiacyl units. For syringyl units the peak around 1328-1317 cm^{-1} is associated with aromatic ring breathing, 1116-1106 cm^{-1} is associated with C-H bond deformation, and 833-827 cm^{-1} is associated with out of plane bending. All samples exhibited a relatively weak signal around the 833-827 cm^{-1} peak, but much stronger signals around the 1328-1318 cm^{-1} and 1116-1106 cm^{-1} peaks. For guaiacyl units, the peaks at 1267 cm^{-1} and 122-1210 cm^{-1} are associated with C-O, C-C, and C=O stretching..The ratio of intensities at the 1327-1317 cm^{-1} peak (syringyl) and 1277 cm^{-1} peak (guaiacyl) are often used to estimate the S/G ratio of biomass [152]. Indeed, the peak around 1317cm^{-1} was well correlated with S unit abundance as measured by NMR. However, the 1317/1277cm^{-1} ratio was not well correlated with the S/G ratio determined by NMR. Instead, the ratio of the 830cm^{-1} and 1230cm^{-1} peak intensity was found to be well correlated with the NMR S/G ratio (CC=0.51, *p*-value=0.1). This correlation appears to be driven by the correlation of the 833cm^{-1} peak with the NMR S/G ratio (CC=0.59). Therefore, the 830cm^{-1} peak may provide a better measurement of condensed S

units that were identified by NMR. In any case, these results provide confidence that the high degree of variability of lignin structure of these leaves observed via NMR is authentic.

6.4.3 Metabolite Profiling

In addition to FTIR and NMR analyses, the 11 foliage samples were subjected to metabolite profiling. Samples were extracted with 80% ethanol, dried under nitrogen, and silylated to generate trimethylsilyl (TMS) derivatives prior to analysis by gas chromatography-mass spectrometry (GC-MS). Sorbitol was used as the internal standard. A representative spectrum is displayed in Figure 6-5. A summary of some of the most abundant metabolites are reported in Table 6-4 on the basis of μg metabolite per g dry weight (DW) in sorbitol equivalents. A full list of quantified metabolites is available in the supplemental information. Sucrose was by far the most abundant metabolite - approximately 3.5 times more abundant than the second most abundant metabolite (tremuloidin). Other highly abundant metabolites include salicortin, salicin, myo-inositol, quinic acid, glucose, and trichocarpin. Although tremuloidin was the second most abundant metabolite on average, two samples (1025-N, 1031-N) observed to be lacking any measurable tremuloidin. Additionally, sample 1031-N was observed to have low levels of salicin, trichocarpin, catechol, and tremulacin, which were highly abundant in most other samples. This is illustrated in the overlaid total ion chromatograms (TICs) of sample 1031-N and LILC-26-4-S displayed in Figure 6. From this figure it can be observed that sample LILC-26-4-S had prominent signals at a retention time (RT) of 16.887 and 17.011 corresponding to trichocarpin and salicortin, respectively. However, these peaks were barely discernable in the 1031-N TIC. A similar phenomenon was observed with the signal at RT=16.224 associated with tremuloidin. A summary of select metabolites from each sample is offered in Table 6-3.

As the samples from the north side of the field (denoted with “-N” in the sample name) were irrigated and samples from the south side of the field (denoted with “-S” in the sample name) were subjugated to drought conditions, one may expect to observed established differences in metabolites associated with drought stress. High variability was observed in levels of tremulacin and salicortin among the samples.

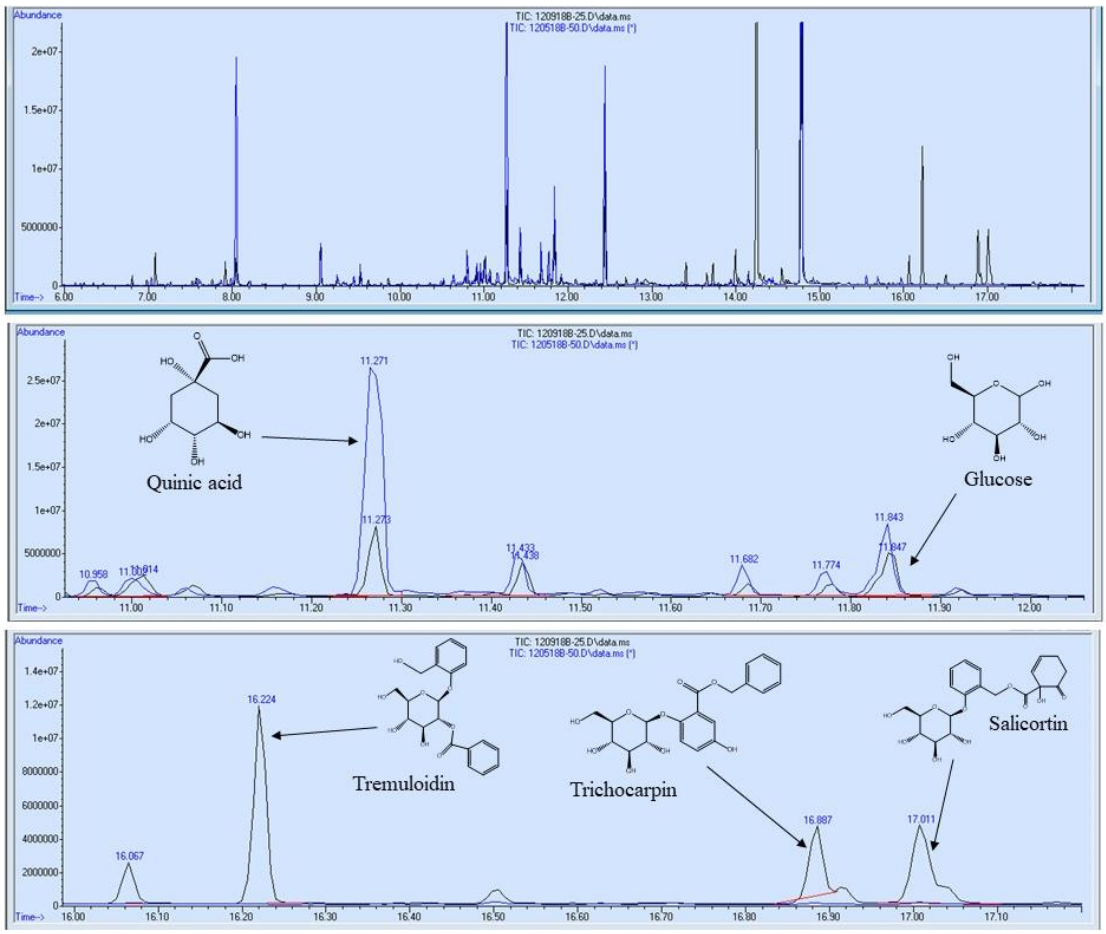


FIGURE 6-5 TIC SPECTRA OF TWO SAMPLES WITH SELECT METABOLITES HIGHLIGHTED

On average tremulacin levels were higher in the irrigated samples and salicortin levels were higher in the drought samples. These phenolic glycosides have been linked to roles in herbivore and/or pathogen defense [194], suggesting there may be biotic stressors impacting these samples. PCA was also utilized to distinguish samples based on the variability of metabolite profiles, with results plotted in Figure 6-6. To avoid biasing the results based on the differences in magnitudes of the metabolite concentrations, data were standardized by the standard deviation. PC1 and PC2 accounted for 42.0% and 19.8% of the variability, respectively. Other PCs accounted for less than 14% (each) of the variability. Like the FTIR PCA, 6-7 samples are clustered closer together in (or just outside) quadrant I, whereas 4-5 samples are more dispersed. Loadings for PC2 indicate that it is highly driven by shikimic acid, fructose, glucose, galactose, and raffinose. Loadings for PC1 are less differentiable, but are highly influenced by salicylic acid, catechol, caffeic acid, stearic acid, and maleic acid. PCA did not distinguish samples by treatment, again indicating that this was not a major contributor to metabolite variability.

The metabolite profiles were also correlated with the lignin traits elucidated by NMR to explore potential relationships between metabolites and lignin structure in leaves. The resulting Pearson correlation coefficients are tabulated in Table 6-4. Several metabolites were found to be correlated with various lignin traits. Among these correlated metabolites were the salicyloids, salicin and salicylic acid. Specifically, both metabolites were negatively correlated with the S/G ratio. Additional salicin derivatives such as benzyl-salicylic acid-2-O-glucoside, salicyl-coumaroyl-glucoside, and salicyl alcohol also exhibited strong correlations to lignin S/G ratio. In *Populus* sp., salicin biosynthesis has a benzoic acid route [195], which can be derived through the shikimate/chorismite pathway [196] or the phenylalanine/cinnamate route [197]. Current research indicates that the production pathways of salicyloids and lignin are not competitive processes, though tradeoffs between the two processes have been hypothesized [198]. The correlation between salicyloids and the lignin S/G ratio would provide weight to this hypothesis, as increased salicyloid content is correlated with increased ratio of guaiacyl units. Although no *p*-hydroxybenzoate (PB) units were observed in the lignin of these leaf tissues, stem tissue of *Populus* usually contains ~5% PB units.

TABLE 6-3 SUMMARY OF METABOLITE ABUNDANCE AS DETERMINED BY GC-MS

Abundance (µg/g DW in sorbitol equivalents)											
Metabolite	**1121-S**	**1181-S**	**9589-S**	**9860-N**	**9953-N**	**1031-N**	**9591-S**	**1025-N**	**LILC-26-4-S**	**856-S**	**425-N**
Sucrose	189837	260895	271979	222162	457405	146732	145955	426783	378962	224151	407349
Tremuloidin	70907	57741	95449	81725	209857	0	21878	0	139440	39062	176411
Salicin	37348	51195	63689	45474	97508	103	24818	25613	101144	44351	94061
Myo-inositol	19369	37636	30932	23191	28732	18209	20249	40563	33033	33720	41203
Quinic acid	59883	57465	62693	29131	28392	55523	31092	143311	24148	29405	8004
Glucose	41119	18283	20227	8872	21190	13444	28488	129027	22380	8564	7702
Trichocarpin	2921	14313	19400	8845	20102	64	3822	2127	15094	6280	32742
Malic acid	2976	1542	4371	5919	18463	3062	2409	11002	7676	767	5311
Shikimic acid	4641	5524	7534	5404	24291	6718	8727	31191	16158	5627	4326
Citric acid	2755	391	5849	3718	5051	1248	1349	4973	4779	626	4186
2-phenethyl-glucoside	2608	4131	6153	3666	12858	4114	1798	7280	13685	7653	9432
Catechol	1932	1565	1997	1816	5674	18	472	1499	4326	2057	7035
Fructose	5528	3976	4749	2477	4357	4766	7720	25158	6884	2640	1814
1,2-cyclohexanediol glucoside	2993	1768	1268	1763	6040	222	634	1965	6951	5473	8631
5-oxo-proline	675	1797	1427	1111	3301	982	975	2276	2996	895	2070
Catechin	1028	2323	1369	907	883	1013	1619	5670	433	3135	1702

TABLE 6-3 CONTINUED

Abundance (µg/g DW in sorbitol equivalents)											
1181-S	**9589-S**	**9860-N**	**9953-N**	**1031-N**	**9591-S**	**1025-N**	**LILC-26-4-S**	**856-S**	**425-N**	1657	2222
Digalactosylglycerol	1855	3009	1509	1413	2828	524	3743	2656	7403	2496	2164
Tremulacin	870	3474	3377	800	880	0	310	0	1412	12667	30895
Threonic acid	632	717	1060	1159	3383	1360	874	3809	1885	550	1050
Galactose	2253	952	690	550	2277	1597	2430	16068	1239	560	526
Salicortin	215	4188	3441	6307	5513	164	1497	622	12591	50949	31845
Monogalactosylglycerol	1788	2005	1124	725	3481	188	2786	1045	8974	1545	1888
Phosphate	881	1011	1492	311	2912	1050	1557	3434	1434	546	475
Glyceric acid	810	763	797	494	3092	920	508	3009	1391	583	369

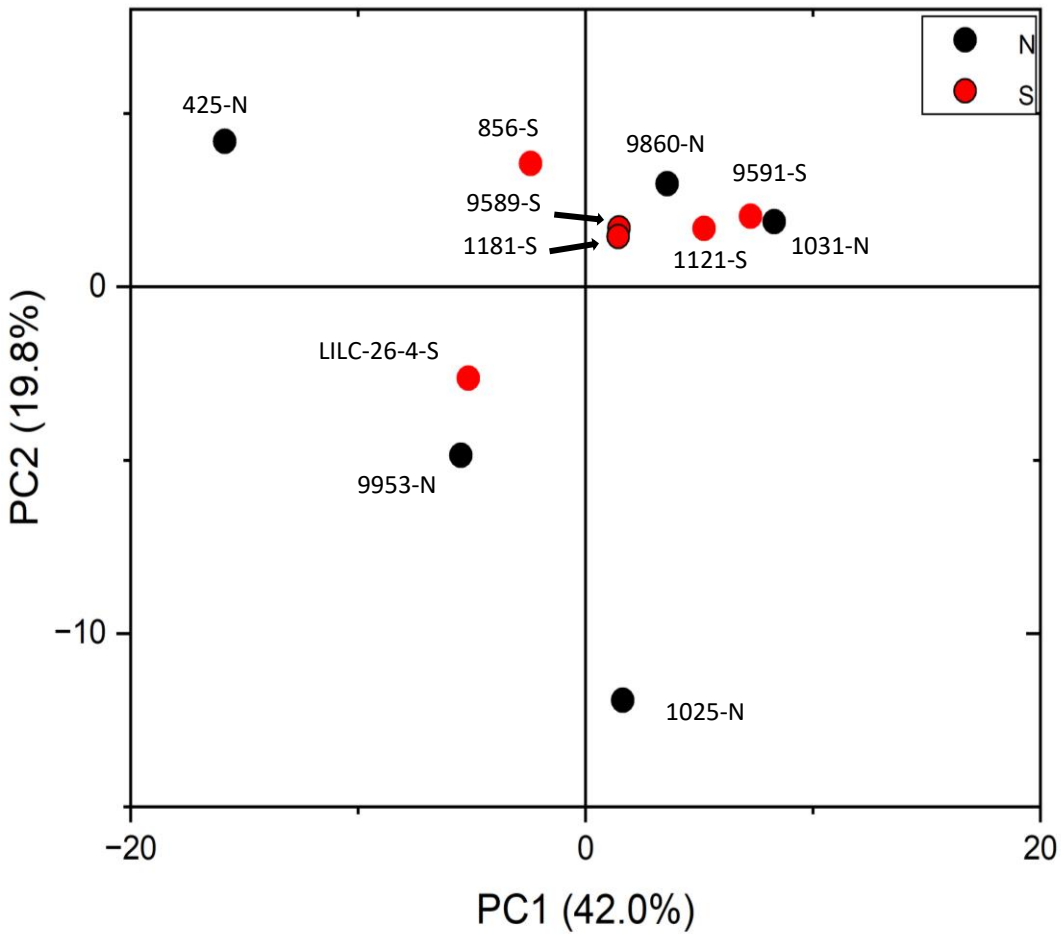


FIGURE 6-6 PRINCIPAL COMPONENT ANALYSIS OF METABOLITE ABUNDANCE DATA

A future study examining salicyloids and benzoate incorporation of lignin may shed additional light on potential trade-offs. Another metabolite found to be correlated with lignin structure was quinic acid. Quinic acid can be a precursor of monolignol synthesis [199]. However, its incorporation is associated with the phenylpropanoid pathway product p-coumaroyl CoA, which represents a major branchpoint in the pathway and can be directed to various pathways to produce flavonoids, monolignols, or a number of other compounds [200]. The flavonoid quercetin, which was also observed to be associated with lignin S/G ratio, is derived by converting p-coumaryl CoA to chalcone, though malonyl CoA is also a substrate [201]. A similar relationship has been reported previously, as N-acetylserotonin methyltransferase (*MsASMT1*) was shown to impact both lignin S/G ratio and glycosides of quercetin when overexpressed in alfalfa [202]. Multiple enzymes catalyze reactions of p-coumaroyl CoA. hydroxycinnamoyl CoA:shikimate hydroxycinnamoyl transferase (HCT) has higher specificity toward shikimic acid and is associated with lignin biosynthesis [reviewed in [200]]. Hydroxycinnamoyl CoA:quinate hydroxycinnamoyl transferase (HQT) utilizes quinic acid more efficiently, and is more closely associated with the production of chlorogenic acids (reviewed in [200]).

6.4.4 Conclusion

In summary, these results provide an extensive characterization of lignin structure and metabolite abundance in *Populus* foliage tissue, of which there is currently limited knowledge. A surprisingly high degree of variability was observed in the cell wall structure, especially lignin, through HSQC NMR and FTIR. Specifically, condensed syringyl structures were observed in most samples, and these levels appear to be independent of treatment. A high degree of variability of metabolite abundance was also observed by GC-MS analysis, especially in tremuloidin, trichocarpin, and salicortin. These results demonstrate differences between foliage and more well-studied stem tissue and highlight previously unexplored nuance and variability associated with foliage tissue of poplar.

TABLE 6-4 PEARSON CORRELATION COEFFICIENTS OF SELECT METABOLITES AND LIGNIN STRUCTURAL TRAITS

Metabolite	S	G	S/G	β-O-4	β-5	β-β
Sucrose	-0.35	0.35	-0.45	0.15	0.33	0.26
Tremuloidin	-0.59	0.59	-0.48	0.48	0.53	0.55
Salicin	-0.69	0.69	-0.64	0.55	0.57	0.56
Myo-inositol	-0.23	0.23	-0.49	0.07	0.03	0.17
Quinic acid	0.65	-0.65	0.38	-0.77	-0.55	-0.70
Glucose	0.39	-0.39	0.17	-0.51	-0.25	-0.41
Trichocarpin	-0.54	0.54	-0.56	0.41	0.35	0.54
α-linolenic acid	-0.57	0.57	-0.41	0.64	0.60	0.35
Quercetin	-0.55	0.55	-0.59	0.46	0.51	0.41
Salicylic acid	-0.62	0.62	-0.54	0.51	0.39	0.66
Salicyl-salicylic acid-2-O-glucoside	-0.64	0.64	-0.49	0.65	0.62	0.39
Salicyl-coumaroyl-glucoside	-0.66	0.66	-0.55	0.57	0.42	0.52
Caffeic acid	-0.55	0.55	-0.63	0.40	0.38	0.55
Gallocatechin	0.72	-0.72	0.54	-0.74	-0.62	-0.70

Chapter 7:

Novel candidate genes for lignin structure identified through genome-wide association study of naturally varying *Populus trichocarpa*

Sections of this chapter was originally published as:
Nathan Bryant, Jin Zhang, Kai Feng, Mengjun Shu, Raphael Ployet, Jin-Gui Chen, Wellington Muchero, Chang Geun Yoo, Timothy J. Tschaplinski, Yunqiao Pu, & Arthur J. Ragauskas. (2023). Novel candidate genes for lignin structure identified through genome-wide association study of naturally varying Populus trichocarpa. Frontiers in Plant Science, 14. https://doi.org/10.3389/fpls.2023.1153113
Arthur J. Ragauskas, Wellington Muchero, and Jin-Gui Chen conceived the study. Wellingon Muchero and Jin-Gui Chen collected the material. Nathan Bryant, Chang Geun Yoo, and Yunqiao Pu performed the experiments. Jin Zhang, Kai Feng, Raphael Ployet, and Mengjun Shu performed the GWAS analysis. Nathan Bryant, Mengjun Shu, and Raphael Ployet analyzed the data. Nathan Bryant, Jin Zhang, and Raphael Ployet drafted the manuscript. All authors reviewed, revised, and approved the manuscript. *Frontiers* publishes under a creative common license, meaning authors retain copyright and material can be reprinted with citation of the original publication.

7.1 Abstract

Understanding compositional variability of biomass, and especially lignin, is crucial to the development and implementation of financially viable, integrated biorefineries. In this study, a population of 409 three-year-old, naturally varying *Populus trichocarpa* genotypes were characterized by heteronuclear single quantum coherence (HSQC) nuclear magnetic resonance (NMR) to elucidate twelve structural properties of lignin. This phenotypic analysis revealed *p*-hydroxybenzoate units were not normally distributed and exhibited an overall negative trend with the syringyl:guaiacyl ratio. A subsequent genome-wide association study (GWAS) was conducted using approximately 8.3 million single nucleotide polymorphisms (SNPs) and a 0.05 minor allele frequency. The GWAS identified 756 genes that were significantly associated ($-\log_{10}$(*p*-value)>6) with at least one lignin phenotype. A network analysis was utilized to integrate additional lines of evidence for the involvement of identified genes in lignin biosynthesis. Several promising candidate genes were identified, many of which have not previously been

reported to be associated with lignin or cell wall biosynthesis. These results provide a resource for gaining insights into the molecular mechanisms of lignin biosynthesis and new targets for future genetic improvement in poplar.

7.2 Introduction

Poplar (*Populus* sp.) is a promising lignocellulosic biomass feedstock due to its fast growth, ability to grow on marginal land, high cellulose content, and relatively low lignin content [127, 203]. However, lignocellulosic biomass exhibits tremendous variability in cell wall traits, such as composition and structure. Cell wall structure and composition depend on factors such as environmental conditions (i.e., drought, cold stresses), biomass type (i.e., woody vs. herbaceous), tissue (i.e., stem vs. leaf), and genetic variations. One cell wall component that demonstrates a high degree of variability is lignin. Lignin is a complex and heterogeneous biopolymer that accounts for 18-32% (dry weight) of *Populus* [127, 204]. It serves several important biological functions, including water transport, providing mechanical strength, and response to environmental stresses [205]. However, since it acts as a natural barrier against pathogens, it is one of the main factors contributing to biomass recalcitrance to biological conversion to biofuels. Lignin is typically comprised of three primary monomers synthesized through the general phenylpropanoid and monolignol specific pathways: sinapyl alcohol, coniferyl alcohol, and *p*-coumaryl alcohol. When exported to the apoplast, these monolignols are oxidized by laccases and/or peroxidases, and then undergo radical cross-coupling reactions. Once coupled into the lignin polymer, these alcohols are identified as syringyl (S), guaiacyl (G), and *p*-hydroxyphenyl (H) units, respectively [206]. It has also been shown that lignin demonstrates plasticity by incorporating non-canonical monolignols into the polymer, such as *p*-hydroxybenzoates (PB) [207]. These units are incorporated into the lignin polymer *via* a variety of interunit bonds, such as aryl ether (β-*O*-4), resinol (β-β), and phenylcoumaran (β-5), among others. The structure of lignin has potential bioenergy implications, as evidenced by the lignin S/G ratio association with biomass digestibility and conversion to biofuels [71]. Lignin structure can also influence lignin valorization. For instance, acetaminophen, the active ingredient in Tylenol, can be synthesized from the PB moiety [208].

The formation of the secondary cell wall requires coordination of many metabolic pathways [93], presenting challenges in linking phenotypes to genetic mutations. Additionally, complex traits such as lignin are often controlled by several multigenic families. Genome-wide association studies (GWAS) are powerful tools for identifying polymorphic loci that contribute to phenotypic variation and sometimes trace-back to the genes or biological mechanisms involved. Due to the large sample size required for GWAS, high throughput techniques are typically utilized to analyze lignin traits. However, high throughput methods such as pyrolysis molecular beam mass spectroscopy (Py-MBMS) or near-infrared (NIR) spectroscopy provide limited information on lignin composition. These methods are generally used to estimate only the relative lignin to sugars ratio within the cell walls, or the S/G ratio for lignin itself. Genomic association mapping has been successfully employed on *Populus* for bioenergy traits including lignin content and S/G ratio [209, 210], biomass yield [211], and cell wall sugars [212]. Consequently, the genetic basis of most lignin traits remains understudied by GWAS, and new methods for characterization of lignin phenotypes are urgently required. In comparison to Py-MBMS or NIR, the analytical technique heteronuclear single quantum coherence (HSQC) NMR, is more time and labor intensive, but has the potential to provide substantially more information on lignin structure.

In this study, we performed a deep phenotyping of the lignin polymer of 409 *P. trichocarpa* genotypes by HSQC NMR. By performing a detailed phenotyping of twelve lignin traits in a large population of poplar trees, we found that the lignin composition is highly variable across individuals, with evidence of incorporation of non-canonical monolignols into the polymer of lignin. Subsequently, a GWAS analysis enabled the identification of novel candidate genes that could explain the diversity in lignin composition. Most of the candidate genes identified were not previously reported to be associated with lignin or cell wall biosynthesis. This provides a resource for gaining insights into the molecular mechanisms of lignin biosynthesis and new targets for future genetic improvement in poplar.

7.3 Materials and methods

7.3.1 Biomass preparation

Wood samples for this study were collected from three-year-old *Populus trichocarpa* grown in a common garden in Corvallis, OR (44°34′14.81″N 123°16′33.59″W). Site establishment and management practices were previously described by Muchero et al. [213]. One-centimeter diameter increment cores were collected at breast height for the 409 genotypes in January 2013. Cores were stored in zip-lock bags at -20^{0}C before processing. Wood cores were air dried at room temperature before they were ground using Wiley Mini-Mills (Swedesboro, NJ) with a 20-mesh screen. Ground samples were stored in glass vials at room temperature.

7.3.2 Lignin sample preparation

Each sample was Soxhlet extracted using toluene/ethanol (2:1, *v:v*) for at least eight hours and subsequently air-dried for at 24 hours. Approximately 500 mg of each extractives-free sample was ball-milled for two hours at 600 RPM (at five-minute intervals) on a Retsch planetary ball mill. The ball-milled biomass was then subjected to enzymatic hydrolysis with cellulase (Sigma-Aldrich) in a sodium acetate buffer at 37^{o}C and 200 RPM for 48 hours. The solid lignin enriched residues were then separated *via* centrifugation and lyophilized for 48 hours for NMR analysis.

7.3.3 NMR analysis

Lignin structure was analyzed by 2D HSQC NMR with a Bruker Avance III 500-MHz spectrometer. Approximately 40 mg of lignin enriched residue was dissolved in 0.5 mL of DMSO-d_6 in a 5 mm NMR tube and sonicated for one hour. The Bruker pulse sequence hsqcetgpsip2.2 was utilized on a N_2 cryoprobe with the following parameters: spectra width of 12 ppm in the ^{1}H dimension with 1024 data points; spectra width of 220 ppm in the ^{13}C dimension with 256 increments and 32 scans. The HSQC spectra were analyzed with Bruker TopSpin 3.5pl6 software. The DMSO-d_6 solvent peak (δ_C/δ_H at 39.5/2.49) was used to calibrate the chemical shifts. At least annually, the repeatability and experimental error of the HSQC measurement are quantified by analyzing a standard *Populus* sample. Lignin was isolated from the standard *Populus* sample per the method described above, and three separate samples were analyzed by the same pulse sequence. For this most recent analysis, the standard deviation of the

three samples ranged from 0.1 (for H unit) to 1.4 (β-*O*-4 linkage). The coefficient of variation (CV) may be considered a better measurement of variability, since it is defined as the ratio of the standard deviation to the mean expressed as a percentage. The CVs of the standard *Populus* samples ranged from 1.4% (S unit) to 18.9% (β-5 linkage). A study utilizing ^{1}H NMR for biomarker analysis identified that larger peaks exhibited a CV of 5-10%, whereas smaller peaks had CV in the 15-30% range [214], which is consistent with our observed measurements.

7.3.4 Genome-wide association study

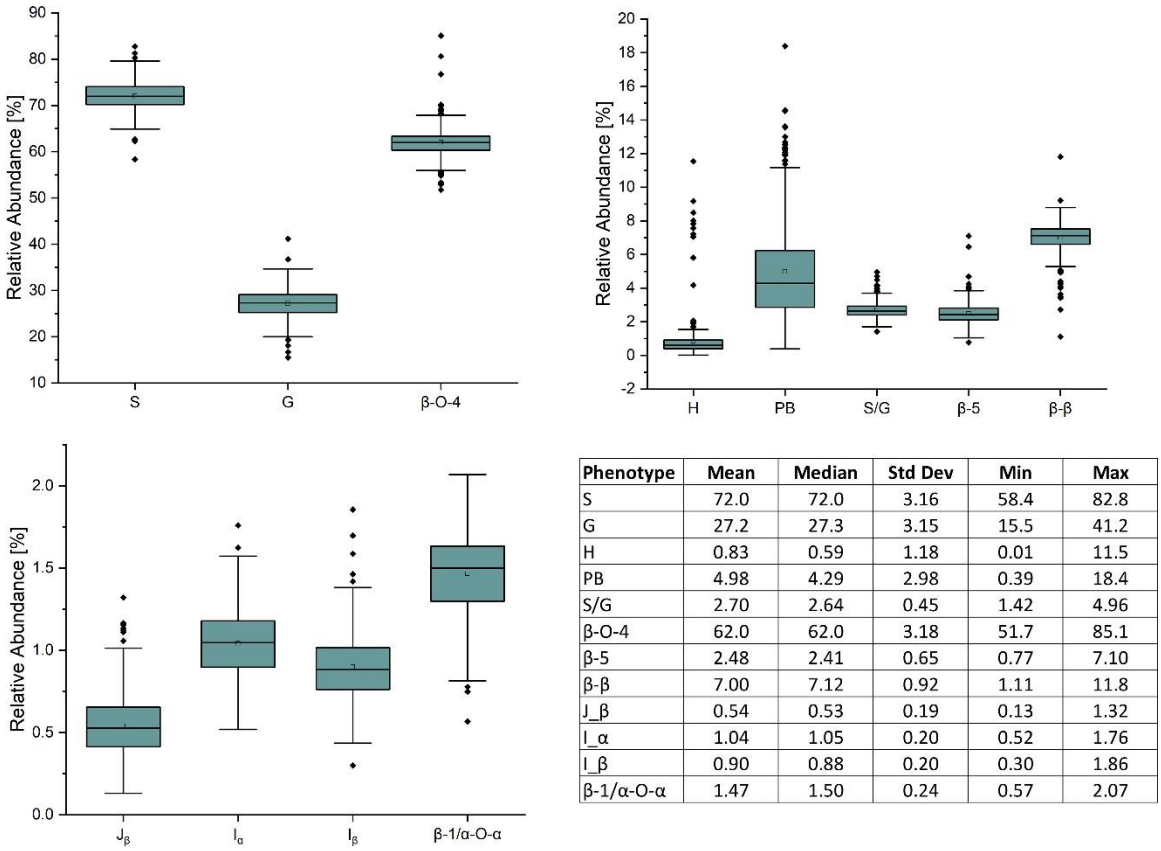
Whole genome resequencing, single nucleotide polymorphism (SNPs)/nucleotide insertions and deletions (indels) calling and SnpEff analysis for the 917 individuals of this *Populus* GWAS population was previously described by Zhang et al. [99]. The *P. trichocarpa* Nisqually-1 reference genome v3.1 was used for read alignment and variant calling. The resulting SNP and indel dataset is available at http://bioenergycenter.org/besc/gwas/. This study utilized genotypic data for a subset of 409 genotypes from this dataset. To assess genetic control, we used the GEMMA software to calculate kinship for the *Populus* GWAS population as the correction factor for genetic background effects [99]. Genotype-to-phenotype associations were performed using 8,301,860 SNP and indel variants with minor allele frequencies > 0.05. The HSQC spectra from 2D HSQC NMR were used as phenotypes. Statistical significance of associations was evaluated using the Storey's Q-value threshold. Deviation of p-values from normality was assessed using quantile-quantile (Q-Q) plots. Candidate gene identification and RNAseq mapping for co-expression analysis were performed using the *P. trichocarpa* v3.1 reference genome. RNAseq of xylem tissue of 378 *Populus trichocarpa* transgenics plants knockdown or overexpressing monolignol genes and transcription factors involved in the regulation of cell wall biosynthesis were downloaded from the Sequence Read Achieve (SRA; accession: PRJNA314500) [100]. Library quality was assessed using FastQC (v0.11.9; https://www.bioinformatics.babraham.ac.uk/projects/fastqc/), residual adapters and low-quality reads were trimmed using Trimmomatic v0.39 [101] reads were mapped to the reference genome using STAR v 2.7.6a (default parameters and --outFilterMultimapNmax 100 [102] and transcript per million (TPM) values were extracted for all annotated genes using Stringtie [103]. 18

samples with low mapping rates (<80% of mapped reads) were excluded for the subsequent analysis. Co-expression of candidate genes with 86 phenylpropanoids and lignin-related genes was estimated by calculating pairwise Pearson correlation coefficient (PCC) across 360 samples using the function rcorr() from the Hmisc R package [104, 105]. For each potential candidate gene, multiple individual scores were calculated: (a) significance threshold of $-\log_{10}(p\text{-value})=6$, 7, and 8 were assigned an individual score of 1, 2, and 3, respectively; (b) connectivity with SNPs was scored according to $\log_{10}$(number of connected SNPs); (c) connectivity of the SNPs with phenotypes; number and average value of significant co-expression associations ($|PCC| \geq 0.5$, FDR < 0.001) with lignin-related genes. All individual scores were scaled to obtain values ranging from 0 to 1. These individual scores were summed to obtain a final overall score, which was utilized to prioritize candidate genes for consideration (higher scoring genes were considered best candidates).

7.4 Results and discussion

7.4.1 Lignin chemistry

409 unique poplar genotypes were analyzed by HSQC NMR to elucidate structural information for twelve lignin phenotypes, including S units, G units, H unit, PB units, S/G ratio, cinnamyl alcohol end groups (I_α, I_β), cinnamyl aldehyde end groups (J_β), β-O-4 aryl ether linkages, β-5 phenylcoumaran linkages, β-β resinol linkages, and β-1/α-O-α spirodienone linkages. with results summarized in Figure 7-1. As expected, the lignin of the *P. trichocarpa* population is comprised primarily of S and G units, with an average of approximately 72 and 27 units per 100 S+G+H units, respectively. The quantity of S units was measured to be as low as 58.4% and as high as 82.8%. Similarly, G units ranged from 15.5% to 41.2%. This resulted in a population average S/G of 2.70, though values ranged from 1.42 to 4.96. The next most abundant phenotype was PB. While PB content averaged 4.87%, levels across the population varied significantly ranging from near background (0.39%) to a non-trivial 18.4% - approximately the lower limit of G unit content. However, within ±0.1 of the average S/G ratio (i.e., 2.60-2.80), PB content ranges from near background levels (1.33%) to 14.6%. To further explore this relationship, the samples were subdivided into a low S/G ratio fraction (<2.70) and a high S/G fraction (>2.70), as



Phenotype	Mean	Median	Std Dev	Min	Max
S	72.0	72.0	3.16	58.4	82.8
G	27.2	27.3	3.15	15.5	41.2
H	0.83	0.59	1.18	0.01	11.5
PB	4.98	4.29	2.98	0.39	18.4
S/G	2.70	2.64	0.45	1.42	4.96
β-O-4	62.0	62.0	3.18	51.7	85.1
β-5	2.48	2.41	0.65	0.77	7.10
β-β	7.00	7.12	0.92	1.11	11.8
J_β	0.54	0.53	0.19	0.13	1.32
I_α	1.04	1.05	0.20	0.52	1.76
I_β	0.90	0.88	0.20	0.30	1.86
β-1/α-O-α	1.47	1.50	0.24	0.57	2.07

FIGURE 7-1 POPULATION DISTRIBUTION DATA OF THE TWELVE LIGNIN FEATURED DETERMINED BY HSQC NMR

shown in Figure 7-2 (right). In the low S/G ratio sample fraction, the correlation between PB and S/G ratio remains statistically significant, with an average PB content of 5.5%. However, in the high S/G ratio sample fraction, this correlation does not hold, and the average PB content is slightly lower at 4.0%. Additionally, 21 of 24 high PB outliers (i.e., PB >11.3%) appeared in the lower range of S/G ratios, with no high PB outlier coming from a sample with an S/G ratio above 2.77. It was confirmed that the H unit constituted only a minor fraction of the lignin polymer, averaging just 0.91%. However, like PB, H unit levels were highly variable, ranging from nondetectable to as high as 11.5%. The β-O-4 aryl ether linkage was by far the most abundant interunit linkage, averaging approximately 62 per 100 aromatic units. It was also the most variable linkage, measuring as high as 85.1% and as low as 51.7%. The β-β resinol linkages were shown to make up a minority of interunit linkages, averaging 7 per 100 aromatic units, with an upper limit of 11.8% and a lower limit of 1.11%. On the other hand, β-5 phenylcoumaran, which averaged 2.48%, ranged from as low as 0.77% to as high as 7.10%. The spirodienone (β-1/α-O-α) linkage was present in small but detectable quantities, averaging 1.47%, but not exceeding 2.07%. The population statistics for these phenotypes are summarized in the table in Figure 7-1. Altogether, the large variations observed in both the subunit content and the type linkages, suggest that a number of polymorphisms segregating within the population can drastically affect the structure of the polymer of lignin.

7.4.2 Genome-wide association study (GWAS) of phenotypic variation

To identify the genomic loci controlling lignin phenotypes described above, we performed a GWAS using 409 unrelated *P. trichocarpa* accessions that had genotypic information represented in a panel of > 8.3 million single nucleotide polymorphisms (SNPs) and nucleotide insertions and deletions (indels) as described in the materials and methods. Associations with the phenotypes β-O-4, β-β, β-5, β-1/α-O-α, S, G, H, I, PB, and S/G ratio were tested in this analysis. The GWAS was conducted at increasing significance thresholds (i.e, $-\log_{10}$(p-value)=6, 7, & 8) to differentiate the strength of the associations. At the threshold of $-\log_{10}$(*p*-value)>6, the GWAS identified 756 genes that were significantly associated with at least one phenotype.

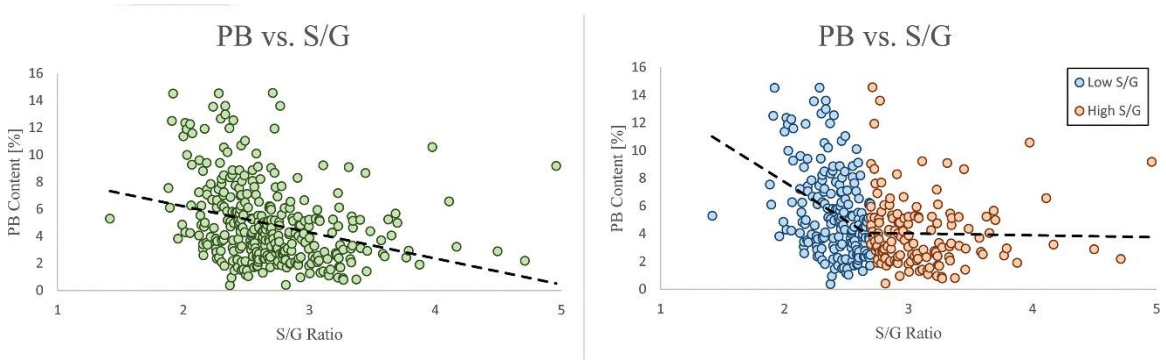


FIGURE 7-2 PB ABUNDANCE IS NEGATIVELY CORRELATED WITH THE S/G RATIO

A significant enrichment in cell wall related GO terms was detected for the highest significance thresholds ($-\log_{10}(p\text{-value})=7$ and $-\log_{10}(p\text{-value})=8$). The enrichment in cell wall related genes (relative to total number of genes) identified at each of these levels were 1.52, 1.87, and 2.48, respectively, when considering Poplar gene annotations. When considering Arabidopsis gene annotations, a similar trend was observed with the enrichment increasing from 2.70 to 5.79 between the $-\log_{10}(p\text{-value}) = 6$ and $-\log_{10}(p\text{-value}) = 8$ thresholds. The increase in enrichment for specific GO terms for higher GWAS significance thresholds provides evidence that the approach is successful at capturing the most likely causal genes for phenotyping variations in lignin composition, as opposed to a simple random sampling of genes from the genome. At the $-\log_{10}(p\text{-value})=6$ threshold, 32 of the 756 genes identified by GWAS have previously been associated with cell wall biosynthesis in *Populus* or *Arabidopsis*. A network analysis is offered in Figure 7-3 to visualize the relationships between select GWAS identified genes and their known relationship to phenylpropanoid biosynthesis genes. The GWAS identified several genes (Potri.006G169600, Potri.001G045100 and Potri.003G059200), which are predicted to encode key enzymes of the phenylpropanoids pathway: 4-coumarate:CoA ligase 2 (4CL2), cinnamoyl CoA reductase 1 (CCR1), and caffeoyl shikimate esterase (CSE), respectively. In addition, possible homologs of AtTRA2 (Potri.003G161900) and AtMAT4 (Potri.008G099300) were found associated with H-lignin and PB content. In previous studies, perturbations in these two genes induced drastic changes in lignin content and structure in Arabidopsis [215, 216]. In addition to effector genes directly involved in metabolic pathways, the GWAS highlighted higher level regulators as potential candidates, including major transcriptional regulators. PtrMYB074 (Potri.015G082700), strongly associated with H-lignin content in our GWAS, is a master regulator of secondary cell wall formation in poplar, and was shown to directly regulate the biosynthesis of lignin in wood forming tissues [217]. Potri.001G346600 encoding a possible homolog of AtMYB21 was also found strongly associated with H-lignin content.
In Arabidopsis AtMYB21 promotes flavanol biosynthesis through the regulation of FLS1, and was related to stress response and hormonal signaling [218].

Interestingly, numerous candidate genes highlighted by this approach are predicted to be involved in the biosynthesis of cell wall polysaccharides, such as Potri.002G135500 a possible

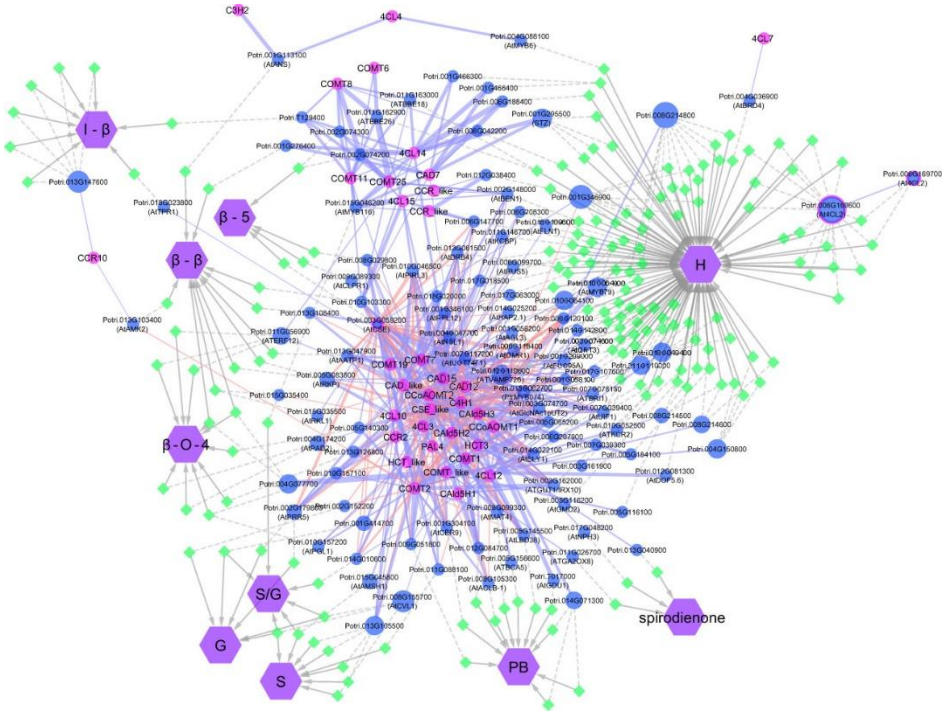


FIGURE 7-3 NETWORK ANALYSIS OF GENOME WIDE ASSOCIATION STUDY RESULTS

homolog of ATRAB6A involved in cellulose biosynthesis [219] , Potri.003G162000 a possible homolog of AtIRX10 involved in xylan biosynthesis [220], or Potri.013G082200 and Potri.003G074600 that are predicted to be a cellulose synthase (AtCSLD3) and a pectin lyase (AtQRT3), respectively.

The GWAS pointed to several other genes which have not been previously associated with lignin or cell wall biosynthesis. To highlight top candidates, all genes were scored based on criteria such as strength of association (*p*-value of nearby SNPs), connectivity with SNPs, and co-expression with phenylpropanoid genes computed from external datasets (Table 7-1). One of the highest scored candidate genes was Potri.008G155500, which is annotated as 20S proteasome beta subunit D1 (PBD1). This candidate gene was highly associated with 16 SNPs across the S, G, and S/G phenotypes. Another highly scored candidate gene was Potri.006G176600, which is annotated as XBAT32. The XBAT32 associations are peculiar in that this gene exhibits strong associations with the S and G phenotypes, but is not strongly associated with the S/G ratio, as presented by the green highlighted Manhattan plots in Figure 7-4. Together with major regulators such as PtrMYB074, and possible homologs of key biosynthetic enzymes such as AtMAT4 and AtTRA2, as mentioned previously, other candidate genes were found strongly co-expressed with lignin biosynthesis genes across transcriptomes of lignin perturbed poplar transgenic plants[221]. Notably, this approach highlighted candidates for the formation of PB. Glutamine dumper 1 (GDU1; Potri.T017000), associated with the PB phenotype, was found to be co-expressed with CAld5H3, a key gene in the lignin biosynthesis pathway that is known to influence S/G ratio. Among the other candidates associated with the PB phenotype, Potri.005G145500 was found negatively co-expressed with PtrCOMT2 and two CSE-encoding genes, involved in lignin biosynthesis in poplar. Potri.005G145500 is a potential ortholog of a group of LBD transcription factors AtLBD37/38/39 that were shown to regulate anthocyanin biosynthesis in Arabidopsis [222]. The candidate genes reported here, as well as other potential candidate genes of interest, are summarized in Table 7-1. All together, these results demonstrate that the GWAS performed highlighted multiple classes of candidate genes for the biosynthesis of the different moieties and linkages that constitute the lignin polymer.

TABLE 7-1 SELECT CANDIDATE GENES IDENTIFIED BY GENOME WIDE ASSOCIATION STUDY

Phenotype	Gene ID	*p*-value	Connectivity with SNPs	Average co-expression with phenylpropanoid genes	Annotation (Arabidopsis alias)
H	Potri.006G169600	8.89E-11	17	0.63	4-coumarate:CoA ligase 2 (4CL2)
H	Potri.001G045100	3.63E-08	2	0.62	cinnamoyl CoA reductase 1 (CCR1)
β-5	Potri.003G059200	8.74E-07	28	0.67	lysophospholipase 2 (CSE)
S	Potri.008G155500	1.16E-07	16	0.58	20S proteasome beta subunit D1 (PBD1)
G		3.69E-09			
S/G		5.46E-08			
S	Potri.006G176600	1.06E-09	27	-	XB3 ortholog 2 in Arabidopsis thaliana (XBAT32)
G		7.88E-09			
S/G		4.52E-07			
PB	Potri.T017000	7.88E-08	1	0.58	glutamine dumper 1 (GDU1)
PB	Potri.T017100	7.88E-08	1	-	glutamine dumper 2 (GDU2)
β-β	Potri.004G077700	7.75E-08	4	0.64	P-loop containing nucleoside triphosphate hydrolases superfamily protein
β-O-4		1.11E-24			

TABLE 7-1 CONTINUED

Phenotype	Gene ID	*p*-value	Connectivity with SNPs	Average co-expression with phenylpropanoid genes	Annotation (Arabidopsis alias)
H	Potri.015G082700	3.32E-08	1	0.66	PtrMYB074 (AtMYB50)
β-β	Potri.019G040900	3.72E-08	4	0.57	MYB domain protein 105
PB	Potri.004G073900	7.97E-07	1	-	Pectin lyase-like superfamily protein
PB	Potri.008G099300	8.78E-07	1	0.66	S-adenosylmethionine synthetase family protein (MAT4)
H	Potri.010G064000	1.51E-07	4	0.59	MYB domain protein 79
β-O-4	Potri.004G174200	5.18E-08	2	0.61	proteasome alpha subunit D2 (PAD2)
β-O-4	Potri.008G011100	1.99E-07	1	-	Plant invertase/pectin methylesterase inhibitor superfamily
PB	Potri.014G142000	3.75E-07	1	0.51	galacturonosyltransferase 15 (GAUT15)
I_β	Potri.005G163900	1.91E-07	9	-	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein (OSU1)

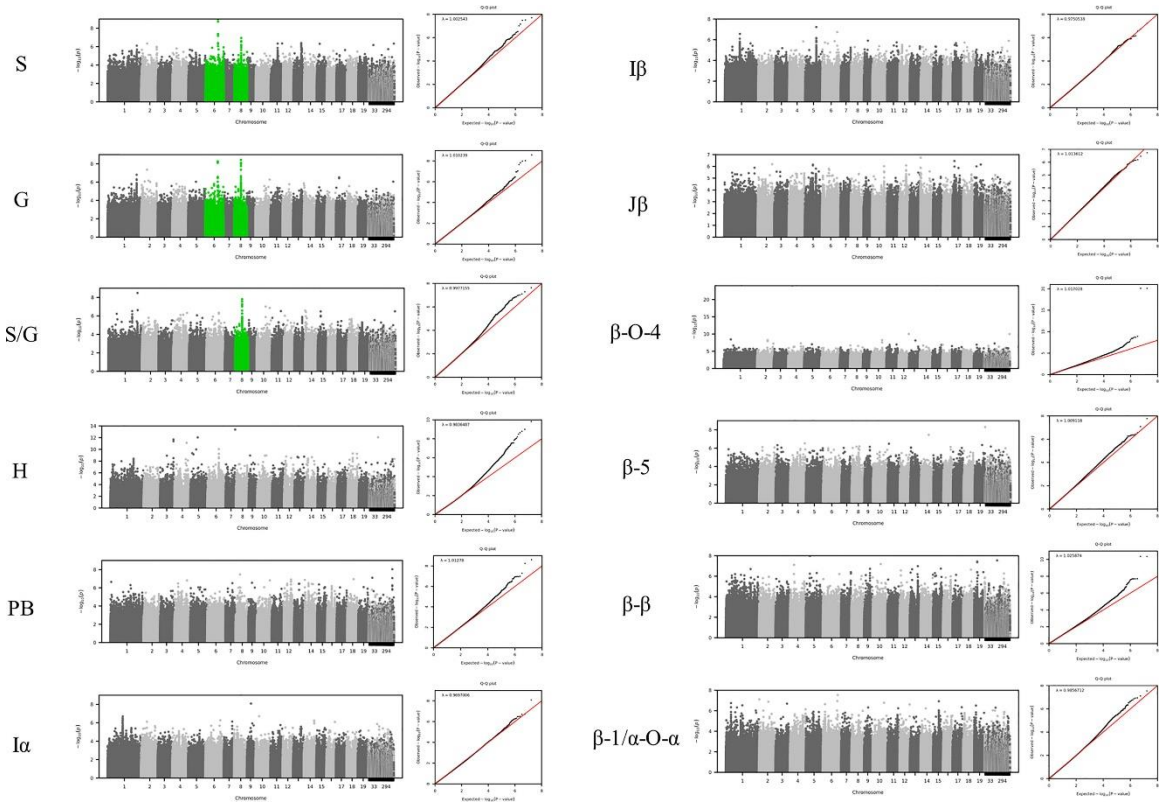


FIGURE 7-4 MANHATTAN AND QQ PLOTS GENERATED FROM GENOME WIDE ASSOCIATIONS

The abundance of PB could be of special interest to biorefinery considerations, either by utilizing PB derivatives as value-added products [208] or as a method of improving biomass deconstruction by increasing the number of ester-linked phenolics [223]. For the 409 genotypes considered in this study, PB is negatively correlated with S/G ratio (Figure 7-2, left), which is consistent with previous observations [224, 225]. More recently, a similar negative correlation was observed in a collection of 316 *P. trichocarpa* analyzed by saponification and HPLC [226]. It has previously been hypothesized that PB (and other acylated units) are produced as a method of promoting syringyl-rich lignin. However, a negative correlation between PB and S units suggests this is not the case. Indeed, PB is a non-canonical monolignol that remains poorly understood. Therefore, GWAS associations related to PB may be considered especially interesting and shed new light underlying the genetic mechanisms underlying PB biosynthesis. As previously mentioned, one of the highest scored candidate genes associated with PB is GDU1. This is due, in part, to its co-expression to CAld5H3, which is known to catalyze a key reaction step in the lignin biosynthesis pathway and therefore heavily influence the S/G ratio. Because there is a clear correlation between PB and S/G, one may attribute this to being an artifact of CAld5H3 expression, rather than GDU1. However, manipulation of CAld5H3 in rice did not impact p-coumarate (the analogous PB ester in grasses) [227]. It has previously been shown that GDU1 is localized at the plasma membrane and is involved with nonselective amino acid export from plant cells into the apoplast [228]. Export of amino acids (including phenylalanine) increased when GDU genes were overexpressed in Arabidopsis [229]. However, the transport mechanism(s) by which lignin monomers are transferred from the cytosol to the apoplast remain unresolved [230]. Additionally, the association of glutamine synthesis with ammonia removal following phenylalanine and tyrosine conversion to *trans*-cinnamic acid and *p*-coumaric acid, respectively, early in the lignin pathway, has recently been reported in *Brachypodium distachyon* (Barros et al. 2021). Given such, GDU1 may play a role in preventing excessive accumulation of glutamine associated with the synthesis of lignin precursors. As another point of emphasis, glutamine dumper 2 (Potri.T017100) was also associated with the PB phenotype. Based on these factors, GDU1 could be further evaluated for potential impacts of PB incorporation into the lignin polymer.

The S/G ratio is perhaps the most well-studied lignin trait. Yet, the GWAS analyses identified strong associations with several genes which have not been previously linked to lignin biosynthesis. This is not entirely surprising, as the formation of the secondary cell wall requires coordination of many metabolic pathways [93], and as previously mentioned, one highly associated candidate gene is the 20S proteasome gene PBD1. This candidate gene, located on chromosome 8, is strongly associated with the S, G, and S/G ratio. The 20S proteasome has been shown to degrade client proteins to amino-acid residues [231]. It is also the proteolytic core of the 26S proteasome, which mediates proteolysis and plays a key role in the regulation of critical cellular processes, such as transcriptional control, cell cycle progression, and stress response. Staszczak et al. [232], using *in vivo* blocking of proteasome function, indicated that the proteasomal pathway is involved in the regulation of activity of some ligninolytic enzymes (such as laccase) under nutrient deprivation in lignin-degrading Basidiomycete *Phlebia radiata*. The 26S proteasome pathway has been implicated in aspects of secondary cell wall biosynthesis, such as in cotton fiber development [233]. Transgenic lines exhibited differences in expression of both cellulose and lignin biosynthesis genes, resulting in increased levels of lignin or lignin-like phenolics [233]. Another strong association identified by the GWAS analysis was XBAT32. The association of XBAT32 is interesting, as it was observed to be associated with the S and G phenotypes independently but was not observed to be associated with the S/G ratio. This is quite peculiar since the S/G ratio is quite literally the quotient of the S and G unit measurements. Of the three primary monolignols found in poplar, H units are typically present in low abundance (approximately 1%), and as a result, the ratio of S and G units tend to vary proportionally. XBAT32 was initially identified as a regulator of lateral root development [234]. It was later shown to regulate ethylene biosynthesis [235] and, by extension, a response to abiotic stresses [236]. Another XBAT protein, XBAT35, has also been shown to be involved in ethylene signaling [237]. A similar gene identified in cotton, GhXB32A, was also shown to function in response to stress [238]. It is well-documented that stress can induce changes to lignin properties [239]. It is therefore quite reasonable to implicate the stress related XBAT32 gene could be influencing lignin structure.

In conclusion, the lignin of 409 unique, natural variant *P. trichocarpa* genotypes were analyzed by HSQC NMR. A subsequent GWAS analysis identified 756 SNPs significantly associated among the twelve lignin phenotypes. The GWAS results include putative lignin and cell wall biosynthesis related genes. Subsequent gene ontology analyses show that cell wall related term enrichment increases with GWAS significance levels. These results provide evidence that the GWAS analyses identified causal genes, rather than randomly sampling the genome. Several candidate genes not previously associated with lignin or cell wall biosynthesis were identified by GWAS, including GDU1, PBD1, and XBAT32. These GWAS results can be used as targets for future work investigating lignin structure, and the functional characterization of these genes may reveal novel genetic mechanisms controlling lignin biosynthesis.

Chapter 8:

Novel candidate genes for lignin structure revealed by NMR analysis of switchgrass GWAS panel

8.1 Abstract

Switchgrass is a phenotypically diverse C4 grass that has been designated as a U.S. Department of Energy bioenergy crop due to its high cellulose content, stress tolerance, and ability to grow on marginal land. The secondary cell wall contains lignin – a complex biopolymer that exhibits a high degree of variability and plasticity. Lignin structure is often cited as a major barrier in the biological conversion of biomass to biofuel. The formation of lignin requires the coordination of numerous metabolic pathways. Yet, there are many potential unknown genetic mechanisms influencing lignin biosynthesis. To address this knowledge gap, we have integrated with deep phenotyping capabilities of heteronuclear single quantum coherence nuclear magnetic resonance with the power of genome wide association studies (GWAS) to elucidate genetic associations with lignin structure. Four GWAS methodologies identified associations with 297 genes across 12 lignin phenotypes, including several putative genes influencing lignin biosynthesis. The genes identified by this GWAS can be used as targets for influencing lignin and/or cell wall structure for improvements in switchgrass.

8.2 Introduction

Switchgrass is a phenotypically diverse C4 grass that exhibits high cellulose content, tolerance to biotic and abiotic stresses, and can also grow on marginal land [240]. Switchgrass also exhibits high yield potential, with top end estimates on the order of 23-26 Mg ha^{-1} [241]. Switchgrass has evolved into upland, lowland, and coastal ecotypes, with morphological adaptations based on their unique geographical regions [242].

Lignin is the second most abundant terrestrial biopolymer and one of the three major components of the secondary cell wall (along with cellulose and hemicellulose). It is mainly comprised of three primary phenylpropane units: syringyl (S), guaiacyl (G), and *p*-hydroxyphenyl (H) [29]. However, lignin demonstrates incredible plasticity with the ability to incorporate many different monomers as building blocks [243]. Lignin moieties exist at varying levels of abundance and will form a variety of interunit linkages when incorporated into the lignin polymer. As such, the resulting lignin structure is highly complex and irregular. The U.S. Department of Energy has identified biomass feedstock variability as a major barrier to commercialization of biofuels [8].

The structure of lignin is often cited as a major barrier in converting biomass to biofuels - a term referred to as “recalcitrance”. Therefore, there is great interest in elucidating the structure of lignin. In the context of GWAS, a large sample size is required. This leads to the utilization of high throughput methods of phenotyping. Previous switchgrass GWAS analyses have utilized high throughput methods such as pyrolysis molecular beam mass spectroscopy (Py-MBMS) [244] or near infrared (NIR) spectroscopy [245] to quickly phenotype hundreds of samples. Each of these techniques has associated advantages and disadvantages [246]. However, such high throughput techniques can analyze few (if any) lignin phenotypes beyond the S/G ratio. An analytical method that is able to elucidate a comparatively large number of lignin phenotypes is heteronuclear single quantum coherence (HSQC) nuclear magnetic resonance (NMR) [247], though it is not traditionally considered a high throughput technique due to sample preparation being lengthy and labor intensive, in addition to the long pulse sequence (~6h). In addition to detecting S and G units, HSQC NMR can elucidate structural features such as acylated monolignols, interunit linkages, and end units.

The formation of the secondary cell wall requires coordination of many metabolic pathways [35], presenting challenges in linking phenotypes to genetic mutations. Genome wide association studies (GWAS) have demonstrated great promise for identifying the loci that contribute to phenotypic variation. However, complex traits, such as lignin, are often controlled by multiple genes. While single-locus GWAS have been used extensively [248-252], they may be insufficient to detect such polygenic traits. To address this shortcoming, multi-locus GWAS methods have been developed and are becoming more commonly utilized. In this study, for the first time, HSQC NMR will be employed for population level phenotyping the lignin of switchgrass. The resulting GWAS analyses are expected to elucidate novel associations between SNPs and lignin phenotypes in switchgrass.

8.3 Materials and methods

8.3.1 Plant materials

A diverse panel of 308 tetraploid *Panicum virgatum* genotypes, composed of upland, lowland, and coastal ecotypes, mostly from the Juenger GWAS panel [242], was established in April 2019

at UGA's Iron Horse Farm (33.728430, − 83.303594) in Watkinsville, GA (Figure 8-1). Field layout followed an established method [253], using a randomized honeycomb design with 3m between linear plants. In December 2019, aboveground plant biomass was harvested at 10cm above ground level, subsampled, chipped, and dried at 50º C. Samples were milled using a Wiley Model 4 mill with a 1mm (18 mesh) screen.

8.3.2 Phenotyping

A total of 308 unique genotypes were phenotyped, and 25 genotypes were randomly analyzed in duplicate as technical replicates. Dried and knife milled biomass samples were prepared for NMR similar to published procedures [254]. Briefly, each sample was Soxhlet extracted using toluene/ethanol (2:1, *v:v*) for at least eight hours and subsequently air dried for at 24 hours. The extractives free biomass was then ball milled to a fine powder on a Retsch planetary ball mill at 600 RPM for 2h in 5-minute intervals to avoid overheating. A 48h enzymatic hydrolysis was then performed using cellulase (Sigma-Aldrich) and sodium acetate buffer (pH 4.8). The lignin enriched residue was then separated *via* centrifugation, washed twice with DI water, and lyophilized for 48 hours. The resulting enzyme lignin (EL) was used for NMR analysis. Lignin structure was analyzed by 2D HSQC NMR with a Bruker Avance III 500-MHz spectrometer. Approximately 40 mg EL was combined with 0.6 mL of DMSO-d_6 in a 5mm NMR tube. The Bruker pulse sequence hsqcetgpsip2.2 was utilized on a N_2 cryoprobe with the following parameters: spectra width of 12 ppm in the ^{1}H dimension with 1024 data points; spectra width of 220 ppm in the ^{13}C dimension with 256 increments, and 32 scans. The HSQC spectra were analyzed with Bruker TopSpin 3.5pl6 software. The DMSO-d_6 solvent peak (δ_C/δ_H at 39.5/2.49) was used to calibrate the chemical shifts.

8.3.3 SNP variant calling and genome wide association analysis

Methods for single nucleotide polymorphism (SNP) variant calling follow Lovell et al. (2021) [242]. Briefly, whole-genome resequencing of 287 genotypes was performed using Illumina genetic analyzers at the Department of Energy (DOE) Joint Genome Institute. We applied multiple filtering criteria on the raw SNP dataset: removed SNPs with more than 10% missing genotypes, removed genotypes with more than 10% missing SNPs, and removed SNPs that had a

FIGURE 8-1 ARIAL VIEW OF GWAS PANEL FIELD SITE

severe departure from Hardy Weinberg Equilibrium (SNPs with HWE *p*-value > 1E-50 removed). We further filtered SNPs with minor allele frequency (MAF) < 0.05 and SNPs with linkage disequilibrium (LD) $r^2 \geq 0.7$, which provided a total of 4,691,989 SNPs corresponding to 285 genotypes for GWAS analysis.

GWAS analysis of all 22 HSQC phenotypic traits was performed using one single locus and three multilocus methods available in the GAPIT3 R package [255]. We first removed outliers in the phenotypic data using the median absolute deviation (MAD) method where any phenotypic values with MAD > 6 were removed for a given trait. The single-locus method used a general mixed linear model with phenotypic data, a SNP data matrix (4,691,989 SNPs), a genetic relationship matrix (GRM), and the first two SNP principal components (PCs) of the SNP data matrix (covariates). The VanRaden method [256] is used to estimate the genetic relationship matrix to control for background genetic variation. The mixed linear model in the GAPIT3 software package in R [255] was used to test for the association between SNPs and the variation in the 22 HSQC phenotypic traits using the following general mixed linear model:

$$Y = X\beta + Zu + \alpha + \varepsilon \quad (8\text{-}1)$$

where Y is the N×1 phenotypic vector with N being the sample size, β is the M×1 vector of additive SNP effects with M being the number of SNPs, X is an N×M incidence matrix, u is an N×1 vector of random effects of the genotypes, Z is an N×N genetic relationship matrix, α is an N×1 vector of random effects that include the genetic relationship matrix (GRM), and ε is an N×1 vector of residual random effects.

The three multilocus methods used either (i) only the fixed effect models (Bayesian-information and LD iteratively nested keyway, BLINK) or (ii) the combination of fixed and random effects models (fixed and random model circulating probability unification, FarmCPU; and multiple loci mixed model (MLMM). BLINK [98] uses two fixed effect models (FEM) – the first FEM model generates a set of pseudo–Quantitative Trait Nucleotides (QTNs), and the second FEM optimizes the selection of pseudo QTNs using the pseudo QTNs from the first FEM model. These two

FEM models are used iteratively but stop when there is no change in the pseudo QTNs in the second FEM model. For the association test of a given SNP, significant SNPs that are not in LD are used as covariates. Below is the first FEM model:

$$y_i = S_{i1}b_1 + S_{i2}b_2 + \cdots + S_{ik}b_k + S_{ij}d_j + e_i \quad (8\text{-}2)$$

where, y_i is the phenotypic value of the i^{th} individual, S_{i1}, …, S_{ik} are the genotypes of the k QTNs, b_1, …, b_k are the corresponding effects of the QTNs, S_{ij} is the genotype of the i^{th} individual and j^{th} SNP, d_j is the j^{th} SNP effect, and e_i is the residual.
The second FEM model (below) is the reduced version of Equation (8-2) such that the $S_{ij}d_j$ term, which tests for the association of the SNP with the phenotypic trait, is removed:

$$y_i = S_{i1}b_1 + S_{i2}b_2 + \cdots + S_{ik}b_k + e_i \quad (8\text{-}3)$$

Fixed and random model circulating probability unification (FarmCPU) uses the combination of fixed and random and fixed effect models iteratively. The first model is a simple mixed linear model which uses a kinship matrix (random effect) and SNP PCs from the SNP dataset as covariates to control for population structure. The associated markers obtained from this model are used to optimize the kinship matrix in the second model using maximum likelihood estimation [255].

Multiple loci mixed model (MLMM) is an extension of the MLM model where the significantly associated markers are used as covariates in the GWAS model [255].

We used False Discovery Rate (FDR) method as implemented in the GAPIT3 R package [255] to control for multiple testing due to a large number of independent association tests. We considered SNPs passing the FDR threshold of 0.05 (FDR *p*-value ≤ 0.05) as significant and the SNPs passing the FDR threshold of 0.2 (FDR *p*-value ≤ 0.2) as suggestive associations. Our final set of significant/suggestive significant SNP sets included the union of GWAS output from all four GWAS methods. We obtained *Panicum virgatum* candidate genes using the nearest gene left and right of the significant/suggestive significant SNPs.

8.3.4 Markov clustering on phenotypic traits

We first removed outliers in the phenotypic data using the median absolute deviation (MAD) method where any phenotypic values with MAD > 6 were removed for a given trait. We applied Markov clustering, an unsupervised graph clustering algorithm, to the predictive phenotype network for 22 HSQC traits to identify the functional modules. The phenotype network used the edge threshold of 0.1. For Markov clustering, we used an inflation factor of 2.5 to control the granularity of the phenotypic groups. The resulting four distinct clusters are represented by different colors in the network. Traits within each clusters show high degree of interconnectedness given the topology of the predictive phenotype network.

8.3.5 Functional partitioning

We fed GWAS seed genes (*Arabidopsis* orthologs of the switchgrass genes) into the *Arabidopsis* multiplex network and generated the rank vectors corresponding to each of the GWAS seed genes. For a given GWAS seed gene, the rank vector contains all genes in the multiplex network ranked against that seed gene. We performed pairwise spearman rank correlation on the ranked vectors and converted to a pairwise distance matrix. Using the distance matrix thus obtained we constructed the hierarchical dendrogram showing the functional relationship of the GWAS genes (Miller et al. unpublished). The multiplex network contains following nine network layers:

- AraNet_CoEx - Arabidopsis Co-expression network
- KO - Arabidopsis Knock Out layer.
- Merged_PPI - Protein-Protein Interaction
- PEN_Aerial - Predictive Expression Network Aerial
- PEN_Leaf - Predictive Expression Network Leaf
- PEN_Root - Predictive Expression Network Root
- PEN_Seedling - Predictive Expression Network Seedling
- PEN_Shoot - Predictive Expression Network Shoot Regulatory

8.4 Results and discussion

8.4.1 Lignin phenotyping

A total of 22 signals were analyzed in the 2D HSQC NMR spectra. The integrated values of the $S_{2/6}$ and oxidized $S'_{2/6}$ signals were combined to calculate the total S unit content. Two signals corresponding to ferulates (FA_6, FA_2), four signals corresponding to tricin (T_3, $T_{2'/6'}$, T_6, T_8), and four signals associated with cinnamyl alcohol/aldehyde end units (I_α, I_β, J_α, J_β) were also integrated for analysis. Signals were integrated relative to the $S_{2/6}$ signal and all phenotypes are expressed on an S+G+H basis.

Among the three primary monolignols, the relative abundance of G was highest (~65%), followed by S (~29%). The average H unit content was measured to be 4.9%, resulting in a population average S:G:H ratio of approximately 29:66:5. *p*-coumarate (*p*CA) content was relatively abundant (average ~28%), but was also highly variable, ranging from approximately 11%-45%. Ferulates content averaged 3% to 5%, depending on which signal is considered, whereas each tricin signal integrated to ~1% abundance. A benefit of HSQC NMR is it enables the observation of lignin interunit linkages. Among interunit linkages, the β-O-4 ether linkage was the most abundant, averaging ~50%. Dibenzodioxocin (DBDO) and β-5 were also present at lower levels, averaging approximately 9% and 6%, respectively. β-β linkages were present in trace amounts, averaging just over 1%. A summary of each lignin phenotype is provided in the form of box plots in Figure 8-2.

Correlations among the phenotypic traits were also examined (Figure 8-3). The S/G ratio was most positively correlated with pCA (CC=0.57) and β-O-4 linkages (CC=0.37), and most negatively correlated with H units (CC=-0.29), $T_{2'/6'}$ (CC=-0.36), and β-5 linkages (CC=-0.50). Principal component analysis (PCA) was performed to examine variability in the NMR data (Appendix C). Twenty-two NMR signals were used in the analysis. The top two principal components PC1 and PC2 accounted for 27.1% and 18.4% of the variability, respectively. Top loadings for PC1 include S units, G units, S/G ratio, and three of the four tricin signals ($T_{2'/6'}$, T_6, T_8). The top loadings for PC2 include the cinnamyl alcohol and aldehyde end units (I_α, I_β, J_α, J_β), β-β linkages, and the FA_2 signal.

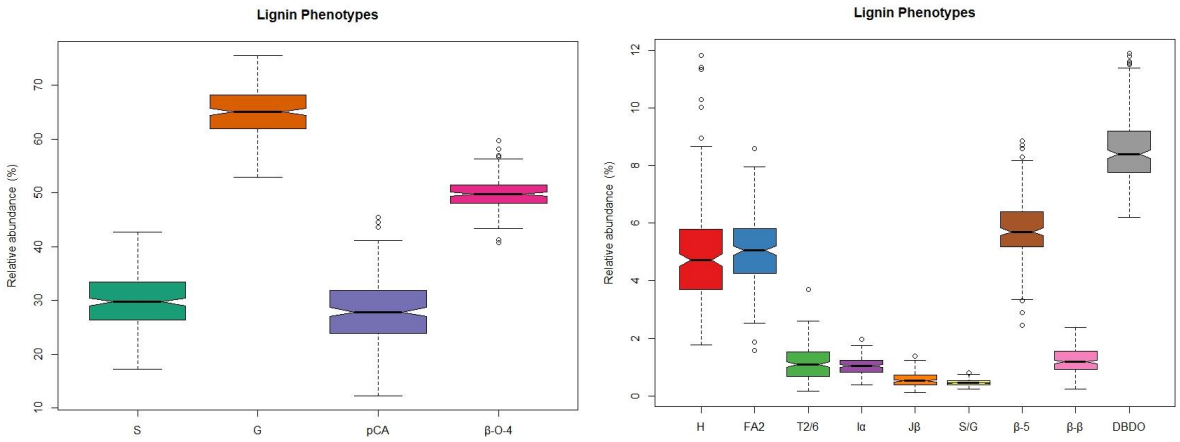


FIGURE 8-2 SUMMARY OF LIGNIN PHENOTYPE DISTRIBUTIONS

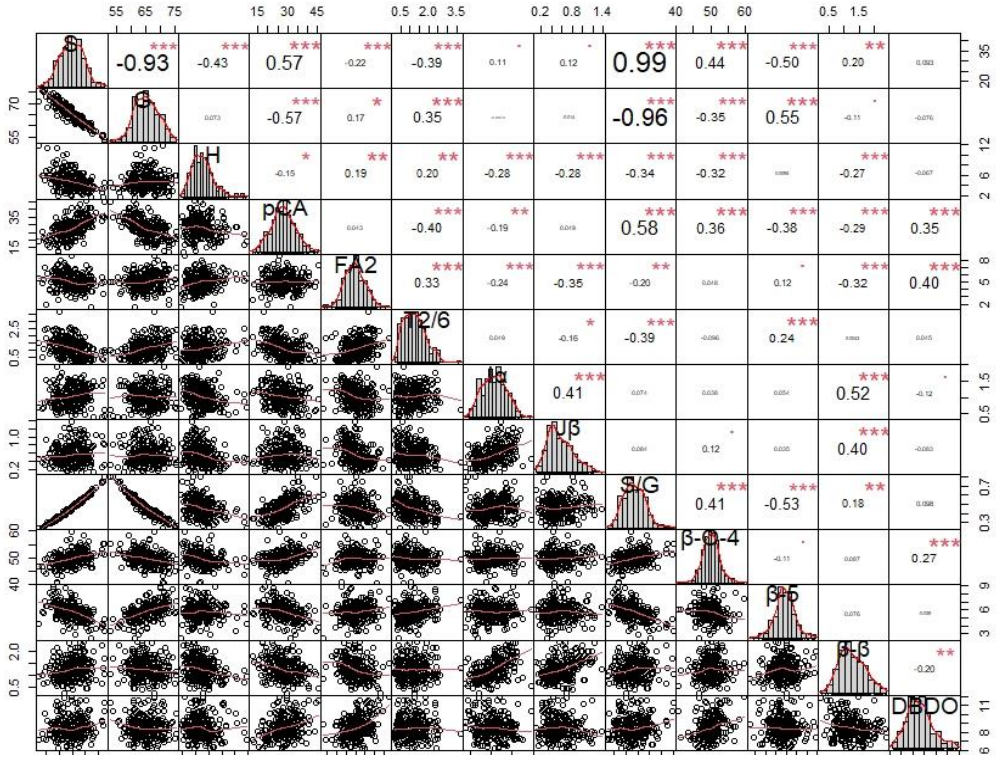


FIGURE 8-3 TRENDS AND CORRELATIONS OF LIGNIN PHENOTYPES

As the number of genotypes required for a sufficiently powered GWAS is relatively high (usually several hundred at minimum), most GWAS efforts rely on high throughput phenotyping techniques such as pyrolysis molecular beam mass spectroscopy (Py-MBMS) or near-infrared (NIR) spectroscopy [257].

Although HSQC NMR is a lower throughput technique, it provides an abundance of lignin phenotypic traits, and is especially useful in elucidating interunit linkages compared to other analytical methods. Indeed, from a literature review, there have been few studies that have examined the lignin structure of large numbers of switchgrass samples by HSQC NMR, making direct comparisons of lignin traits elusive. The S/G ratio of this GWAS population was measured to be 0.46 with a standard deviation of 0.11. Other studies that have analyzed switchgrass S/G ratio by techniques such as Py-MBMS or thioacidolysis typically report values in the range of 0.6-0.8 [258, 259]. It has previously been shown that while the S/G ratio determined by NMR does not exactly match the S/G ratio determined by Py-MBMS or thioacidolysis, the measurements do exhibit good correlation [260]. The *p*CA content varied greatly, ranging from 10.8%-45.4% with an average of 27.6%. This is likely due to *p*CA existing as as a pendant unit, which results in comparatively long relaxation times, and therefore tends to be overestimated by NMR. *p*CA was also positively correlated with S units. This is not unanticipated, as *p*CA has been shown to acylate the γ-carbon of S units [261]. However, the analogous *p*-hydroxybenzoate ester in *Populus* has been shown to be negatively correlated with S untis [78, 262]. The most abundant interunit linkage ratio was determined to be the β-O-4 aryl ether linkage, averaging 49.9%. DBDO, which serves as a main branching point in lignin, was the second most abundant interunit linkage with an average 8.6% abundance. Interestingly, DBDO was positively correlated with both *p*CA (CC=0.39) and FA_2 (CC=0.43) content. Each of the four tricin signals indicated this moiety was present at levels of approximately 1%. Tricin was observed to be negatively correlated with the S/G ratio, which has been observed previously [263]. Tricin is a flavone (part of the shikimate pathways but not the lignin biosynthesis pathway), and its incorporation into the lignin polymer is an excellent demonstration of lignin plasticity. Tricin only attaches via β-O-4 linkages and thus each lignin chain can contain only one tricin moiety. Er go, tricin is typically described as nucleation site. The lignin S/G ratio has been found to be

correlated with its molecular weight [74, 264], however a correlation between tricin and molecular weight was not observed by Eloy et al. [263]. As previously discussed, population level phenotypic traits of lignin is not extensively reported for switchgrass. Therefore, this lignin from this GWAS panel can be considered the largest, most extensively characterized switchgrass population to date.

8.4.2 GWAS

GWAS is an especially useful tool for determining the genetic basis of lignification, as the formation of the secondary cell wall requires coordination of many metabolic pathways. Complex traits (such as lignin) are usually polygenic, with potentially thousands of genes controlling traits. Additionally, there exist several methods available for conducting GWAS analyses. For this study, four GWAS models were employed. The single locus method mixed linear model (MLM) was used. Three multilocus methods were also used: multi-locus mixed-model (MLMM), fixed and random model circulating probability unification (FarmCPU), and Bayesian-information and linkage-disequilbrium iteratively nested keyway (BLINK). Manhattan plots for each of these four methods associated with the S unit phenotype are displayed in Figure 8-4. The four GWAS methods identified 460 SNPs associated with 15 lignin phenotypes below an FDR<0.2 threshold. These SNP associations were assigned to 199 unique genes. Of these 199 genes, 130 had orthologs annotated in *Arabidopsis thaliana* and 134 had homologs annotated in *Oryza sativa*.

The GWAS results identified several bona fide candidate genes that have been definitively proven to impact lignification. One of these is Pavir.1NG376500 (annotated as MYB4 in *Arabidopsis*), which was associated with the S unit phenotype, has previously been shown to impact the S/G ratio in switchgrass [265]. The S unit phenotype was also associated with Pavir.1NG376500 (WRKY12), which increased S unit (and total) lignin content in *Medicago truncatula* [266]. While this result is already known, it provides confidence that the GWAS has indeed identified authentic candidate genes.

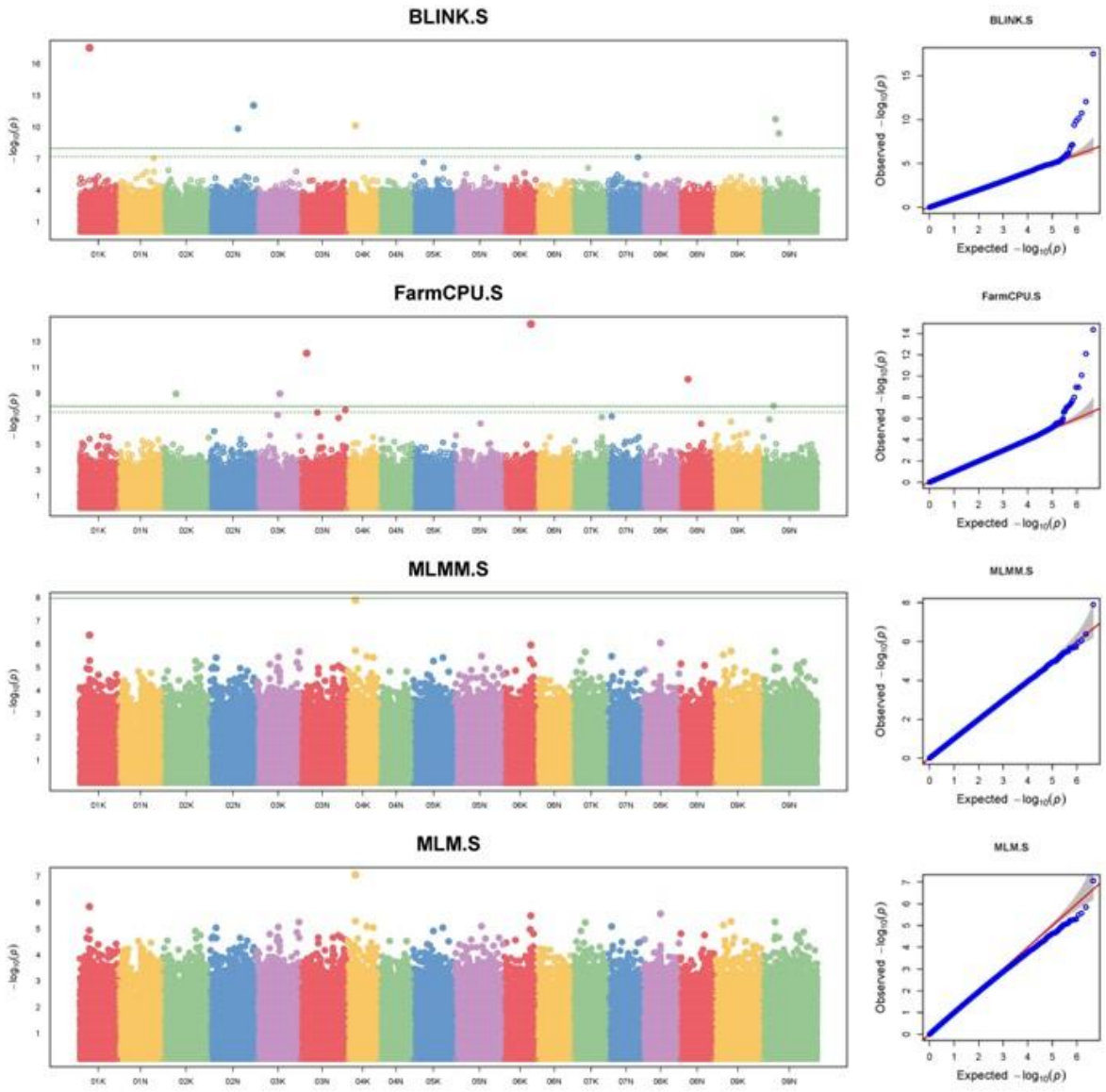


FIGURE 8-4 MANHATTAN AND QQ PLOTS OF S UNIT GWAS CORRELATIONS

While the GWAS identified authentic causal loci influencing lignin biosynthesis, unverified genes should be examined for additional lines of evidence (LOE) to invoke confidence in valid candidate gene identification as opposed to false positive associations. LOE can come from factors such as strength of association (*p*-value), multiple methods of detection, and support from current literature. For instance, one of the strongest associations identified was the *p*CA lignin phenotype with the SNP position Chr09K_28154416 (p-value = 5.85E-19), which corresponds to the genes Pavir.9KG166900 and Pavir.9KG167413. These are annotated as *Arabidopsis* orthologs AT1G71880 (sucrose -proton symporter 1) and AT1G69780 (Homeobox-leucine zipper protein family), respectively. The strong association between this SNP position and *p*CA levels indicate that one of these genes may influence the biosynthesis of *p*CA. Multiple methods of detection can also contribute to LOE. For instance, while the Pavir.9KG166900 and Pavir.9KG167413 genes were identified as candidate genes by the BLINK method, they were likewise identified by the FarmCPU method, again with a strong *p*-value (*p*-value = 9.25E-15). Detection by multiple GWAS methods provides more confidence that an identified association may be an authentic causal gene. Another LOE that may indicate an authentic candidate gene is support from current literature. As previously mentioned, MYB4 was found to be associated with S units. However, interestingly, AT4G38620 (MYB4) was also associated with the dibenzodioxocin linkage. While several studies have investigated MYB4 transgenic plants, none have specifically analyzed the dibenzodioxocin linkage. This represents an opportunity to explore potentially unexamined impacts of a gene well known to impact lignin biosynthesis.
The β-O-4 ether bond of lignin is of special interest since it is often referenced as the most readily cleavable linkage, and higher abundance of β-O-4 tends to promote delignification and sugar release. GWAS identified 16 genes associated with the β-O-4 phenotype, all of which were identified by the MLMM method. Of these 16 identified genes, 10 had Arabidopsis orthologs available, an additional three had Maize orthologs available, while three lacked annotation. One interesting gene associated with the β-O-4 phenotype was Pavir.5NG413000, which is annotated in *Arabidopsis* as AT4G21700 (Protein of unknown function [DUF2921]) and classified as a trans-Golgi network protein highly expressed in root tissue [267]. However, a predicted functionally related protein of AT4G21700 is GAUT10, which is a galacturonosyltransferase.

Other GAUT family proteins have previously been shown to impact lignin and cell wall structure [268, 269], which suggests that AT4G21700 could have a similar function. Other members of the AT4G21700 gene neighborhood (genes that reside in similar genomic region) are involved in protein transport and membrane trafficking. It should be noted that monolignols are synthesized in the cytoplast, and it remains unresolved how they are translocated to the apoplast where they participate in polymerization. Additionally, while biological control has been demonstrated in the monolignol coupling of lignans [270], current evidence indicates that lignification occurs without biological (i.e., enzymatic) control. Therefore, while it not expected that genes such as Pavir.5NG413000 exhibit direct control over monolignol coupling, it is possible that it may influence conditions that promote the formation of the β-O-4 linkage.

Multiple WRKY transcription factors (TFs) were also identified in the GWAS results. The MLMM method identified tricin to be associated with Pavir.7NG317000, which has the *Arabidopsis* ortholog (AT1G30650) annotated as WRKY14. Additionally, the BLINK method identified an S unit association with Pavir.1NG376500, which is annotated as WRKY12 in *Arabidopsis* (ortholog AT2G44745). In maize, these two genes are annotated WRKY37 and WRKY36, respectively. Genes in the WRKY domain contain a highly conserved WRKYGQK motif and a C_2H_2 zinc finger sequence at their C termini. WRKY genes constitute one of the largest transcription factors in plants and play roles in stress defense, development, and metabolism [271]. Various WRKY TFs have been shown to influence the production of secondary metabolites, including lignin [272-274]. Specifically, WRKY12 has been shown to influence the production of S unit in *Medicago truncatula*. It is important to note that lignin is derived from the phenylpropanoid pathway and therefore competes for carbon flux with other phenolic based secondary metabolites (i.e., flavonoids, tannins, etc.) [275]. The expression of WRKY transcription factors was found to be regulated by a R2R3 MYB and bHLH transcription factors [276], illustrating the complex regulatory mechanisms underlying cell wall biosynthesis. Understanding the complex regulatory mechanisms is crucial for elucidating and controlling lignin and cell wall formation [277].

8.4.3 Functional partitioning

Markov clustering was performed to functionally partition the genes identified by GWAS, with the 22 NMR phenotypes segregated into four major clusters. Cluster 1 consists of the phenotypes associated with primary monolignol signals (S, G, S/G ratio, H), coumarate and ferulate signals, and three interunit linkage signals (β-5, β-O-4, DBDO). Cluster 2 consists of the two xylan signals. Cluster 3 includes signals associated with cinnamyl end units (alcohol and aldehyde), along with the β-β interunit linkage. Cluster 4 exclusively contained the four tricin signals. Functional partitioning works on the premise that causal genes are likely to be functionally connected to one another. At the 0.55 significance threshold, random walk with restart (RWR) identified 125 genes in Cluster 1, two genes in Cluster 2, 59 genes in Cluster 3, and 26 genes in Cluster 4.

True positive causal genes are likely to be functionally related to one another. Therefore, functional partitioning clades can be examined to explore functionally similar genes. For instance, the Cluster 1 RWR returned an interesting clade which includes MYB4 (AT4G38620), NAC domain containing protein 35 (AT2G02450), and WRKY12 (AT2G44745). Similar to MYB4, WRKY12 is a transcription factor that regulates the expression of numerous downstream genes that influence secondary cell wall formation [35]. Likewise, NAC (NAM, ATAF, and CUC) transcription factors are involved in numerous biological processes, such as growth, development, and stress response [278]. Therefore, the genes included in this clade may exhibit similar transcriptional regulation of several genes involved in secondary cell wall formation. Utilizing a more discriminating significance threshold of 0.45, MYB4 is functionally related to Homeobox-leucin zipper protein family (AT1G69780), Protein with RING/U-box and TRAF-like domains (AT5G37870), Chalcone and stilbene synthase family protein (AT5G13930), arginine decarboxylase 2 (AT4G34710), and basic helix-loop-helix (bHLH) DNA-binding superfamily protein (AT3G06120). AT5G13930 particularly may be a gene of interest, as this gene plays a key role in the biosynthesis of flavonoids – which, like lignin, are derived from the phenylpropanoid pathways [279]. Indeed, an Arabidopsis mutant lacking AT5G13930 was defective in the gene encoding flavanone 3'-hydroxylase (F3'H), which catalyzes the production

of quercetin and isorhamnetin derivatives [280]. Again using the 0.45 significance threshold, WRKY12 was found to be functionally similar to cyclophilin-like peptidyl-prolyl cis-trans isomerase family protein (AT2G21130), AWPM-19-like family protein (AT1G04560), and DPP6 N-terminal domain-like protein (AT1G21680). AT1G04560 and AT2G21130 are both associated with the *p*CA phenotype, with AT2G21130 identified by both the FarmCPU and BLINK methods. Interestingly, at the 0.55 significance threshold, a clade in Cluster 3 is also observed to contain a NAC domain containing protein (NAC058; AT3G18400) and a WRKY DNA-binding protein (WRKY14; AT1G30650). The other genes in this clade include Protein of unknown function 642 (AT1G29980), Phosphate-responsive 1 family protein (AT4G08950), IQ-domain 19 (AT4G14750), DHHC-type zinc finger family protein (AT3G56930), and transducing/WD40 repeat-like superfamily protein (AT4G38480). These results suggest that certain NAC, WRKY, and/or functionally related genes may be functioning in groups or complexes to regulate aspects of secondary cell wall formation, which in turn influences lignin phenotypes.

8.5 Conclusion

Despite being designated by the U.S. Department of Energy as a bioenergy crop [281], studies examining switchgrass lignin by genome wide association are limited. To this end, we have integrated the deep phenotyping of HSQC NMR with the power of GWAS to elucidate genetic associations with lignin structure. A combination of four GWAS methodologies identified associations with 297 genes across 12 lignin phenotypes. This includes several genes known to influence lignin biosynthesis, indicating the GWAS returned authentic candidate genes. The genes identified by this GWAS can be used as targets for influencing lignin and/or cell wall structure for improvements in switchgrass.

Chapter 9:

GWAS directed characterization of lignin and cell wall phenotypes

9.1 Abstract

There are many aspects of lignification that remain a mystery. Our previous poplar GWAS study suggested that there may be novel candidate genes that influence lignin biosynthesis, such as the molybdate transporter MOT1. A previous sampling effort revealed a lognormal distribution of molybdenum (Mo) content across a population of natural variant poplar. This same distribution trend was observed with *p*-hydroxybenzoate (PB) units. Additionally, genotypes with differential Mo content (used as a proxy for MOT1 expression) exhibited differential levels of PB, S/G ratio, lignin content, and xylose content. These results suggest that MOT1 may play a role in the lignin metabolism.

Also, from our previous poplar GWAS results, we have the unique opportunity to select genotypes with specific characteristics for additional studies. Here, we strategically selected genotypes with similar lignin structure – except for PB content – which differed significantly between these paired samples. The cell wall architecture of these samples was extensively examined to test the hypothesis that PB impact cell wall structure and/or sugar release (polysaccharide accessibility). Further analysis of quantitative PB content, the lignin carbohydrate complex, compositional analysis, and biomass crystallinity revealed that cell wall structures between high/low PB genotypes were not significantly different. A subsequent sugar release experiment indicated that polysaccharide accessibility was not dependent on PB content.

Both studies described in this chapter demonstrate the benefits and some of the numerous potential work it can enable. Similar studies can be extended to other concepts with poplar or could be extended to consider samples from the switchgrass GWAS to achieve novel insights about cell wall structure and/or genetic mechanisms underlying lignin biosynthesis.

9.2 Introduction

There are many aspects of lignification that remain a mystery. As complex traits such as lignin are highly polygenic, there may be yet undiscovered genes that influence lignin structure. The poplar GWAS results suggested that the MOT1 gene may influence lignin biosynthesis. Additionally, the benefits and biological function of certain lignin moieties, such as *p*-hydroxybenzoate (PB), remain poorly understood. With a well characterized library of over 400

unique poplar genotypes, we are provided a unique opportunity to examine the impact of specific lignin moieties such as PB.

The *Populus* GWAS results identified a correlation between β-β linkages and a molybdate transporter (MOT1). Molybdenum (Mo) is an essential micronutrient for plants and serves as a cofactor for enzymes involved in nitrate assimilation, sulfite detoxification, abscisic acid biosynthesis, and purine degradation [282]. For instance, the nitrogen reductase enzyme requires Mo, and therefore Mo is associated with nitrogen assimilation. A deficiency or surplus of Mo can impact plant growth and development [283]. The presence of molybdenum transporters was discovered surprisingly recently, as it was previously classified as a sulfate transporter (though does still belong to the sulfate transporter superfamily) [284]. The effects of molybdenum are far reaching, exhibited by Mo deficiency in mutant *Arabidopsis* impacting differential expression of 187 genes [285]. It should be noted that although MOT1 was found to be correlated with the β-β linkage phenotype, there is no definitive evidence of biological (i.e., protein or enzyme) control over the type of interunit linkages formed by monolignols [286]. Rather, monolignol coupling is under "simple" chemical control – that is, dependent upon availability of reactants (monolignols), stabilizing post-coupling reactions, physical conditions, etc. Therefore, part of this chapter aims to evaluate the potential impacts of MOT1 on lignin structure. Since the propagation of transgenic plants takes an extended amount of time, naturally variant genotypes with differential molybdenum levels – which is highly correlated with MOT1 expression [284] – were instead used to explore correlations between molybdenum content and lignin structure.

As has been stated numerous times in this dissertation, lignin is a complex biopolymer. Additionally, while lignin is often cited as a major contributor toward biomass recalcitrance, this is a multifaceted issue, and the influence of individual lignin moieties is difficult to elucidate. This is exemplified in a recent study which sought to study the effects of PB in poplar [77]. While the introduction of a bacterial chorismate pyruvate lyase gene into transgenic poplar did increase the amount of PB, it also induced changes to lignin content and S/G ratio. This is not unexpected, as PB is correlated with syringyl (S) units [78, 262]. However, it then becomes unclear whether improved delignification from pretreatment is due to changes in PB level, lignin

content, or the S/G ratio – or perhaps even an interaction between these factors. Based on our previous work [262] we have the unique opportunity to strategically select samples from a well characterized library of over 400 unique *Populus* genotypes for further studies that target a single lignin trait. That is, the library can be examined and samples selected such that they exhibit differences in a single lignin phenotype, with most other lignin phenotypes being controlled to similar levels. Based on the recent interest in the PB moiety, we strategically selected genotypes which exhibited differential levels of PB, but otherwise retained similar lignin structure (S/G ratio, β-O-4 linkages, etc.). As S/G ratio is often correlated with various cell wall properties and recalcitrance, paired samples were selected at two different S/G ratios. These samples were extensively examined to test the hypothesized impact of *p*-hydroxybenzoate on cell wall properties and sugar release.

9.3 Materials and methods

9.3.1 HSQC NMR

Enzyme lignin of each sample was characterized by a previous study. Briefly, biomass was Soxhlet extracted for at least 8 hours using toluene:ethanol (2:1, v:v) and subsequently air dried. Biomass was then ball milled in a Retsch PM100 planetary ball at 600RPM for 2h using 5-minute intervals. Ball milled biomass was then subjected to enzymatic hydrolysis with cellulase in a sodium acetate buffer (pH 4.8) for 48h. The resulting enzyme lignin (EL) was centrifuged and washed with deionized (repeated twice) and subsequently lyophilized. EL dissolved in DMSO-d_6 and analyzed on a Bruker Avance III 500 MHz spectrometer. For HSQC analysis of whole cell wall (WCW) biomass, ball milled material was dissolved directly in DMSO-d_6/HMPA-d_{18} (4:1, v:v) in a 5mm NMR tube. For both EL and WCW, the Bruker pulse sequence hsqcetgpsip2.2 was utilized on a N_2 cryoprobe with the following parameters: spectra width of 12 ppm in the ^{1}H dimension with 1024 data points; spectra width of 220 ppm in the ^{13}C dimension with 256 increments and 32 scans. The HSQC spectra were analyzed with Bruker TopSpin 3.5pl6 software. The DMSO-d_6 solvent peak (δ_C/δ_H at 39.5/2.49) was used to calibrate the chemical shifts.

9.3.2 Compositional analysis

Prior to performing compositional analysis, biomass was dried under vacuum at 45°C for at least 48 hours. Compositional analysis was performed by using a two-step acid hydrolysis method according to the literature. Briefly, extractives free biomass was hydrolyzed at 30°C in 72% sulfuric acid for 1h. The mixture was then diluted to 4% sulfuric acid with deionized water and autoclaved for 121°C for 1h. The soluble and insoluble fractions were separated by vacuum filtration through a glass filter. The carbohydrate fraction was measured by an Agilent 1200 series high performance liquid chromatography (HPLC) equipped with a refractive index detector and an 87H column (Bio-Rad Laboratories Inc. Hercules, CA) at 60°C and a flow rate of 0.5 mL min^{-1} with 5mM H_2SO_4 as the mobile phase. Glucose and xylose standards were used for calibration.

9.3.3 CASA lignin content

Due to limited material availability for some samples, another method was required to estimate lignin content. The cysteine assisted sulfuric acid (CASA) method was selected for lignin content estimation since it has been shown to be consistent with several types of biomass, including *Populus* [287]. CASA lignin analysis was performed per the prescribed procedure. Briefly, a stock solution of 0.1 g mL^{-1} of L-cysteine in 72% sulfuric acid was prepared. In a 4mL glass vial, 5-10 mg of extractives free biomass was combined with 1.0 mL of stock solution and stirred at room temperature for at least 1h. The completely dissolved biomass mixture was then transferred to a volumetric flask and diluted to 50mL with deionized water. The absorbance of the diluted solution was measured at 283nm in a 1cm quartz cell using a UV spectrophotometer against a blank containing the stock solution. The lignin content was calculated based the Beer-Lambert law using formula (1). All tests were performed in duplicate. CASA lignin results were compared with a well characterized reference *Populus* biomass (BESC-131).

$$CASA\_L(\%) = \frac{Abs_{283} \times V}{\varepsilon \times m_s \times L} \times 100 \quad (9\text{-}1)$$

where Abs_{283} is the UV absorbance at $\lambda = 283nm$; V is the total volume (liters) of the diluted solution; m_s is the mass of the extractives free biomass (grams); L is the path length (1cm); ε is the UV-absorption coefficient of lignin at $\lambda = 283nm$, and the recommended 11.23 g^{-1} L cm^{-1} was used here.

9.3.4 Sugar release

The glucose and xylose release of these extractives free, unpretreated, natural *P. trichocarpa* variants were determined by enzymatic hydrolysis. Enzymatic hydrolysis was performed with extractives free biomass at 2% (w/v) loading. The enzyme blend CTec2 (sigma SAE0020), which contains cellulases, β-glucosidases, and hemicellulase, was added at a loading of 24 filter paper units per gram glucan. Sodium azide was added as an antimicrobial agent at a final concentration of 0.0004 g/mL. An enzyme blank and a substrate blank were also prepared to account for glucose stabilizers sometimes found in cellulase. Samples were placed in an incubator shaker at 50°C and 150 RPM at 9:30AM on 5/24/2023. All samples were prepared in duplicate. Aliquots of 0.5mL were taken at 4h, 24h, 48h, 72h, and 120h and each hydrolysate was filtered through a 0.45mm syringe filter. Hydrolysates were analyzed by an Agilent 1200 series high performance liquid chromatography (HPLC) equipped with a refractive index detector and an 87H column (Bio-Rad Laboratories Inc. Hercules, CA) at 60°C and a flow rate of 0.5 mL min^{-1} with 5mM H_2SO_4 as the mobile phase. Glucose and xylose standards were used for calibration.

9.3.5 Chemical analysis of *p*-hydroxybenzoate

The amount of *p*-hydroxybenzoate (PB) was determined by an established method (Goacher et al., 2021) with slight modifications. Briefly, 1 mL of 2M sodium hydroxide and 100µL of 1 mg/mL *o*-coumaric acid (internal standard) was added to 20mg of extractive-free ball milled biomass. Samples were incubated at 30°C for 24h and the reaction was subsequently terminated by the addition of 100 mL of 72% sulfuric acid. Samples were then incubated on ice for 5 minutes. The supernatant was collected by centrifugation and filtered through a 0.45mm nylon syringe filter prior to high-performance liquid chromatography (HPLC) analysis. Samples were analyzed by an Agilent 1200 series HPLC. For each sample, 10µL was injected into a Symmetry

C18 column (4.6 x 50mm, 5mm particle size) maintained at 35°C. Adequate peak resolution was achieved using 75:25 (v:v) ratio of eluent B (0.1% trifluoroacetic acid in 70:30 acetonitrile:methanol) in eluent A (0.1% triflouroacetic acid in water) at a flow rate of 0.4mL/min. Spectra were integrated at the UV maxima of 255nm. A five point calibration curve was created with 4-hydroxybenzoate (Sigma Aldrich). The PB measurement obtained by HPLC was normalized by lignin content as determined by the CASA method for comparison to HSQC NMR results.

9.3.6 Biomass crystallinity by NMR

The crystallinity index (CrI) of cellulose was measured by ^{13}C cross-polarization/magic angle spinning (CP/MAS) NMR as described in previous work. Briefly, xtractives free whole cell wall biomass was packed into a 4mm cylindrical ceramic MAS rotor. Solid-state NMR analysis was carried out on a Bruker Avance 400 MHz spectrometer operating at frequencies of 100.55MHz for ^{13}C in a Bruker double-resonance MAS probe at spinning speeds of 8kHz. CP/MAS experiments were carried out with a 5ms (90^{o}) proton pulse, 1.5ms contact pulse, 4s recycle delay, and 4k scans.

9.4 Results and discussion

9.4.1 Characterization of MOT1 genotypes

The Populus GWAS identified promising associations with several interesting candidate genes. One such gene is AT2G25680, which is associated with β-β interunit linkages. AT2G25680 is annotated as molybdate transporter 1 (MOT1) in *Arabidopsis*. Two *Populus* genes, Potri.006G245900 and Potri.006G246000, have similar homology with AtMOT1. Potri.006G245900 is most highly expressed in roots, modestly expressed in leaves, and barely expressed in xylem (Figure 9-1). Potri.006G246000 likewise is most highly expressed in roots, but also has appreciable expression both leaves and xylem.

Examination of the xylem molybdenum (Mo) content reported across the population reveals a non-normal distribution of Mo levels (p-value < 0.001).

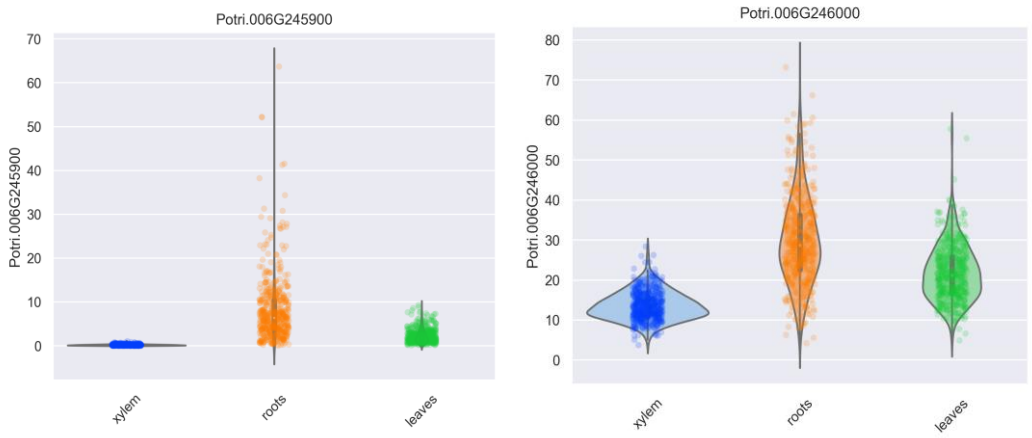


FIGURE 9-1 DISTRIBUTION OF POTRI.006G24500 AND POTRI.006G246000

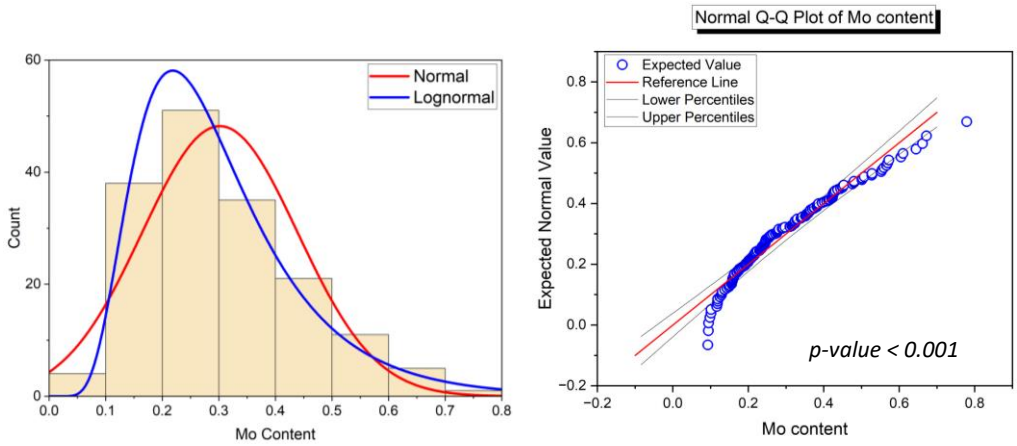


FIGURE 9-2 POPULATION DISTRIBUTION OF MOLYBDENUM CONTENT

Rather, Mo levels of the population are more appropriately modeled by a lognormal distribution (Figure 9-2). Similar distribution trends were observed in the p-hydroxybenzoate (PB) levels in the original GWAS phenotyping effort [262]. While this relationship may be serendipitous, there could also be a potentially unknown link between molybdenum/MOT1 and PB biosynthesis.

As a molybdate transporter, MOT1 is closely tied to molybdenum content [284]. From field sampling efforts, the molybdenum content of xylem, root, and leaf tissue from *Populus* genotypes from the Clatskanie field site has previously been analyzed. For this study, core samples were taken from 14-year-old, clonally replications, naturally variant population of *Populus trichocarpa* from a field site located in Clatskanie, OR. Six genotypes were selected for sampling. Three genotypes were measured to have lower molybdenum content (BESC-81, GW-9776, BESC-13), and three genotypes were measured to have higher molybdenum content (BESC-843, BESC-329, KLNA-20). While expression of MOT1 (either Potri.006G245900 or Potri.006G246000) was not explicitly analyzed here, Mo content has been closely associated with MOT1 [284], and therefore Mo is used here as a proxy for MOT1 expression.
As HSQC NMR was the phenotyping used for the GWAS which generated the MOT1 candidate gene result, these samples were first analyzed by HSQC NMR. The HSQC NMR results of one low Mo level sample and one high Mo level sample are displayed in Figure 9-3. A comparison of these spectra immediately revealed drastic differences in the $PB_{2/6}$ contour (blue ellipse). The low Mo genotype (GW-9776) had an extremely high PB content of 24.4%, whereas the high Mo had PB levels barely above background (0.8%). The PB content and S/G ratio determined by HSQC NMR for all samples is summarized in Figure 9-4 and follows the previously reported inverse relationship between these two lignin traits [60, 288].

Besides HSQC NMR, all samples were subjected to compositional analysis via two-step acid hydrolysis. Results for Klason lignin, glucose, and xylose content are displayed in Figure 9-5. PB and S/G quantities measured by HSQC NMR are also included in Figure 9-5. Each plot also displays Mo levels as determined by ICP (connected black dots). The low Mo genotypes generally had higher PB levels, with all three of these genotypes exhibiting higher PB levels than BESC-329, and two of three genotypes exhibiting higher PB levels than KLNA-20.

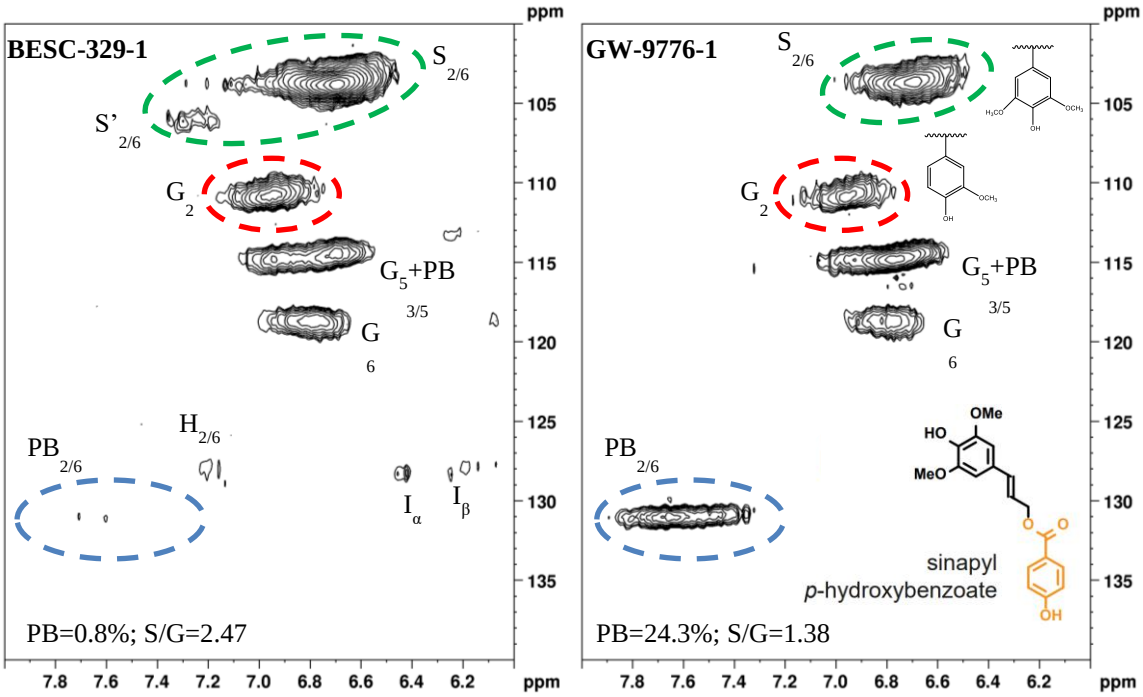


FIGURE 9-3 AROMATIC REGIONS OF HIGH MO GENOTYPE (LEFT) AND LOW MO GENOTYPE (RIGHT)

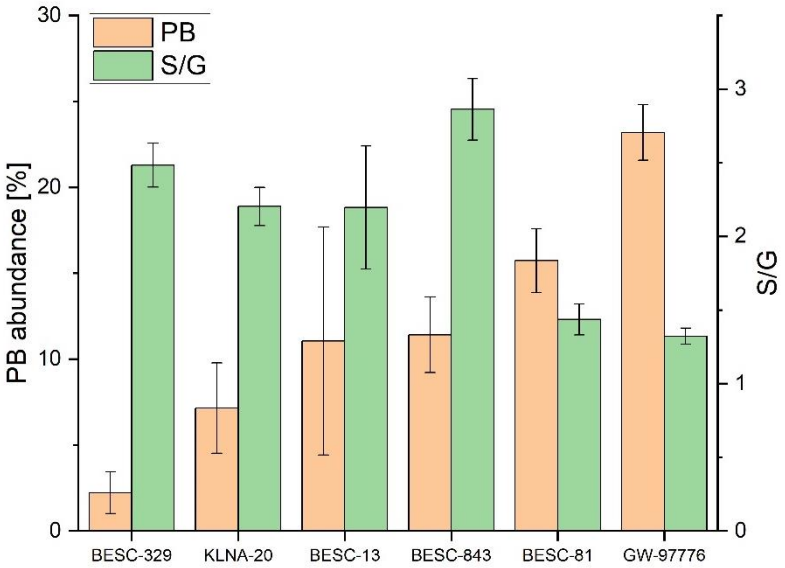
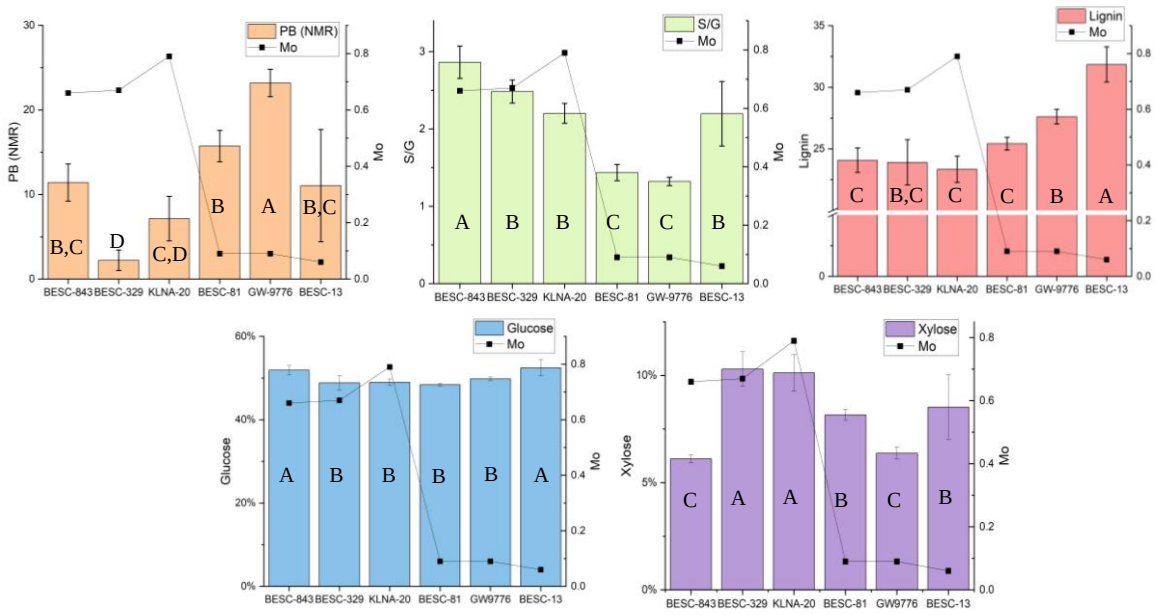


FIGURE 9-4 PB AND S/G LEVELS OF SIX MOT1 GENOTYPES



Connecting letters → ANOVA means comparison (Student's t, α=0.05)

FIGURE 9-5 LEVELS OF PB, S/G RATIO, LIGNIN, GLUCOSE, AND XYLOSE COMPARED TO MOLYBDENUM CONTENT

BESC-13 was found to contain similar PB levels as KLNA-20, though this may be in part to the large standard deviation derived from relatively high variability in the analysis of replicates. A similar trend was observed with Kalson lignin, as low Mo genotype BESC-13 had higher lignin content than all three high Mo genotypes. Additionally, low Mo genotype GW-9776 had higher lignin content than two of three high Mo genotypes.

Glucose levels were mostly consistent across all genotypes, with one low Mo genotype (BESC-13) and one high Mo genotype (BESC-843) having higher glucose levels than the other genotypes. However, differences were observed in xylose content. Two low Mo genotypes (BESC-81, BESC-13) exhibited differential xylose content from two high Mo genotypes (BESC-329, KLNA-20). One each low Mo and high Mo genotypes (GW-9776, BESC-843) had similar lower xylose levels.

While there is not a clear delineation of lignin/cell wall properties between all low Mo and high Mo genotypes, there are several important distinctions to consider. First, the actual expression of MOT1 was not verified for these samples. While analyzed Mo content was used as a proxy for MOT1 expression, the relationships reported here would be more adequately characterized by MOT1 expression rather than Mo content. Additionally, candidate genes were generated from GWAS utilized 3-year-old *Populus trichocarpa* genotypes grown at the Corvallis, OR field site. The samples analyzed here, although are clonal replicates, were 14-years old and grown at the Clatskanie, OR field site. Therefore, environmental conditions and age could be influencing cell wall properties over MOT1 expression. Generating transgenic Populus plants is a long and complex process. Therefore, analyzing genotypes in this manner can provide an initial line of evidence that certain genes of interest will have an effect. These results provide confidence that manipulating levels of MOT1 (as indicated by Mo content) will indeed result in changes to cell wall structure – specifically PB content in lignin.

9.4.2 Characterization genotypes with variable PB levels

The lignin of *Populus* is decorated by *p*-hydroxybenzoate (PB), which is an ester-linked pendent moiety. It is analogous to *p*-coumarate found acylating the lignin of herbaceous species.

Although the subject of numerous conjectures, the biosynthesis and biological function of PB remains to be fully elucidated. The presence of PB is of special interest as it has the potential to positively contribute to the biorefinery as a value-added platform chemical. Indeed, in the past few years there has been a renewed flurry of interest in the PB moiety, including elucidating a previously unknown acyltransferase shown to influence PB production [289, 290]. The presence of PB has also been hypothesized to influence cell wall structure and sugar release [291, 292]. However, this has been difficult to evaluate due to the inherently variable and complex structure of lignin. Additionally, PB is highly correlated with other lignin traits such as the S/G ratio [78, 262]. Due to our previous extensive phenotyping efforts of over 400 unique *Populus trichocarpa* genotypes [262], we have the unique opportunity to strategically select samples that differ primarily by PB content. In addition to HSQC NMR of isolated enzyme lignin, these samples were subjected to whole cell wall HSQC NMR, compositional analysis, quantitative PB, biomass crystallinity, and sugar release analyses to determine the impacts of PB on cell wall architecture. Four genotypes were selected for this analysis: BESC-198, BESC-94, BESC-8, and BESC-65 (Table 9-1). As PB is known to be correlated with S units, samples with a similar S/G ratio were specifically selected. BESC-198 and BESC-94 were determined to have a similar S/G ratio of ~2.0. BESC-8 and BESC-65 had a similar S/G ratio of ~2.8. From the replicate runs of standard *Populus*, the standard deviation of S/G ratio determined by HSQC NMR is 0.05. An effort was made to ensure other lignin phenotypes (i.e., interunit linkages) were similar between samples with a similar S/G ratio. In this way, the primary difference between samples is the PB content. Indeed, the PB content between BESC-198 and BESC-94 was ~10.3%, and the difference between BESC-8 and BESC-65 is ~10.7%. The standard deviation of the PB measurement is 0.37%, indicating that differences >10% represent authentic differences in PB content.

The primary criticism of HSQC NMR for lignin characterization is that it is not fully quantitative. This is especially applicable to the PB moiety. As PB exists exclusively as a pendant unit, it inherently has a longer relaxation time and therefore tends to be overestimated by HSQC NMR. Therefore, samples were subjected to mild alkaline hydrolysis to cleave the ester linkages to enable quantitative PB analysis by HPLC. As indicated in Figure 9-6, PB was eluted at approximately 14.25 minutes.

TABLE 9-1 NMR RESULTS OF LIGNIN STRUCTURE OF SELECTED GENOTYPES

Structure	BESC-198	BESC-94	BESC-8	BESC-65	Sample StDev
PB	4.24	14.5	2.90	13.6	0.37
S/G	2.05	1.92	2.80	2.77	0.05
H	0.66	0.61	0.94	1.42	0.05
β-O-4	60.5	59.1	63.2	62.6	1.34
β-5	3.32	3.41	2.68	1.87	0.71
β-β	8.06	7.45	1.11	6.15	0.59

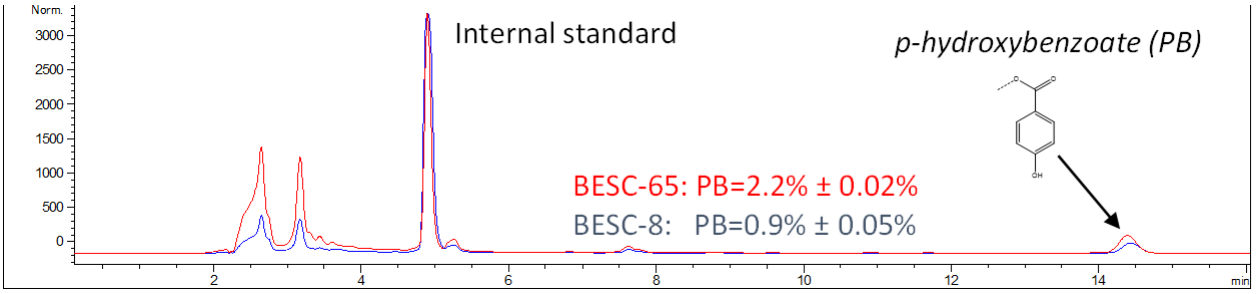


FIGURE 9-6 HPLC CHROMATOGRAM OF *P*-HYDROXYBENZOATE FROM ALKALINE HYDROLYSIS OF BIOMASS

The PB content determined by HPLC was normalized by the lignin content, such that PB results are comparable to the results reported by HSQC NMR. BESC-65 was found to contain 2.2% ± 0.02% PB, whereas BESC-8 was found to contain 0.9% ± 0.05% PB. Similarly, BESC-198 was measured to contain 0.8% ± 0.1% PB, whereas BESC-94 was measured to contain 1.3% ± 0.1% PB. A comparison of NMR and HPLC measurements of PB abundance is provided in Figure 9-7. These HPLC results validate that the paired samples do have differential levels of PB, though the results may not be as drastic as indicated by HSQC NMR. This is consistent with our previous findings, which demonstrated an excellent correlation between HSQC NMR and HPLC measurement of PB content [293].

Compositional analysis was performed on all samples to evaluate potential differences in cellulose, hemicellulose, and lignin content, with results displayed in Figure 9-8. As previously mentioned, it has been postulated that PB levels may have an impact on cell wall structure – such as the abundance of cell wall biopolymers. The amount polysaccharides in the cell wall is also important for accurately performing the sugar release experiment. Additionally, the lignin content of each sample to normalize the PB measured by HPLC. Samples BESC-198 and BESC-94 (S/G ~2.0) were measured to have lignin contents of 23.7% ± 1.6% and 29.7% ± 0.9%, respectively. These levels of lignin content were determined to be significantly different by a pairwise Student's t-test (α=0.05). BESC-198 and BESC-94 were similarly determined to have differential levels of glucose and xylose (Student's t-test, α=0.05). In contrast, BESC-8 and BESC-65 (S/G ~2.8) were measured to have lignin contents of 26.8% ± 2.0% and 24.1% ± 3.4%, respectively. A pairwise Student's t-test (α=0.05) concluded that the differences in these lignin contents were not statistically significant. Similarly, differences in xylose levels were determined to not be statistically significant (α=0.05). However, BESC-65 was found to have differential glucose content compared BESC-8 (α=0.05). However, while BESC-65 (S/G≈2.8; high PB) was measured to have higher glucose content compared to BESC-8 (S/G≈2.8; low PB), BESC-94 (S/G≈2.0; high PB) had lower glucose compared to BESC-198 (S/G≈2.0; low PB).

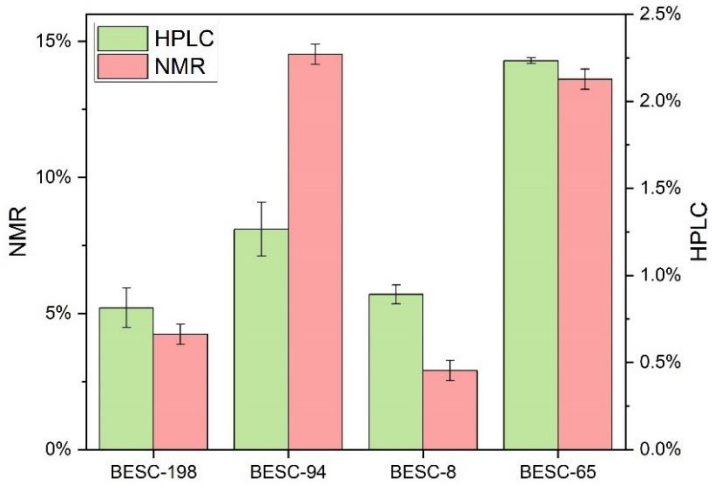


FIGURE 9-7 COMPARISON OF P-HYDROXYBENZOATE LEVELS DETERMINED BY HPLC AND NMR

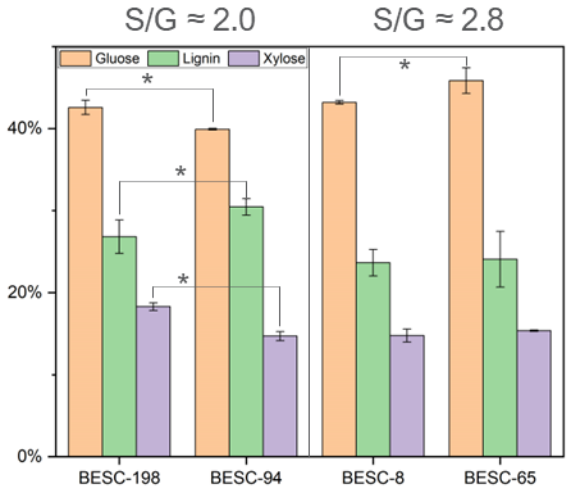


FIGURE 9-8 COMPOSITIONAL ANALYSIS RESULTS FOR LIGNIN, GLUCOSE, AND XYLOSE CONTENT

Therefore, while some differences in cell wall composition were detected between these paired samples, they are inconsistent with PB levels. These results also illustrate the variability and complexity in cell wall phenotypes that exist, even between samples that may appear similar. The measured glucose content from this compositional analysis will be used to inform the enzyme loading for the sugar release experiment such that results can be reported on a normalized g/g_{glucan} basis.

Besides cellulose, hemicellulose, and lignin content, there are also several additional factors that can impact enzymatic digestion of cell wall polysaccharides. “cellulose crystallinity”, but whole cell wall biomass was examined here. Since cellulose (and xylose and lignin) were found to be similar among all samples, this “biomass crystallinity” is argued to be a comparable representative quantity to use for comparison here. Cellulose chains in microfibrils exhibit inter- and intra- chain hydrogen bonding, giving rise to crystalline lattice regions. The regions of cellulose that remain amorphous offer easier access to enzymes which in turn leads to more efficient digestion. Therefore, the ratio of amorphous to crystalline regions of cellulose is often a measurement of great interest. There are several methods for quantifying crystallinity, such as XRD, NMR, IR, Raman, and SFG [294].

One of these factors is crystallinity. Note that crystallinity is often discussed in terms of Here, crystallinity was measured by ^{13}C CP/MAS NMR. This one-dimensional NMR technique gives rise to several cellulose related peaks in the region of 50-110 ppm (Figure 9-9, left). The C_4 carbon of cellulose is present in the 80-90 ppm region. Specifically, the C_4 carbon exhibits two peaks, with the peak around 89 ppm representing the crystalline fraction and the peak around 84 ppm representing the amorphous fraction. The crystallinity index (CrI) is calculated by dividing the crystalline C_4 peak by the sum of the crystalline and amorphous C_4 peaks. The CrI of BESC-94, BESC-198, BESC-65, and BESC-8 were measured to be 37.6%, 35.8%, 39.8%, and 38.4%, respectively (Figure 9-9, right). The standard deviation of the CrI measurement was ~1.0%. A one-way ANOVA analysis determined that the differences in CrI were not statistically significant across these samples (α=0.05). These results suggest that (i) PB levels are not

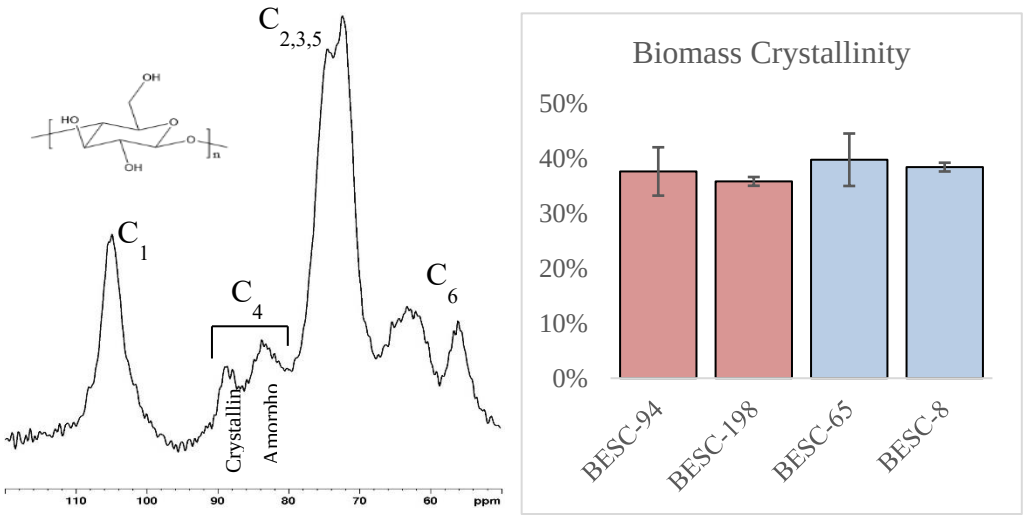


FIGURE 9-9 NMR SPECTRA (LEFT) AND BIOMASS CRYSTALLINITY (RIGHT)

associated with CrI, and (ii) potential differences in sugar release will not be due to variable levels of CrI.

An important qualification in the selection of these samples is they were chosen based on the initial analysis of isolated enzyme lignin (after cellulase treatment to remove polysaccharides). One of the ways PB may impact cell wall architecture is though the presence of different sugars or linkages in the lignin-carbohydrate complex (LCC). However, this information is lost along with the polysaccharides when enzymatic hydrolysis is performed. Therefore, whole cell wall (WCW) HSQC NMR was performed to investigate the LCC region for difference between low and high PB genotypes. A WCW NMR comparison between low PB and high PB genotypes revealed that signals in both samples were similar (Figure 9-10). This is yet additional evidence that variable levels of PB are not associated with differences in cell wall structure. Also, the levels of S, G, H, PB, etc. were found to be similar between the WCW and isolated enzyme lignin samples, indicating there is no appreciable contamination by cell wall amino acids or enzymes used for hydrolysis, as has been previous reported [117].
It has been hypothesized that PB may facilitate sugar release [10, 295, 296]. There have been attempts to evaluate the impact on PB on sugar release. However, attempts to manipulate PB content (through genetic engineering) often results in changes to other cell wall components. For instance, introducing a bacterial chorismate pyruvate lyase gene into poplar increased PB levels, but also induced changes to lignin content and S/G ratio [77]. It then becomes convoluted which factor(s) contributed to the improved sugar release. This illustrates the unique opportunity available here by strategically selected paired samples which have differential PB content, but otherwise similar cell wall properties.

Sugar release experiments were performed with the enzyme blend CTec2, which contains cellulases, β-glucosidases, and hemicellulase, was added at a loading of 24 filter paper units per gram glucan. Aliquots were taken at 4h, 24h, 48h, and 72h, and results are displayed in Figures 9-11 and 9-12. Samples BESC-94 and BESC-198 (S/G ~2.0) exhibited similar trends in glucose and xylose release, with the high PB genotype (BESC-94) releasing slightly more of each sugar. BESC-9 and BESC-65 (S/G ~2.8) exhibited higher sugar release than the S/G ~2.0 samples,

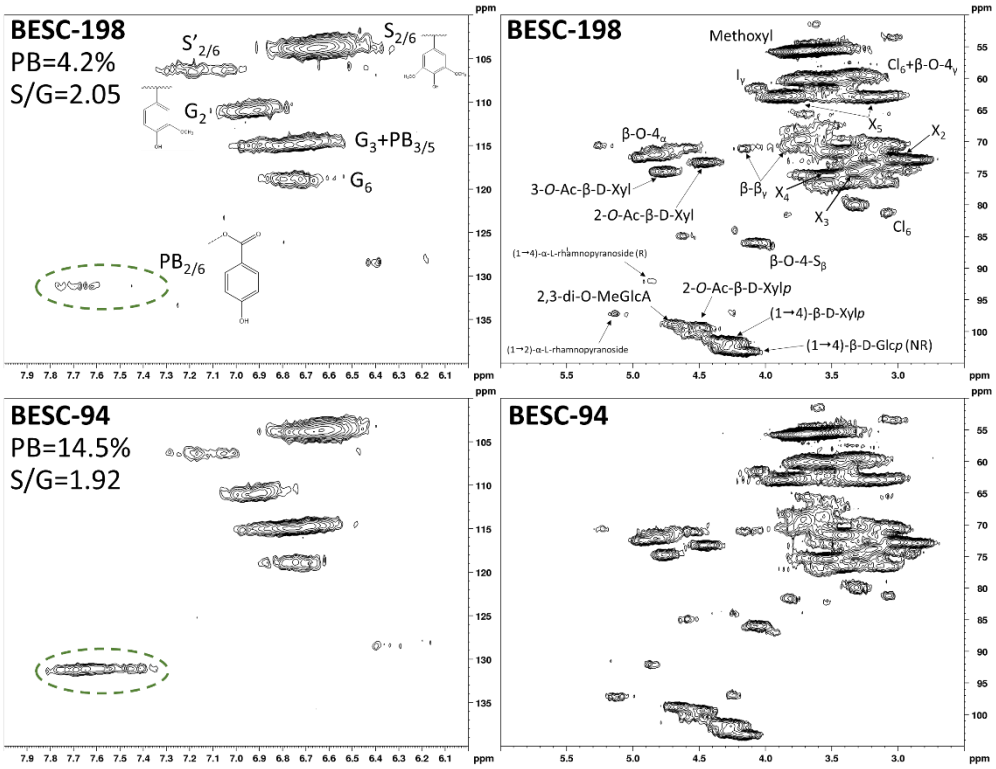


FIGURE 9-10 HSQC NMR SPECTRA OF ANOMERIC AND LINKAGE REGIONS OF LOW AND HIGH PB GENOTYPES

which is expected as a higher S/G ratio is associated with increased sugar release. However, as indicated in Figures 9-11 and 9-12, the low PB genotype (BESC-8) exhibited much higher glucose and xylose release than its high PB (BESC-65) counterpart. If either the low PB or high PB genotypes had exhibited higher sugar release across both pairs of samples, this would have been evidence that PB levels impact sugar release. However, sugar release appears not to be impacted by difference in PB content. This difference in sugar release could potentially be influenced by differences in other lignin phenotypes, despite efforts made to control these for similar levels. For instance, BESC-8, which exhibited the highest sugar release (glucose and xylose), had fewer β-β linkage (1.11%) compared to all other genotypes (6.15%-8.06%). As condensed linkages are known to impact sugar release, this could be a more influential factor than PB in sugar accessibility/release.

This work sought to explore the hypothesized relationship between the PB lignin moiety and various cell wall properties. Samples were strategically selected from a population of >400 naturally variant *Populus trichocarpa* such that the primary difference between samples were PB levels. Differences in PB content were determined quantitatively through mild alkaline hydrolysis and HPLC, confirming the initial HSQC NMR results. Compositional analysis determined that all samples contained similar levels of glucose, xylose, and lignin. Biomass crystallinity as determined by ^{13}C CP/MAS NMR was likewise measured to be similar among all samples. Whole cell wall NMR also indicated that there were no significant differences in the LCC region between low PB and high PB genotypes. Finally, sugar release was found to not be correlated with PB levels. Taken together, these results provide evidence that PB does not impact cell wall architecture or sugar release. This suggests that PB levels could be increased in *Populus* (i.e., through genetic engineering) for biorefinery purposes without detrimental effects to the cell wall. Due to the large poplar and switchgrass populations characterized by our previous efforts, similar experiments could be conducted for other lignin moieties (i.e., p-coumarate, ferulate, tricin, etc.) while controlling for numerous variables.

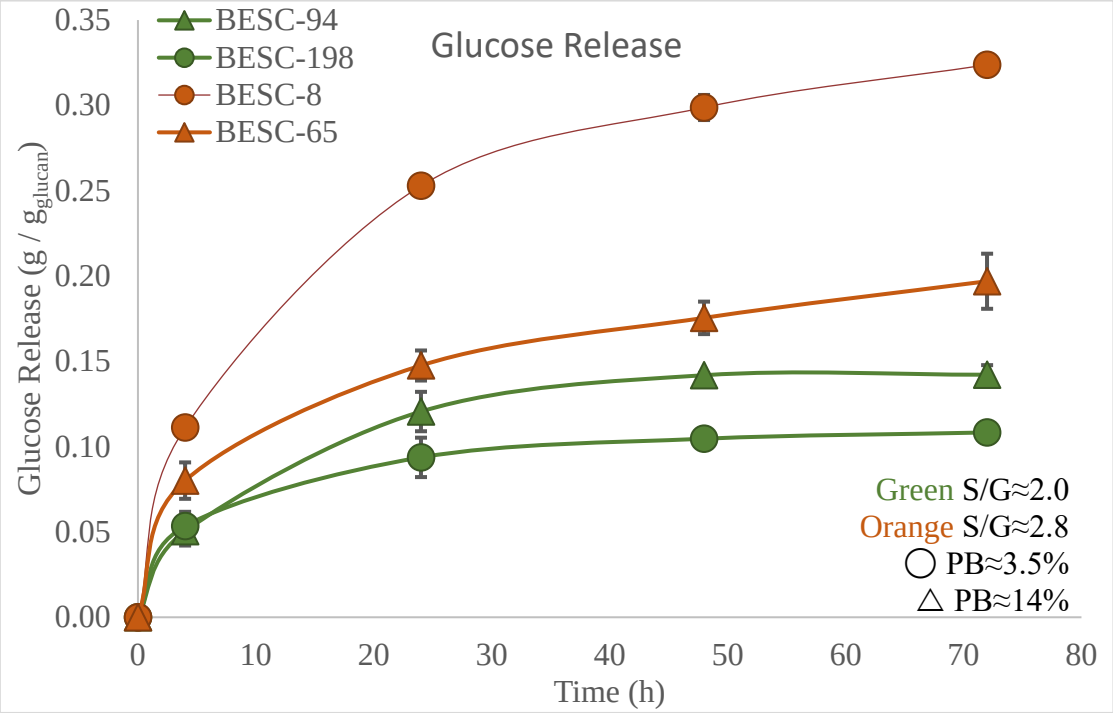


FIGURE 9-11 GLUCOSE RELEASE OF HIGH AND LOW PB GENOTYPES

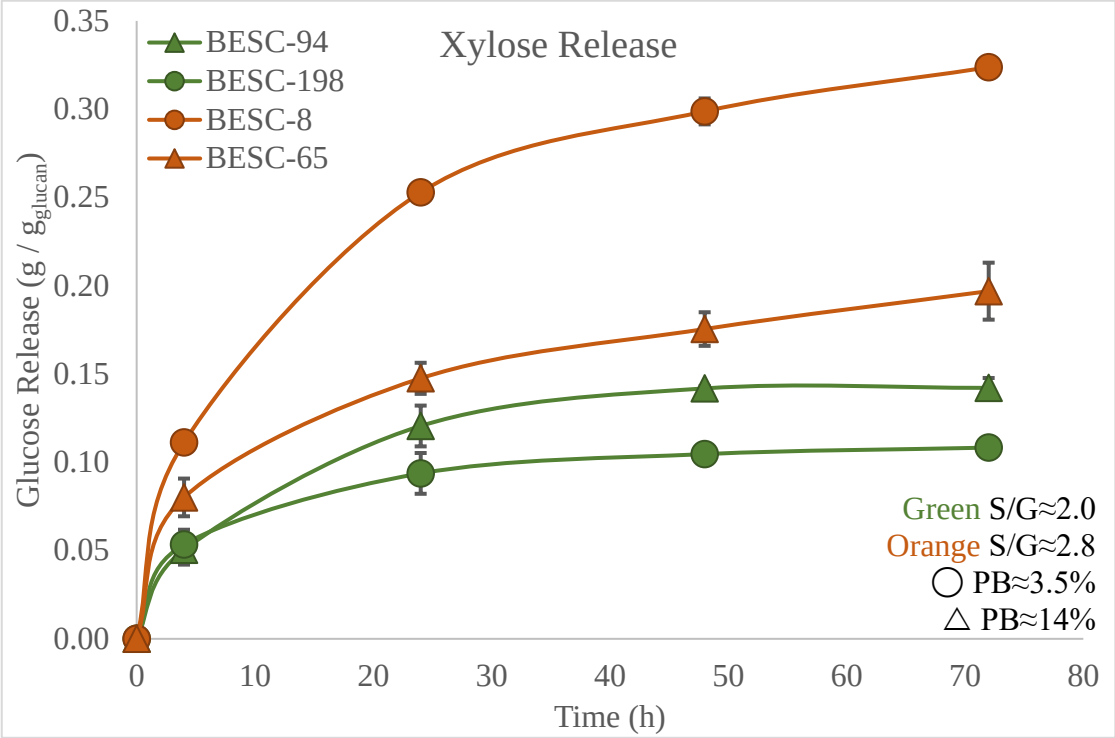


FIGURE 9-12 XYLOSE RELEASE OF HIGH AND LOW PB GENOTYPES

Chapter 10:
Conclusions and future work

10.1 Conclusions

The secondary cell wall, and especially lignin, plays a vital role in plant development and conversion to biofuels. Characterization of lignin is crucial to understanding its impact in these endeavors. This dissertation extensively characterized the lignin to examine the potential contamination effects of pretreatment chemicals, explore differences in plants impacted by drought and fungal infection, and to elucidate novel genes that may influence lignin biosynthesis.

It is well documented that impurities and degradation products can interfere with the analysis of biomass. An NMR evaluation of twenty-nine compounds commonly used in biomass pretreatment revealed that there were identifiable chemical shifts with some signals, such as those associated with the ionic liquid 1-butyl-3-methylimidazolium acetate. There was also a clear overlap in the methoxy signal between biomass and ethanol. Both observations could lead to potential misinterpretation of these signals in the NMR spectra, compromising the analysis of lignin.

It has been shown that biomass yield is the most important driver in determining the financial viability of the biorefinery. Biomass yield can be negatively affected by abiotic stress such as drought. The leaves from irrigated and drought treated poplar trees were examined to determine the impact of drought on lignin structure. However, tremendous variability was observed both in lignin and metabolites that was not associated with drought treatment, providing unique insights on this less well studied leaf tissue. Another set of poplar leaves currently under investigation by the Ragauskas group is also showing similar levels of variability. Biomass yield can also be affected by biotic stress such as fungal infection. The fungus *Septoria* is of particular concern due to its ability to kill an entire tree and spread quickly through a population. A study on poplar infected by the *Septoria* fungus revealed that infected regions tended to have a higher S/G ratio and lower PB content than the corresponding healthy region. This could suggest that these phenotypes – or genes in their biosynthesis pathway – may play a role in stress response. This is especially interesting for PB, where the biosynthesis pathway has yet to be fully elucidated.

The formation of the secondary cell wall requires the coordination of many metabolic pathways. Additionally, complex traits such as lignin are highly polygenic. Therefore, there are likely many

genes that are unknowingly influencing lignin phenotypes. Deep phenotyping was performed on natural variant populations of poplar and switchgrass to quantify several lignin traits. These lignin phenotypes were incorporated into a genome wide association study to identify associations with causal alleles. These associations provide new genetic targets for future work on lignin structure, and functional characterization of these previously unknown genes could reveal novel genetic mechanisms controlling lignin biosynthesis. These results can be used to improve these bioenergy crops for utilization toward biofuel and bio-based products.

To illustrate the utility of these results, natural variant *Populus* genotypes were selected to examine the effect of GWAS identified candidate gene MOT1. Genotypes with differential levels of molybdenum (a proxy for MOT1 expression), where shown to have differential levels of *p*-hydroxybenzoate (PB) and S/G ratio, suggesting that MOT1 may play a role in lignin biosynthesis. Due to the recent interest in the PB moiety, well characterized *Populus* genotypes were selected such that lignin traits were similar – except for PB. Extensive characterization of these high/low PB samples revealed no significant differences in lignin/cell wall structure, and PB was not correlated with polysaccharide accessibility.

These results contribute to the body of knowledge of how lignin biosynthesis competes for carbon flux and provide a basis for elucidating novel genetic mechanisms underlying lignin biosynthesis.

10.2 Future work

In many regards, these results – especially the GWAS results – can be viewed as a steppingstone, fueling future research for years. The logical next step is the functional characterization of genes identified in the GWAS results, with the goal of being able to tune lignin for applications such as lignin-first or lignin-last processing. As lignin is involved in many biological processes, genes identified by GWAS may also lead to improvements in biomass yield, biotic and/or abiotic stress resistance. Synergistic effects may even be realized by targeting double or triple mutant lines. These efforts are high risk/high reward, as they will require extensive teamwork with geneticists and biologists, in addition to the phenotyping capabilities demonstrated here. However, successful functional characterization of some of these GWAS identified candidate genes may

fundamentally change the current understanding of the genetic mechanisms underlying lignin biosynthesis.

In a more practical sense, there is a significant amount of information that can be obtained from the currently characterized set of poplar and switchgrass. In the poplar GWAS work, 12 phenotypes were quantified from 409 genotypes (>4,900 elements). Similarly, the switchgrass GWAS work quantified 19 phenotypes from 312 genotypes (>5,900 elements). These results were obtained from HSQC NMR on isolated lignin, which is a time and labor-intensive process. However, it has the tremendous benefit of elucidating many (i.e., 12-19) lignin traits simultaneously, the power of which cannot be matched by any other analytical method. It would be of enormous benefit to the field if there existed a method that had the benefits of being high throughput and high power. This opportunity now exists. Analytical techniques such as FTIR and NIR are high throughput but have the drawback of convoluted spectra. By combining FTIR/NIR spectra of lignin with known HSQC NMR results, a model could be developed for high throughput analysis of poplar and switchgrass biomass. Such a model could save weeks or months of experimental time, enabling high resolution results with a high throughput technique.

The GWAS populations could also fuel additional fundamental work such as elucidating relationships among additional cell wall properties. As stated numerous times in this dissertation, lignin (and the entire secondary cell wall) is a very complex structure. With the lignin monolignols and interunit linkages now well characterized, additional characterization of lignin or cell wall properties could be conducted to explore relationships with lignin structure. These include:

- Compositional analysis could be performed to evaluate the relationship of lignin structure to lignin, cellulose, and hemicellulose content. Alternatively, the newly reported cysteine assisted acetic acid procedure could be utilized as a high throughput method to evaluate lignin content.
- ^{13}C CP/MAS solid state NMR could be used to characterize crystallinity. This could be performed on native biomass (biomass crystallinity) or isolated cellulose (cellulose crystallinity).

- ^{31}P NMR could be performed to elucidate aromatic and aliphatic hydroxyl groups of lignin to test for influence from lignin structure.
- Alkaline hydrolysis could be conducted to quantify ester linked groups of lignin. As these pendent groups are often overestimated by HSQC NMR, this would provide a quantifiable validation of this measurement.
- Select genotypes could be evaluated for potentially valorizing properties. For instance, lignin is an excellent radical scavenger, and insight could be gained by measuring antioxidant properties of variable lignin structures. Additionally, various lignin structures could be tested for thermal properties by DSC/TGA for evaluation in potential uses such as 3D printing.

It should be noted that although some of these relationships have been explored, they are often reported with a limited sample size (typically <10 samples). Additionally, samples may under under-characterized and/or not representative of the natural variability of the population. There is also the opportunity to explore these relationships in both poplar and switchgrass populations, which offer a drastically differential lignin biopolymer.

While characterizing lignin is essential fundamental work, the ultimate end goal is to valorize lignin to improve the economic viability of the biorefinery. Indeed, lignin has been explored as applications such as resins [297], battery electrodes/binders [298], fuels [299], and healthcare products [300], among many others. In this area of research, there are active efforts to produce and utilize lignin nanoparticles. One potential application of lignin nanoparticles is as a UV absorber in sunscreen. Lignin is known to be a good UV absorber. However, there are numerous properties of lignin that influence its antioxidant activity, such as phenolic hydroxyl content, molecular weight and polydispersity [301]. Elucidating structure-property relationships to enhance lignin UV absorbance could enable lignin to replace current chemical UV absorbers. Chemical UV absorbers have been associated with coral reef degradation [302, 303], and other compounds in sunscreen that may be absorbed through the skin have been shown to act as hormone disrupters [304]. Conversely, lignin has no inherent environmental proliferation problems, and thus is more eco-friendly. Therefore, utilizing lignin in this application represents

a double-bonus benefit. Along these lines, lignin nanoparticles could also be implemented in eco-friendly applications such as food preservatives and nanopesticides.

References:

1. Sarwer, A., et al., *Suitability of Biofuels Production on Commercial Scale from Various Feedstocks: A Critical Review.* ChemBioEng Reviews, 2022. **9**(5): p. 423-441.
2. Holechek, J.L., et al., *A global assessment: can renewable energy replace fossil fuels by 2050?* 2022. **14**(8): p. 4792.
3. Jeswani, H.K., A. Chilvers, and A.J.P.o.t.R.S.A. Azapagic, *Environmental sustainability of biofuels: a review.* 2020. **476**(2243): p. 20200351.
4. Jeswani, H.K., A. Chilvers, and A. Azapagic, *Environmental sustainability of biofuels: a review.* J Proceedings of the Royal Society, 2020. **476**(2243): p. 20200351.
5. Reid, W.V., M.K. Ali, and C.B.J.G.c.b. Field, *The future of bioenergy.* 2020. **26**(1): p. 274-286.
6. Khan, N., K. Sudhakar, and R.J.S. Mamat, *Role of biofuels in energy transition, green economy and carbon neutrality.* 2021. **13**(22): p. 12374.
7. Ragauskas, A.J., et al., *The path forward for biofuels and biomaterials.* 2006. **311**(5760): p. 484-489.
8. Energy, U.D.o., *Biorefinery Optimization Workshop Summary Report*. 2016, Office of Energy Efficiency and Renewable Energy Bioenergy Technologies ….
9. Li, M., Y. Pu, and A.J.J.F.i.c. Ragauskas, *Current understanding of the correlation of lignin structure with biomass recalcitrance.* 2016. **4**: p. 45.
10. Studer, M.H., et al., *Lignin content in natural Populus variants affects sugar release.* 2011. **108**(15): p. 6300-6305.
11. Abu-Omar, M.M., et al., *Guidelines for performing lignin-first biorefining.* 2021. **14**(1): p. 262-292.
12. Williams, C.L., et al., *Sources of biomass feedstock variability and the potential impact on biofuels production.* 2016. **9**: p. 1-14.
13. Gundekari, S., J. Mitra, and M. Varkolu, *Classification, characterization, and properties of edible and non-edible biomass feedstocks*, in *Advanced Functional Solid Catalysts for Biomass Valorization*. 2020, Elsevier. p. 89-120.
14. Hon, D.N.S., *Cellulose: a random walk along its historical path.* Cellulose, 1994. **1**(1): p. 1-25.
15. Heinze, T.J.C.c., n. properties: fibers, and a. materials, *Cellulose: structure and properties.* 2016: p. 1-52.
16. Hallac, B.B., A.J.J.B. Ragauskas, Bioproducts, and Biorefining, *Analyzing cellulose degree of polymerization and its relevancy to cellulosic ethanol.* 2011. **5**(2): p. 215-225.
17. McNamara, J.T., J.L. Morgan, and J.J.A.r.o.b. Zimmer, *A molecular description of cellulose biosynthesis.* 2015. **84**: p. 895-921.
18. Park, S., et al., *Cellulose crystallinity index: measurement techniques and their impact on interpreting cellulase performance.* 2010. **3**: p. 1-10.
19. Himmel, M.E., et al., *Biomass recalcitrance: engineering plants and enzymes for biofuels production.* 2007. **315**(5813): p. 804-807.
20. Gírio, F.M., et al., *Hemicelluloses for fuel ethanol: a review.* 2010. **101**(13): p. 4775-4800.
21. Scheller, H.V. and P.J.A.r.o.p.b. Ulvskov, *Hemicelluloses.* 2010. **61**: p. 263-289.

22. Rennie, E.A. and H.V.J.C.o.i.b. Scheller, *Xylan biosynthesis.* 2014. **26**: p. 100-107.
23. Kumar, M., L. Campbell, and S.J.J.o.e.b. Turner, *Secondary cell walls: biosynthesis and manipulation.* 2016. **67**(2): p. 515-531.
24. Ebringerová, A., et al., *Hemicellulose.* 2005: p. 1-67.
25. Pauly, M., et al., *Hemicellulose biosynthesis.* 2013. **238**: p. 627-642.
26. McCann, M.C. and N.C.J.J.o.E.B. Carpita, *Biomass recalcitrance: a multi-scale, multi-factor, and conversion-specific property.* 2015. **66**(14): p. 4109-4118.
27. Luo, Y., et al., *The production of furfural directly from hemicellulose in lignocellulosic biomass: A review.* 2019. **319**: p. 14-24.
28. Boerjan, W., J. Ralph, and M.J.A.r.o.p.b. Baucher, *Lignin biosynthesis.* 2003. **54**(1): p. 519-546.
29. Vanholme, R., et al., *Lignin biosynthesis and structure.* 2010. **153**(3): p. 895-905.
30. Herrmann, K.M.J.P.p., *The shikimate pathway as an entry to aromatic secondary metabolism.* 1995. **107**(1): p. 7.
31. Boerjan, W., J. Ralph, and M. Baucher, *Lignin biosynthesis.* J Annual review of plant biology, 2003. **54**(1): p. 519-546.
32. Dixon, R.A. and J.J.O.B. Barros, *Lignin biosynthesis: old roads revisited and new roads explored.* 2019. **9**(12): p. 190215.
33. Vanholme, R., et al., *Lignin biosynthesis and structure.* J Plant physiology, 2010. **153**(3): p. 895-905.
34. Wang, J.P., et al., *Flux modeling for monolignol biosynthesis.* 2019. **56**: p. 187-192.
35. Zhang, J., et al., *Recent advances in the transcriptional regulation of secondary cell wall biosynthesis in the woody plants.* 2018. **9**: p. 1535.
36. Hu, W.-J., et al., *Repression of lignin biosynthesis promotes cellulose accumulation and growth in transgenic trees.* 1999. **17**(8): p. 808-812.
37. Bryant, N.D., et al., *Transgenic poplar designed for biofuels.* 2020. **25**(9): p. 881-896.
38. Tolbert, A., et al., *Characterization and analysis of the molecular weight of lignin for biorefining studies.* 2014. **8**(6): p. 836-856.
39. Wang, H., et al., *Lignin modification improves the biofuel production potential in transgenic Populus tomentosa.* 2012. **37**(1): p. 170-177.
40. Voelker, S.L., et al., *Antisense down-regulation of 4CL expression alters lignification, tree growth, and saccharification potential of field-grown poplar.* 2010. **154**(2): p. 874-886.
41. Xiang, Z., et al., *Field-grown transgenic hybrid poplar with modified lignin biosynthesis to improve enzymatic saccharification efficiency.* 2017. **5**(3): p. 2407-2414.
42. Min, D., et al., *The elucidation of the lignin structure effect on the cellulase-mediated saccharification by genetic engineering poplars (Populus nigra L.× Populus maximowiczii A.).* 2013. **58**: p. 52-57.
43. Min, D., et al., *The cellulase-mediated saccharification on wood derived from transgenic low-lignin lines of black cottonwood (Populus trichocarpa).* 2012. **168**: p. 947-955.
44. Xiang, Z., et al., *Wood characteristics and enzymatic saccharification efficiency of field-grown transgenic black cottonwood with altered lignin content and structure.* 2015. **22**: p. 683-693.

45. Ralph, J., et al., *Effects on lignin structure of coumarate 3-hydroxylase downregulation in poplar.* 2012. **5**: p. 1009-1019.
46. Zhou, X., et al., *Preliminary study of cell wall structure and its mechanical properties of C3H and HCT RNAi transgenic poplar sapling.* 2018. **8**(1): p. 10508.
47. Miller, Z.D., et al., *Altered lignin content and composition in transgenic Populus trichocarpa results in a decrease of modulus of elasticity.* 2018. **13**(4): p. 7698-7708.
48. Van Acker, R., et al., *Different routes for conifer-and sinapaldehyde and higher saccharification upon deficiency in the dehydrogenase CAD1.* 2017. **175**(3): p. 1018-1039.
49. Özparpucu, M., et al., *The effect of altered lignin composition on mechanical properties of CINNAMYL ALCOHOL DEHYDROGENASE (CAD) deficient poplars.* 2018. **247**: p. 887-897.
50. Saleme, M.d.L.S., et al., *Silencing CAFFEOYL SHIKIMATE ESTERASE affects lignification and improves saccharification in poplar.* 2017. **175**(3): p. 1040-1057.
51. Van Acker, R., et al., *Improved saccharification and ethanol yield from field-grown transgenic poplar deficient in cinnamoyl-CoA reductase.* 2014. **111**(2): p. 845-850.
52. Yoo, C.G., et al., *Significance of lignin S/G ratio in biomass recalcitrance of Populus trichocarpa variants for bioethanol production.* ACS Sustainable Chemistry Engineering, 2018. **6**(2): p. 2162-2168.
53. Zhou, S., et al., *Chemical Pulping Advantages of Zip-lignin Hybrid Poplar.* 2017. **10**(18): p. 3565-3573.
54. de Vries, L., et al., *p HBMT1, a BAHD-family monolignol acyltransferase, mediates lignin acylation in poplar.* J Plant Physiology, 2022. **188**(2): p. 1014-1027.
55. Zhao, Y., et al., *Monolignol acyltransferase for lignin p-hydroxybenzoylation in Populus.* J Nature Plants, 2021. **7**(9): p. 1288-1300.
56. Ralph, J., *Hydroxycinnamates in lignification.* J Phytochemistry Reviews, 2010. **9**: p. 65-83.
57. Sannigrahi, P., et al., *Poplar as a feedstock for biofuels: a review of compositional characteristics.* Biofuels, Bioproducts, & Biorefining, 2010. **4**(2): p. 209-226.
58. David, K. and A.J. Ragauskas, *Switchgrass as an energy crop for biofuel production: a review of its ligno-cellulosic chemical properties.* ACS Sustainable Chemistry, 2010. **3**(9): p. 1182-1190.
59. Yan, J., et al., *Chemical compositions of four switchgrass populations.* Biomass & Bioenergy

2010. **34**(1): p. 48-53.
60. Bryant, N., et al., *Novel candidate genes for lignin structure identified through genome-wide association study of naturally varying Populus trichocarpa.* J Frontiers in Plant Science, 2023. **14**.
61. Ralph, J., C. Lapierre, and W. Boerjan, *Lignin structure and its engineering.* Current opinion in biotechnology, 2019. **56**: p. 240-249.
62. Samuel, R., et al., *Structural changes in switchgrass lignin and hemicelluloses during pretreatments by NMR analysis.* Polymer Degradation Stability, 2011. **96**(11): p. 2002-2009.

63. Smith, D.C., *p-Hydroxybenzoate groups in the lignin of aspen (Populus tremula).* Journal of the Chemical Society, 1955: p. 2347-2351.
64. Cesarino, I., et al., *Building the wall: recent advances in understanding lignin metabolism in grasses.* Acta Physiologiae Plantarum, 2016. **38**: p. 1-14.
65. Tolbert, A., et al., *Characterization and analysis of the molecular weight of lignin for biorefining studies.* Biofuels, Bioproducts, & Biorefining, 2014. **8**(6): p. 836-856.
66. Henriksson, G., et al., *Lignin-carbohydrate network in wood and pulps: a determinant for reactivity.* Biotechnology for biofuels, 2007.
67. Zhao, Y., et al., *Lignin-carbohydrate complexes (LCCs) and its role in biorefinery.* 2020. **253**: p. 120076.
68. Tarasov, D., M. Leitch, and P. Fatehi, *Lignin–carbohydrate complexes: properties, applications, analyses, and methods of extraction: a review.* Biotechnology for biofuels, 2018. **11**(1): p. 1-28.
69. Fry, S.C., *Plant cell walls. From chemistry to biology*. 2011, Oxford University Press.
70. Takahashi, N., T.J.W.S. Koshijima, and Technology, *Ester linkages between lignin and glucuronoxylan in a lignin-carbohydrate complex from beech (Fagus crenata) wood.* 1988. **22**: p. 231-241.
71. Li, M., Y. Pu, and A.J. Ragauskas, *Current understanding of the correlation of lignin structure with biomass recalcitrance.* J Frontiers in chemistry, 2016. **4**: p. 45.
72. Samuel, R., et al., *Investigation of the fate of poplar lignin during autohydrolysis pretreatment to understand the biomass recalcitrance.* 2013. **3**(16): p. 5305-5309.
73. Samuel, R., et al., *Structural characterization and comparison of switchgrass ball-milled lignin before and after dilute acid pretreatment.* Applied Biochemistry & Biotechnology, 2010. **162**: p. 62-74.
74. Yoo, C.G., et al., *Significance of lignin S/G ratio in biomass recalcitrance of Populus trichocarpa variants for bioethanol production.* 2018. **6**(2): p. 2162-2168.
75. Mansfield, S.D., K.Y. Kang, and C.J.N.P. Chapple, *Designed for deconstruction–poplar trees altered in cell wall lignification improve the efficacy of bioethanol production.* 2012. **194**(1): p. 91-101.
76. Taboada, A., et al., *Digestibility of silages in relation to their hydroxycinnamic acid content and lignin composition.* 2010. **90**(7): p. 1155-1162.
77. Mottiar, Y., et al., *Metabolic engineering of p-hydroxybenzoate in poplar lignin.* 2023. **21**(1): p. 176-188.
78. Mottiar, Y. and S.D.J.F.i.P.S. Mansfield, *Lignin p-hydroxybenzoylation is negatively correlated with syringyl units in poplar.* 2022. **13**.
79. Matsushita, Y.J.J.o.W.S., *Conversion of technical lignins to functional materials with retained polymeric properties.* 2015. **61**(3): p. 230-250.
80. Sannigrahi, P., et al., *Fundamentals of biomass pretreatment by fractionation.* 2013: p. 201-222.
81. Sun, R.C.J.C., *Lignin source and structural characterization.* 2020. **13**(17): p. 4385-4393.
82. Zhang, C.J.A.f.-c.t. and f. prospect, *Lignocellulosic ethanol: technology and economics.* 2019.

83. Ragauskas, A.J., et al., *Lignin valorization: improving lignin processing in the biorefinery.* 2014. **344**(6185): p. 1246843.
84. Yu, O. and K.H.J.A.S. Kim, *Lignin to materials: A focused review on recent novel lignin applications.* 2020. **10**(13): p. 4626.
85. Nguyen, N.A., et al., *A path for lignin valorization via additive manufacturing of high-performance sustainable composites with enhanced 3D printability.* 2018. **4**(12): p. eaat4967.
86. Obielodan, J., et al. *Characterization of PLA/lignin biocomposites for 3D printing*. in *2019 International Solid Freeform Fabrication Symposium*. 2019. University of Texas at Austin.
87. Nguyen, N.A., et al., *A path for lignin valorization via additive manufacturing of high-performance sustainable composites with enhanced 3D printability.* J Science advances, 2018. **4**(12): p. eaat4967.
88. Davis, K. and T.S.J.C.o.i.c.b. Moon, *Tailoring microbes to upgrade lignin.* 2020. **59**: p. 23-29.
89. Abu-Omar, M.M., et al., *Guidelines for performing lignin-first biorefining.* J Energy Environmental Science 2021. **14**(1): p. 262-292.
90. Amir, R.M., et al., *Application of Fourier transform infrared (FTIR) spectroscopy for the identification of wheat varieties.* %J Journal of food science, 2013. **50**: p. 1018-1023.
91. Zaky, A.S., et al., *A new HPLC method for simultaneously measuring chloride, sugars, organic acids and alcohols in food samples.* 2017. **56**: p. 25-33.
92. Mackay, T.F., E.A. Stone, and J.F.J.N.R.G. Ayroles, *The genetics of quantitative traits: challenges and prospects.* 2009. **10**(8): p. 565-577.
93. Zhang, J., et al., *Recent advances in the transcriptional regulation of secondary cell wall biosynthesis in the woody plants.* J Frontiers in Plant Science, 2018. **9**: p. 1535.
94. Balding, D.J., *A tutorial on statistical methods for population association studies.* J Nature reviews genetics, 2006. **7**(10): p. 781-791.
95. Yu, J., et al., *A unified mixed-model method for association mapping that accounts for multiple levels of relatedness.* 2006. **38**(2): p. 203-208.
96. Segura, V., et al., *An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations.* 2012. **44**(7): p. 825-830.
97. Liu, X., et al., *Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies.* 2016. **12**(2): p. e1005767.
98. Huang, M., et al., *BLINK: a package for the next level of genome-wide association studies with both individuals and markers in the millions.* 2019. **8**(2): p. giy154.
99. Zhang, J., et al., *Genome-wide association studies and expression-based quantitative trait loci analyses reveal roles of HCT 2 in caffeoylquinic acid biosynthesis and its regulation by defense-responsive transcription factors in Populus.* New Phytologist, 2018. **220**(2): p. 502-516.
100. Matthews, M.L., et al., *A multiscale model of lignin biosynthesis for predicting bioenergy traits in Populus trichocarpa.* 2021. **19**: p. 168-182.
101. Bolger, A.M., M. Lohse, and B.J.B. Usadel, *Trimmomatic: a flexible trimmer for Illumina sequence data.* 2014. **30**(15): p. 2114-2120.

102. Dobin, A., et al., *STAR: ultrafast universal RNA-seq aligner.* 2013. **29**(1): p. 15-21.
103. Pertea, M., et al., *StringTie enables improved reconstruction of a transcriptome from RNA-seq reads.* 2015. **33**(3): p. 290-295.
104. Shi, R., et al., *Towards a systems approach for lignin biosynthesis in Populus trichocarpa: transcript abundance and specificity of the monolignol biosynthetic genes.* 2010. **51**(1): p. 144-163.
105. Sundell, D., et al., *AspWood: high-spatial-resolution transcriptome profiles reveal uncharacterized modularity of wood formation in Populus tremula.* 2017. **29**(7): p. 1585-1604.
106. Thomas, V.A., et al., *Comparative evaluation of Populus variants total sugar release and structural features following pretreatment and digestion by two distinct biological systems.* 2017. **10**(1): p. 1-16.
107. Happs, R.M., et al., *Comparison of methodologies used to determine aromatic lignin unit ratios in lignocellulosic biomass.* J Biotechnology for biofuels, 2021. **14**(1): p. 1-16.
108. Meng, X., et al., *An in-depth understanding of biomass recalcitrance using natural poplar variants as the feedstock.* 2017. **10**(1): p. 139-150.
109. Kljun, A., et al., *Comparative analysis of crystallinity changes in cellulose I polymers using ATR-FTIR, X-ray diffraction, and carbohydrate-binding module probes.* 2011. **12**(11): p. 4121-4126.
110. Bhagia, S., et al., *Nanoscale FTIR and Mechanical Mapping of Plant Cell Walls for Understanding Biomass Deconstruction.* J ACS Sustainable Chemistry, 2022. **10**(9): p. 3016-3026.
111. Allison, G.G., *Application of Fourier transform mid-infrared spectroscopy (FTIR) for research into biomass feed-stocks.* J Fourier Transforms–New Analytical Approaches FTIR Strategies, 2011: p. 71-88.
112. Lu, Y., et al., *Structural characterization of lignin and its degradation products with spectroscopic methods.* 2017. **2017**: p. 1-15.
113. Karimi, K. and M.J.J.B.t. Taherzadeh, *A critical review of analytical methods in pretreatment of lignocelluloses: composition, imaging, and crystallinity.* 2016. **200**: p. 1008-1018.
114. Li, M., et al., *31P NMR chemical shifts of solvents and products impurities in biomass pretreatments.* 2018. **6**(1): p. 1265-1270.
115. Kumar, A.K., S.J.B. Sharma, and bioprocessing, *Recent updates on different methods of pretreatment of lignocellulosic feedstocks: a review.* 2017. **4**(1): p. 1-19.
116. Jiang, F., et al., *Degradation profiles of non-lignin constituents of corn stover from dilute sulfuric acid pretreatment.* 2016. **36**(3): p. 192-204.
117. Kim, H., et al., *Characterization and elimination of undesirable protein residues in plant cell wall materials for enhancing lignin analysis by solution-state nuclear magnetic resonance spectroscopy.* 2017. **18**(12): p. 4184-4195.
118. Sun, N., et al., *Complete dissolution and partial delignification of wood in the ionic liquid 1-ethyl-3-methylimidazolium acetate.* 2009. **11**(5): p. 646-655.
119. Zhang, J., et al., *Understanding changes in cellulose crystalline structure of lignocellulosic biomass during ionic liquid pretreatment by XRD.* 2014. **151**: p. 402-405.

120. Weerachanchai, P., et al., *Improvement of biomass properties by pretreatment with ionic liquids for bioconversion process.* 2012. **111**: p. 453-459.
121. Bahrani, S., S. Raeissi, and M. Sarshar, *Experimental investigation of ionic liquid pretreatment of sugarcane bagasse with 1, 3-dimethylimadazolium dimethyl phosphate.* J Bioresource Technology, 2015. **185**: p. 411-415.
122. Cao, Q., et al., *Conversion of hexose into 5-hydroxymethylfurfural in imidazolium ionic liquids with and without a catalyst.* 2011. **346**(7): p. 956-959.
123. Tan, Y.T., G.C. Ngoh, and A.S.M.J.B.t. Chua, *Effect of functional groups in acid constituent of deep eutectic solvent for extraction of reactive lignin.* 2019. **281**: p. 359-366.
124. Jablonský, M., et al., *Deep eutectic solvents: fractionation of wheat straw.* 2015. **10**(4): p. 8039-8047.
125. Wan, Y.L. and Y.J. Mun. *Assessment of natural deep eutectic solvent pretreatment on sugar production from lignocellulosic biomass*. in *MATEC Web of Conferences*. 2018. EDP Sciences.
126. Saravana, P.S., et al., *Green and efficient extraction of polysaccharides from brown seaweed by adding deep eutectic solvent in subcritical water hydrolysis.* 2018. **198**: p. 1474-1484.
127. Sannigrahi, P., A.J. Ragauskas, and G.A. Tuskan, *Poplar as a feedstock for biofuels: a review of compositional characteristics.* Biofuels, Bioproducts and Biorefining, 2010. **4**(2): p. 209-226.
128. Chudy, R., et al., *The economics of dedicated hybrid poplar biomass plantations in the western US.* Biomass and Bioenergy, 2019. **124**: p. 114-124.
129. Feau, N., et al., *Recent advances related to poplar leaf spot and canker caused by Septoria musiva: Minireview/Minisynthèse.* Canadian Journal of Plant Pathology, 2010. **32**(2): p. 122-134.
130. Royle, D. and M. Ostry, *Disease and pest control in the bioenergy crops poplar and willow.* Biomass and Bioenergy, 1995. **9**(1-5): p. 69-79.
131. Cellerino, G.P., *Review of fungal diseases in Poplar.* Food and Agriculture Organization of The United Nations, Rome (FAO). AC492/E, 1999.
132. Ostry, M.E., L.F. Wilson, and S. Harold Jr, *Impact and control of Septoria musiva on hybrid poplars.* 1989.
133. Ostry, M.E., *Susceptibility of Populus species and hybrids to disease in the north central United States.* Plant Disease, 1985: p. 755.
134. Bryant, N.D., et al., *Transgenic poplar designed for biofuels.* Trends in plant science, 2020. **25**(9): p. 881-896.
135. Muchero, W., et al., *Association mapping, transcriptomics, and transient expression identify candidate genes mediating plant–pathogen interactions in a tree.* Proceedings of the National Academy of Sciences, 2018. **115**(45): p. 11573-11578.
136. Foster, A.J., et al., *Transcriptome analysis of poplar during leaf spot infection with Sphaerulina spp.* J PLoS One, 2015. **10**(9): p. e0138162.
137. Bucciarelli, B., et al., *Wound response characteristics as related to phenylpropanoid enzyme activity and lignin deposition in resistant and susceptible Populus tremuloides*

inoculated with Entoleuca mammata (Hypoxylon mammatum). Canadian Journal of Botany, 1998. **76**(7): p. 1282-1289.
138. Søndreli, K., et al., *Outbreak of Septoria Canker caused by Sphaerulina musiva on Populus trichocarpa in Eastern Oregon.* 2020. **104**(12): p. 3266-3266.
139. Sluiter, A., et al., *Determination of structural carbohydrates and lignin in biomass.* 2008. **1617**(1): p. 1-16.
140. Barros, J., et al., *4-Coumarate 3-hydroxylase in the lignin biosynthesis pathway is a cytosolic ascorbate peroxidase.* J Nature Communications, 2019. **10**(1): p. 1994.
141. Chen, F., et al., *A rapid thioacidolysis method for biomass lignin composition and tricin analysis.* Biotechnology for biofuels, 2021. **14**(1): p. 1-9.
142. Goacher, R.E., Y. Mottiar, and S.D. Mansfield, *ToF-SIMS imaging reveals that p-hydroxybenzoate groups specifically decorate the lignin of fibres in the xylem of poplar and willow.* J Holzforschung, 2021. **75**(5): p. 452-462.
143. Faix, O., *Classification of lignins from different botanical origins by FT-IR spectroscopy.* Holzforschung, 1991. **45**: p. 21-28.
144. Colom, X., et al., *Structural analysis of photodegraded wood by means of FTIR spectroscopy.* Polymer degradation and stability, 2003. **80**(3): p. 543-549.
145. Pandey, K.K. and A. Pitman, *FTIR studies of the changes in wood chemistry following decay by brown-rot and white-rot fungi.* International biodeterioration & biodegradation, 2003. **52**(3): p. 151-160.
146. Fengel, D. and G. Wegener, *Wood : Chemistry, Ultrastructure, Reactions*. 1983, Berlin/Boston, GERMANY: De Gruyter, Inc.
147. Pandey, K. and A. Pitman, *FTIR studies of the changes in wood chemistry following decay by brown-rot and white-rot fungi.* International biodeterioration biodegradation, 2003. **52**(3): p. 151-160.
148. Yang, H., et al., *Structural changes of lignins in natural Populus variants during different pretreatments.* 2020. **295**: p. 122240.
149. Xu, F., et al., *Determination of cell wall ferulic and p-coumaric acids in sugarcane bagasse.* 2005. **552**(1-2): p. 207-217.
150. Zhou, G., G. Taylor, and A.J.P.m. Polle, *FTIR-ATR-based prediction and modelling of lignin and energy contents reveals independent intra-specific variation of these traits in bioenergy poplars.* 2011. **7**: p. 1-10.
151. Shi, Z., et al., *Structural characterization of lignin from D. sinicus by FTIR and NMR techniques.* 2019. **12**(3): p. 235-243.
152. Kline, L.M., et al., *Simplified determination of lignin content in hard and soft woods via uv-spectrophotometric analysis of biomass dissolved in ionic liquids.* 2010. **5**(3).
153. Akinosho, H.O., et al., *Elucidating the structural changes to Populus lignin during consolidated bioprocessing with Clostridium thermocellum.* 2017. **5**(9): p. 7486-7491.
154. Kim, H., et al., *Characterization and elimination of undesirable protein residues in plant cell wall materials for enhancing lignin analysis by solution-state nuclear magnetic resonance spectroscopy.* J Biomacromolecules, 2017. **18**(12): p. 4184-4195.
155. Bhagia, S., et al., *Natural genetic variability reduces recalcitrance in poplar.* 2016. **9**(1): p. 1-12.

156. Do, C.-T., et al., *Both caffeoyl Coenzyme A 3-O-methyltransferase 1 and caffeic acid O-methyltransferase 1 are involved in redundant functions for lignin, flavonoids and sinapoyl malate biosynthesis in Arabidopsis.* 2007. **226**(5): p. 1117-1129.
157. Stewart, J.J., et al., *The Effects on Lignin Structure of Overexpression of Ferulate 5-Hydroxylase in Hybrid Poplar1* Plant Physiology, 2009. **150**(2): p. 621-635.
158. Mottiar, Y. and S.D. Mansfield, *Lignin p-hydroxybenzoylation is negatively correlated with syringyl units in poplar.* Frontiers in Plant Science, 2022. **13**.
159. Schnitzler, J.-P. and H.U.J.Z.f.N.C. Seitz, *Rapid responses of cultured carrot cells and protoplasts to an elicitor from the cell wall of Pythium aphanidermatum (Edson) Fitzp.* 1989. **44**(11-12): p. 1020-1028.
160. Frankenstein, C., U. Schmitt, and G.J.A.o.B. Koch, *Topochemical studies on modified lignin distribution in the xylem of poplar (Populus spp.) after wounding.* J Annals of Botany, 2006. **97**(2): p. 195-204.
161. MICHAEL, E.J.P.D., *Susceptibility of Populus species and hybrids to disease in the north central United States.* 1985: p. 755.
162. Newcombe, G. and M.J.P. Ostry, *Recessive resistance to Septoria stem canker of hybrid poplar.* 2001. **91**(11): p. 1081-1084.
163. Liu, Q., L. Luo, and L.J.I.j.o.m.s. Zheng, *Lignins: biosynthesis and biological functions in plants.* 2018. **19**(2): p. 335.
164. Hatakeyama, H. and T. Hatakeyama, *Lignin structure, properties, and applications*, in *Biopolymers*. 2009, Springer. p. 1-63.
165. Vanholme, R., et al., *Lignin biosynthesis and structure.* Plant physiology, 2010. **153**(3): p. 895-905.
166. Ragauskas, A.J., et al., *The path forward for biofuels and biomaterials.* J Science, 2006. **311**(5760): p. 484-489.
167. Richet, N., et al., *The response to daylight or continuous ozone of phenylpropanoid and lignin biosynthesis pathways in poplar differs between leaves and wood.* 2012. **236**(2): p. 727-737.
168. Rutigliano, F.A., et al., *Lignin decomposition in decaying leaves of Fagus sylvatica L. and needles of Abies alba Mill.* 1996. **28**(1): p. 101-106.
169. Xu, T., et al., *Insights into the Molecular Regulation of Lignin Content in Triploid Poplar Leaves.* J International Journal of Molecular Sciences, 2022. **23**(9): p. 4603.
170. Xu, T., et al., *Insights into the molecular regulation of lignin content in triploid poplar leaves.* 2022. **23**(9): p. 4603.
171. Fu, Y., et al., *PtrARF2. 1 is involved in regulation of leaf development and lignin biosynthesis in poplar trees.* 2019. **20**(17): p. 4141.
172. Zhou, S., et al., *Lignin valorization through thermochemical conversion: comparison of hardwood, softwood and herbaceous lignin.* ACS Sustainable Chemistry Engineering, 2016. **4**(12): p. 6608-6617.
173. Bergs, M., et al., *Miscanthus x giganteus stem versus leaf-derived lignins differing in monolignol ratio and linkage.* International journal of molecular sciences, 2019. **20**(5): p. 1200.

174. Jiang, B., et al., *Comparison of the structural characteristics of cellulolytic enzyme lignin preparations isolated from wheat straw stem and leaf.* ACS Sustainable Chemistry Engineering, 2017. **5**(1): p. 342-349.
175. Cabané, M., et al., *Condensed Lignins Are Synthesized in Poplar Leaves Exposed to Ozone.* Plant Physiology, 2004. **134**(2): p. 586-594.
176. Cuchietti, A., et al., *Leaf litter mixtures and neighbour effects: low-nitrogen and high-lignin species increase decomposition rate of high-nitrogen and low-lignin neighbours.* 2014. **82**: p. 44-51.
177. Steffen, K.T., et al., *Differential degradation of oak (Quercus petraea) leaf litter by litter-decomposing basidiomycetes.* 2007. **158**(5): p. 447-455.
178. Tschaplinski, T.J., et al., *Down-regulation of the caffeic acid O-methyltransferase gene in switchgrass reveals a novel monolignol analog.* Biotechnology for biofuels, 2012. **5**(1): p. 1-15.
179. Abraham, P.E., et al., *Transcript, protein and metabolite temporal dynamics in the CAM plant Agave.* Nature Plants, 2016. **2**(12): p. 1-10.
180. Kim, H., et al., *Characterization and elimination of undesirable protein residues in plant cell wall materials for enhancing lignin analysis by solution-state nuclear magnetic resonance spectroscopy.* Biomacromolecules, 2017. **18**(12): p. 4184-4195.
181. Barchet, G.L., et al., *Investigating the drought-stress response of hybrid poplar genotypes by metabolite profiling.* J Tree physiology, 2014. **34**(11): p. 1203-1219.
182. Jia, H., et al., *Comparative metabolomics analysis reveals different metabolic responses to drought in tolerant and susceptible poplar species.* 2020. **168**(3): p. 531-546.
183. Jamil, M., et al., *Inducing drought tolerance in wheat through combined use of l-tryptophan and Pseudomonas fluorescens.* 2018. **55**(2).
184. Khan, N., et al., *UPLC-HRMS-based untargeted metabolic profiling reveals changes in chickpea (Cicer arietinum) metabolome following long-term drought stress.* 2019. **42**(1): p. 115-132.
185. Jin, Z., et al., *Changes in lignin content of leaf litters during mulching.* 2003. **64**(5): p. 1023-1031.
186. Shao, Z., et al., *Modification of the aspen lignin structure during integrated fractionation process of autohydrolysis and formic acid delignification.* International Journal of Biological Macromolecules, 2020. **165**: p. 1727-1737.
187. Martín-Sampedro, R., et al., *Characterization of lignins from Populus alba L. generated as by-products in different transformation processes: Kraft pulping, organosolv and acid hydrolysis.* International journal of biological macromolecules, 2019. **126**: p. 18-29.
188. Faix, O., *Fourier transform infrared spectroscopy*, in *Methods in lignin chemistry*. 1992, Springer. p. 83-109.
189. Abdelkafi, F., et al., *Structural analysis of alfa grass (Stipa tenacissima L.) lignin obtained by acetic acid/formic acid delignification.* Biomacromolecules, 2011. **12**(11): p. 3895-3902.
190. Wang, P., et al., *Structural changes to aspen wood lignin during autohydrolysis pretreatment.* BioResources, 2016. **11**(2): p. 4086-4103.

191. Popova, Y.A., et al., *Comprehensive analysis of the chemical structure of lignin from raspberry stalks (Rubus idaeus L.).* International Journal of Biological Macromolecules, 2020. **164**: p. 3814-3822.
192. Abidi, N., L. Cabrales, and C.H. Haigler, *Changes in the cell wall and cellulose content of developing cotton fibers investigated by FTIR spectroscopy.* J Carbohydrate Polymers, 2014. **100**: p. 9-16.
193. Kline, L.M., et al., *Simplified determination of lignin content in hard and soft woods via uv-spectrophotometric analysis of biomass dissolved in ionic liquids.* J BioResources, 2010. **5**(3).
194. Clausen, T.P., et al., *Chemical model for short-term induction in quaking aspen (Populus tremuloides) foliage against herbivores.* 1989. **15**: p. 2335-2346.
195. Ruuhola, T. and R. Julkunen-Tiitto, *Trade-off between synthesis of salicylates and growth of micropropagated Salix pentandra.* Journal of chemical ecology, 2003. **29**(7): p. 1565-1588.
196. Wildermuth, M.C., *Variations on a theme: synthesis and modification of plant benzoic acids.* Current opinion in plant biology, 2006. **9**(3): p. 288-296.
197. Widhalm, J.R. and N. Dudareva, *A familiar ring to it: biosynthesis of plant benzoic acids.* Molecular plant, 2015. **8**(1): p. 83-97.
198. Abreu, I.N., et al., *A metabolite roadmap of the wood-forming tissue in Populus tremula.* New Phytologist, 2020. **228**(5): p. 1559-1572.
199. Barros, J., et al., *The cell biology of lignification in higher plants.* Annals of Botany, 2015. **115**(7): p. 1053-1074.
200. Silva, N.V.E., P. Mazzafera, and I. Cesarino, *Should I stay or should I go: are chlorogenic acids mobilized towards lignin biosynthesis?* Phytochemistry, 2019. **166**: p. 112063.
201. Winkel-Shirley, B., *Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology.* Plant physiology, 2001. **126**(2): p. 485-493.
202. Cen, H., et al., *Overexpression of MsASMT1 promotes plant growth and decreases flavonoids biosynthesis in transgenic alfalfa (Medicago sativa L.).* Frontiers in plant science, 2020. **11**: p. 489.
203. Bryant, N.D., et al., *Transgenic Poplar Designed for Biofuels.* Trends in Plant Science, 2020.
204. Li, M., Y. Pu, and A.J. Ragauskas, *Current understanding of the correlation of lignin structure with biomass recalcitrance.* Frontiers in chemistry, 2016. **4**: p. 45.
205. Liu, Q., L. Luo, and L. Zheng, *Lignins: biosynthesis and biological functions in plants.* International journal of molecular sciences, 2018. **19**(2): p. 335.
206. Bose, S.K., et al., *Lignin content versus syringyl to guaiacyl ratio amongst poplars.* Bioresource technology, 2009. **100**(4): p. 1628-1633.
207. del Río, J.C., et al., *Lignin monomers from beyond the canonical monolignol biosynthetic pathway: another brick in the wall.* 2020. **8**(13): p. 4997-5012.
208. Ralph, J., S. Karlen, and J. Mobley, *Synthesis of paracetamol (acetaminophen) from biomass-derived p-hydroxybenzamide*. 2019, Google Patents.

209. Porth, I., et al., *P opulus trichocarpa cell wall chemistry and ultrastructure trait variation, genetic control and genetic correlations.* New Phytologist, 2013. **197**(3): p. 777-790.
210. Fahrenkrog, A.M., et al., *Genome-wide association study reveals putative regulators of bioenergy traits in Populus deltoides.* J New Phytologist, 2017. **213**(2): p. 799-811.
211. Allwright, M.R., et al., *Biomass traits and candidate genes for bioenergy revealed through association genetics in coppiced European Populus nigra (L.).* Biotechnology for biofuels, 2016. **9**(1): p. 195.
212. Guerra, F.P., et al., *Association genetics of chemical wood properties in black poplar (Populus nigra).* J New Phytologist, 2013. **197**(1): p. 162-176.
213. Muchero, W., et al., *High-resolution genetic mapping of allelic variants associated with cell wall chemistry in Populus.* BMC genomics, 2015. **16**(1): p. 1-14.
214. Wang, B., A.M. Goodpaster, and M.A. Kennedy, *Coefficient of variation, signal-to-noise ratio, and effects of normalization in validation of biomarkers from NMR-based metabonomics studies.* Chemometrics and Intelligent Laboratory Systems, 2013. **128**: p. 9-16.
215. de Vries, L., et al., *Stacking of a low-lignin trait with an increased guaiacyl and 5-hydroxyguaiacyl unit trait leads to additive and synergistic effects on saccharification efficiency in Arabidopsis thaliana.* Biotechnology for Biofuels, 2018. **11**(1): p. 1-14.
216. Shen, B., C. Li, and M.C. Tarczynski, *High free-methionine and decreased lignin content result from a mutation in the Arabidopsis S-adenosyl-L-methionine synthetase 3 gene.* The Plant Journal, 2002. **29**(3): p. 371-380.
217. Chen, H., et al., *Hierarchical transcription factor and chromatin binding network for wood formation in Populus trichocarpa.* The Plant Cell, 2019. **31**(3): p. 602-626.
218. Zhang, X., et al., *Involvement of the R2R3-MYB transcription factor MYB21 and its homologs in regulating flavonol accumulation in Arabidopsis stamen.* Journal of Experimental Botany, 2021. **72**(12): p. 4319-4332.
219. He, M., et al., *Rab-H1b is essential for trafficking of cellulose synthase and for hypocotyl growth in Arabidopsis thaliana.* Journal of Integrative Plant Biology, 2018. **60**(11): p. 1051-1069.
220. Wu, A.M., et al., *The Arabidopsis IRX10 and IRX10-LIKE glycosyltransferases are critical for glucuronoxylan biosynthesis during secondary cell wall formation.* The Plant Journal, 2009. **57**(4): p. 718-731.
221. Wang, J.P., et al., *Improving wood properties for wood utilization through multi-omics integration in lignin biosynthesis.* Nature communications, 2018. **9**(1): p. 1-16.
222. Rubin, G., et al., *Members of the LBD family of transcription factors repress anthocyanin synthesis and affect additional nitrogen responses in Arabidopsis.* The Plant Cell, 2009. **21**(11): p. 3567-3584.
223. Sibout, R., et al., *Structural redesigning Arabidopsis lignins into alkali-soluble lignins through the expression of p-coumaroyl-CoA: monolignol transferase PMT.* Plant physiology, 2016. **170**(3): p. 1358-1366.
224. Stewart, J.J., et al., *The effects on lignin structure of overexpression of ferulate 5-hydroxylase in hybrid poplar1.* Plant physiology, 2009. **150**(2): p. 621-635.

225. Yoo, C.G., et al., *Significance of lignin S/G ratio in biomass recalcitrance of Populus trichocarpa variants for bioethanol production.* ACS Sustainable Chemistry & Engineering, 2018. **6**(2): p. 2162-2168.
226. Mottiar, Y. and S.D. Mansfield, *Lignin p-Hydroxybenzoylation Is Negatively Correlated With Syringyl Units in Poplar.* 2022. **13**.
227. Takeda, Y., et al., *Regulation of CONIFERALDEHYDE 5-HYDROXYLASE expression to modulate cell wall lignin structure in rice.* J Planta, 2017. **246**(2): p. 337-349.
228. Pratelli, R., et al., *The Ubiquitin E3 Ligase LOSS OF GDU2 Is Required for GLUTAMINE DUMPER1-Induced Amino Acid Secretion in Arabidopsis* Plant Physiology, 2012. **158**(4): p. 1628-1642.
229. Pratelli, R., et al., *Stimulation of Nonselective Amino Acid Export by Glutamine Dumper Proteins.* Plant Physiology, 2009. **152**(2): p. 762-773.
230. Perkins, M., R.A. Smith, and L. Samuels, *The transport of monomers during lignification in plants: anything goes but how?* J Current Opinion in Biotechnology, 2019. **56**: p. 69-74.
231. Tanaka, K., *The proteasome: overview of structure and functions.* Proceedings of the Japan Academy, Series B, 2009. **85**(1): p. 12-36.
232. Staszczak, M., *The 26S proteasome of the lignin-degrading Basidiomycete Phlebia radiata.* Enzyme and microbial technology, 2007. **40**(2): p. 347-353.
233. Feng, H., et al., *GhHUB2, a ubiquitin ligase, is involved in cotton fiber development via the ubiquitin–26S proteasome pathway.* Journal of experimental botany, 2018. **69**(21): p. 5059-5075.
234. Nodzon, L.A., et al., *The ubiquitin ligase XBAT32 regulates lateral root development in Arabidopsis.* The Plant Journal, 2004. **40**(6): p. 996-1006.
235. Prasad, M.E., et al., *Arabidopsis RING E3 ligase XBAT32 regulates lateral root production through its role in ethylene biosynthesis.* Plant physiology, 2010. **153**(4): p. 1587-1596.
236. Prasad, M.E. and S.L. Stone, *Further analysis of XBAT32, an Arabidopsis RING E3 ligase, involved in ethylene biosynthesis.* Plant signaling & behavior, 2010. **5**(11): p. 1425-1429.
237. Carvalho, S.D., et al., *XBAT35, a novel Arabidopsis RING E3 ligase exhibiting dual targeting of its splice isoforms, is involved in ethylene-mediated regulation of apical hook curvature.* Molecular plant, 2012. **5**(6): p. 1295-1309.
238. Ge, D., et al., *Genomics, expression, and function analyses of XB3 family genes in cotton.* Genomics, 2021. **113**(1): p. 245-256.
239. Moura, J.C.M.S., et al., *Abiotic and biotic stresses and changes in the lignin content and composition in plants.* 2010. **52**(4): p. 360-376.
240. Sanderson, M.A., et al., *Switchgrass as a biofuels feedstock in the USA.* 2006. **86**(Special Issue): p. 1315-1325.
241. McLaughlin, S.B., L.A.J.B. Kszos, and bioenergy, *Development of switchgrass (Panicum virgatum) as a bioenergy feedstock in the United States.* 2005. **28**(6): p. 515-535.
242. Lovell, J.T., et al., *Genomic mechanisms of climate adaptation in polyploid bioenergy switchgrass.* 2021. **590**(7846): p. 438-444.

243. Vanholme, R., et al., *Lignin biosynthesis and its integration into metabolism.* 2019. **56**: p. 230-239.
244. Mann, D.G., et al., *Rapid assessment of lignin content and structure in switchgrass (Panicum virgatum L.) grown under different environmental conditions.* 2009. **2**: p. 246-256.
245. Bahri, B.A., et al., *Natural variation in lignin and pectin biosynthesis-related genes in switchgrass (Panicum virgatum L.) and association of SNP variants with dry matter traits.* J BioEnergy research, 2020. **13**: p. 79-99.
246. Lupoi, J.S., et al., *Assessment of lignocellulosic biomass using analytical spectroscopy: an evolution to high-throughput techniques.* 2014. **7**: p. 1-23.
247. Ralph, J., et al., *Solution-state NMR of lignins.* 1999: p. 55-108.
248. TIAN, X.-m., et al., *Association mapping of lignin response to Verticillium wilt through an eight-way MAGIC population in Upland cotton.* 2023. **22**(5): p. 1324-1337.
249. Wang, J., et al., *Genome-wide analysis of seed acid detergent lignin (ADL) and hull content in rapeseed (Brassica napus L.).* 2015. **10**(12): p. e0145045.
250. Nguyen, D.T., et al., *Association mapping identifies quantitative trait loci (QTL) for digestibility in rice straw.* 2020. **13**(1): p. 1-16.
251. Xu, S., et al., *Genome-wide association study identifies novel candidate loci or genes affecting stalk strength in maize.* 2023. **11**(1): p. 220-227.
252. Du, Q., et al., *Genome-wide association studies to improve wood properties: challenges and prospects.* 2018. **9**: p. 1912.
253. Lowry, D.B., et al., *QTL× environment interactions underlie adaptive divergence in switchgrass across a large latitudinal gradient.* 2019. **116**(26): p. 12933-12941.
254. Li, M., et al., *Downregulation of pectin biosynthesis gene GAUT4 leads to reduced ferulate and lignin-carbohydrate cross-linking in switchgrass.* 2019. **2**(1): p. 22.
255. Wang, J., Z.J.G. Zhang, proteomics, and bioinformatics, *GAPIT version 3: boosting power and accuracy for genomic association and prediction.* 2021. **19**(4): p. 629-640.
256. VanRaden, P.M.J.J.o.d.s., *Efficient methods to compute genomic predictions.* 2008. **91**(11): p. 4414-4423.
257. Tuskan, G.A., et al., *Population-level approaches reveal novel aspects of lignin biosynthesis, content, composition and structure.* 2019. **56**: p. 250-257.
258. Harman-Ware, A.E., et al., *A thioacidolysis method tailored for higher-throughput quantitative analysis of lignin monomers.* 2016. **11**(10): p. 1268-1273.
259. Chen, F., et al., *A rapid thioacidolysis method for biomass lignin composition and tricin analysis.* 2021. **14**(1): p. 1-9.
260. Happs, R.M., et al., *Comparison of methodologies used to determine aromatic lignin unit ratios in lignocellulosic biomass.* 2021. **14**(1): p. 1-16.
261. Ralph, J.J.P.R., *Hydroxycinnamates in lignification.* 2010. **9**: p. 65-83.
262. Bryant, N., et al., *Novel candidate genes for lignin structure identified through genome-wide association study of naturally varying Populus trichocarpa.* 2023. **14**: p. 1153113.
263. Eloy, N.B., et al., *Silencing CHALCONE SYNTHASE in maize impedes the incorporation of tricin into lignin and increases lignin content.* 2017. **173**(2): p. 998-1016.

264. Dumitrache, A., et al., *Consolidated bioprocessing of Populus using Clostridium (Ruminiclostridium) thermocellum: a case study on the impact of lignin composition and structure.* 2016. **9**(1): p. 1-14.
265. Shen, H., et al., *Functional characterization of the switchgrass (Panicum virgatum) R2R3-MYB transcription factor PvMYB4 for improvement of lignocellulosic feedstocks.* 2012. **193**(1): p. 121-136.
266. Wang, H., et al., *Mutation of WRKY transcription factors initiates pith secondary wall formation and increases stem biomass in dicotyledonous plants.* 2010. **107**(51): p. 22338-22343.
267. Groen, A.J., et al., *Identification of trans-Golgi network proteins in Arabidopsis thaliana root tissue.* 2014. **13**(2): p. 763-776.
268. Li, M., et al., *Downregulation of pectin biosynthesis gene GAUT4 leads to reduced ferulate and lignin-carbohydrate cross-linking in switchgrass.* J Communications biology, 2019. **2**(1): p. 22.
269. Hao, Z., et al., *Loss of Arabidopsis GAUT12/IRX8 causes anther indehiscence and leads to reduced G lignin associated with altered matrix polysaccharide deposition.* 2014. **5**: p. 357.
270. Davin, L.B., et al., *Stereoselective bimolecular phenoxy radical coupling by an auxiliary (dirigent) protein without an active center.* 1997. **275**(5298): p. 362-367.
271. Tripathi, P., R.C. Rabara, and P.J.J.P. Rushton, *A systems biology perspective on the role of WRKY transcription factors in drought responses in plants.* 2014. **239**: p. 255-266.
272. Naoumkina, M.A., X. He, and R.A.J.B.p.b. Dixon, *Elicitor-induced transcription factors for metabolic reprogramming of secondary metabolism in Medicago truncatula.* 2008. **8**(1): p. 1-14.
273. Guillaumie, S., et al., *The grapevine transcription factor WRKY2 influences the lignin pathway and xylem development in tobacco.* 2010. **72**(1-2): p. 215.
274. Teng, R., et al., *CsWRKY13, a novel WRKY transcription factor of Camellia sinensis, involved in lignin biosynthesis and accumulation.* 2021. **1**(1): p. 1-9.
275. Besseau, S., et al., *Flavonoid accumulation in Arabidopsis repressed in lignin synthesis affects auxin transport and plant growth.* J The Plant Cell, 2007. **19**(1): p. 148-162.
276. Ishida, T., et al., *Arabidopsis TRANSPARENT TESTA GLABRA2 is directly regulated by R2R3 MYB transcription factors and is involved in regulation of GLABRA2 transcription in epidermal differentiation.* 2007. **19**(8): p. 2531-2543.
277. Zhong, R., et al., *The poplar MYB master switches bind to the SMRE site and activate the secondary wall biosynthetic program during wood formation.* 2013. **8**(7): p. e69219.
278. Olsen, A.N., et al., *NAC transcription factors: structurally distinct, functionally diverse.* 2005. **10**(2): p. 79-87.
279. Tohge, T., et al., *Phytochemical genomics in Arabidopsis thaliana: a case study for functional identification of flavonoid biosynthesis genes.* 2007. **79**(4): p. 811-823.
280. Lepiniec, L., et al., *Genetics and biochemistry of seed flavonoids.* 2006. **57**: p. 405-430.
281. McLaughlin, S., et al. *Evaluating physical, chemical, and energetic properties of perennial grasses as biofuels*. in *Proc. Bioenergy*. 1996.

282. Schwarz, G. and R.R.J.A.R.P.B. Mendel, *Molybdenum cofactor biosynthesis and molybdenum enzymes.* 2006. **57**: p. 623-647.
283. Kaiser, B.N., et al., *The role of molybdenum in agricultural plant production.* 2005. **96**(5): p. 745-754.
284. Baxter, I., et al., *Variation in molybdenum content across broadly distributed populations of Arabidopsis thaliana is controlled by a mitochondrial molybdenum transporter (MOT1).* J PLoS Genetics, 2008. **4**(2): p. e1000004.
285. Ide, Y., et al., *Effects of molybdenum deficiency and defects in molybdate transporter MOT1 on transcript accumulation and nitrogen/sulphur metabolism in Arabidopsis thaliana.* 2011. **62**(4): p. 1483-1497.
286. Ralph, J., et al., *Lignins: natural polymers from oxidative coupling of 4-hydroxyphenyl-propanoids.* 2004. **3**: p. 29-60.
287. Lu, F., et al., *A facile spectroscopic method for measuring lignin content in lignocellulosic biomass.* 2021. **23**(14): p. 5106-5112.
288. Mottiar, Y. and S.D. Mansfield, *Lignin p-Hydroxybenzoylation Is Negatively Correlated With Syringyl Units in Poplar.* J Frontiers in Plant Science, 2022. **13**.
289. de Vries, L., et al., *p HBMT1, a BAHD-family monolignol acyltransferase, mediates lignin acylation in poplar.* 2022. **188**(2): p. 1014-1027.
290. Zhao, Y., et al., *Monolignol acyltransferase for lignin p-hydroxybenzoylation in Populus.* 2021. **7**(9): p. 1288-1300.
291. Cai, Y., et al., *Enhancing digestibility and ethanol yield of Populus wood via expression of an engineered monolignol 4-O-methyltransferase.* 2016. **7**(1): p. 11989.
292. Yoo, C.G., et al., *Insights of biomass recalcitrance in natural Populus trichocarpa variants for biomass conversion.* 2017. **19**(22): p. 5467-5478.
293. Bryant, N., et al., *Cell wall response of field grown Populus to Septoria infection.* 2023. **14**: p. 1089011.
294. Kim, S.H., C.M. Lee, and K.J.K.J.o.C.E. Kafle, *Characterization of crystalline cellulose in biomass: Basic principles, applications, and limitations of XRD, NMR, IR, Raman, and SFG.* 2013. **30**: p. 2127-2141.
295. Cai, Y., et al., *Enhancing digestibility and ethanol yield of Populus wood via expression of an engineered monolignol 4-O-methyltransferase.* J Frontiers in Plant Science, 2016. **7**(1): p. 11989.
296. Yoo, C.G., et al., *Insights of biomass recalcitrance in natural Populus trichocarpa variants for biomass conversion.* J Green Chemistry, 2017. **19**(22): p. 5467-5478.
297. Gioia, C., et al., *Lignin-based epoxy resins: Unravelling the relationship between structure and material properties.* 2020. **21**(5): p. 1920-1928.
298. Nirmale, T.C., B.B. Kale, and A.J.J.I.j.o.b.m. Varma, *A review on cellulose and lignin based binders and electrodes: Small steps towards a sustainable lithium ion battery.* 2017. **103**: p. 1032-1043.
299. Azadi, P., et al., *Liquid fuels, hydrogen and chemicals from lignin: A critical review.* J Renewable Sustainable Energy Reviews, 2013. **21**: p. 506-523.
300. Zmejkoski, D., et al., *Bacterial cellulose-lignin composite hydrogel as a promising agent in chronic wound healing.* 2018. **118**: p. 494-503.

301. Yu, H., et al., *Lignin Nanoparticles with High Phenolic Content as Efficient Antioxidant and Sun-Blocker for Food and Cosmetics.* 2023. **11**(10): p. 4082-4092.
302. Downs, C., et al., *Toxicological effects of the sunscreen UV filter, benzophenone-2, on planulae and in vitro cells of the coral, Stylophora pistillata.* 2014. **23**: p. 175-191.
303. Corinaldesi, C., et al., *Impact of inorganic UV filters contained in sunscreen products on tropical stony corals (Acropora spp.).* 2018. **637**: p. 1279-1285.
304. Matta, M.K., et al., *Effect of sunscreen application on plasma concentration of sunscreen active ingredients: a randomized clinical trial.* 2020. **323**(3): p. 256-267.

Appendix:

Appendix A – Supporting information for cell wall response of field grown *Populus* to *Spetoria* infection

TABLE 10-1 LIGNIN COMPOSITION RESULTS FROM THIOACIDOLYSIS AND NMR ANALYSES

		% (Thioacidolysis)				% (NMR)			
Genotype	**Status**	**H**	**G**	**S**	**S/G**	**H**	**G**	**S**	**S/G**
BESC-76	Healthy	0.42	35.29	64.29	1.82	1.23	29.63	69.15	2.33
BESC-76	Healthy	0.32	34.30	65.37	1.91	1.33	27.87	70.81	2.54
BESC-76	Healthy	0.36	33.93	65.71	1.94	0.90	29.65	69.45	2.34
13127	Healthy	0.60	34.21	65.20	1.91	1.11	31.46	67.43	2.14
13127	Reaction Zone	0.76	33.22	66.02	1.99	1.03	31.33	67.65	2.16
13127	Infected	0.64	35.27	64.08	1.82	1.95	29.87	68.18	2.28
HOMC-21-5	Healthy	0.44	33.48	66.08	1.97	1.52	28.13	70.35	2.50
HOMC-21-5	Reaction Zone	0.42	32.86	66.73	2.03	1.50	25.75	72.75	2.83
HOMC-21-5	Infected	0.49	33.96	65.55	1.93	1.81	29.08	69.11	2.38
BESC-144	Healthy	0.57	33.07	66.36	2.01	2.55	28.96	68.49	2.37
BESC-144	Reaction Zone	0.66	35.00	64.33	1.84	2.36	29.95	67.69	2.26
BESC-144	Infected	1.10	34.91	63.99	1.83	1.18	29.43	69.39	2.36
BESC-335	Reaction Zone	0.51	33.26	66.22	1.99	1.35	27.56	71.09	2.58
BESC-335	Infected	0.56	31.52	67.92	2.15	1.93	26.63	71.44	2.68
GW-9763	Healthy	0.89	32.61	66.50	2.04	1.36	28.10	70.54	2.51

TABLE 10-1 CONTINUED

		% (Thioacidolysis)				% (NMR)			
Genotype	**Status**	**H**	**G**	**S**	**S/G**	**H**	**G**	**S**	**S/G**
GW-9763	Healthy	0.42	32.08	67.50	2.10	1.98	26.24	71.77	2.73
GW-9763	Healthy	0.42	32.62	66.96	2.05	1.93	28.92	69.16	2.39
Note: samples BESC-144 Infected and GW-9763 healthy yielded low level of internal standard and lignin									

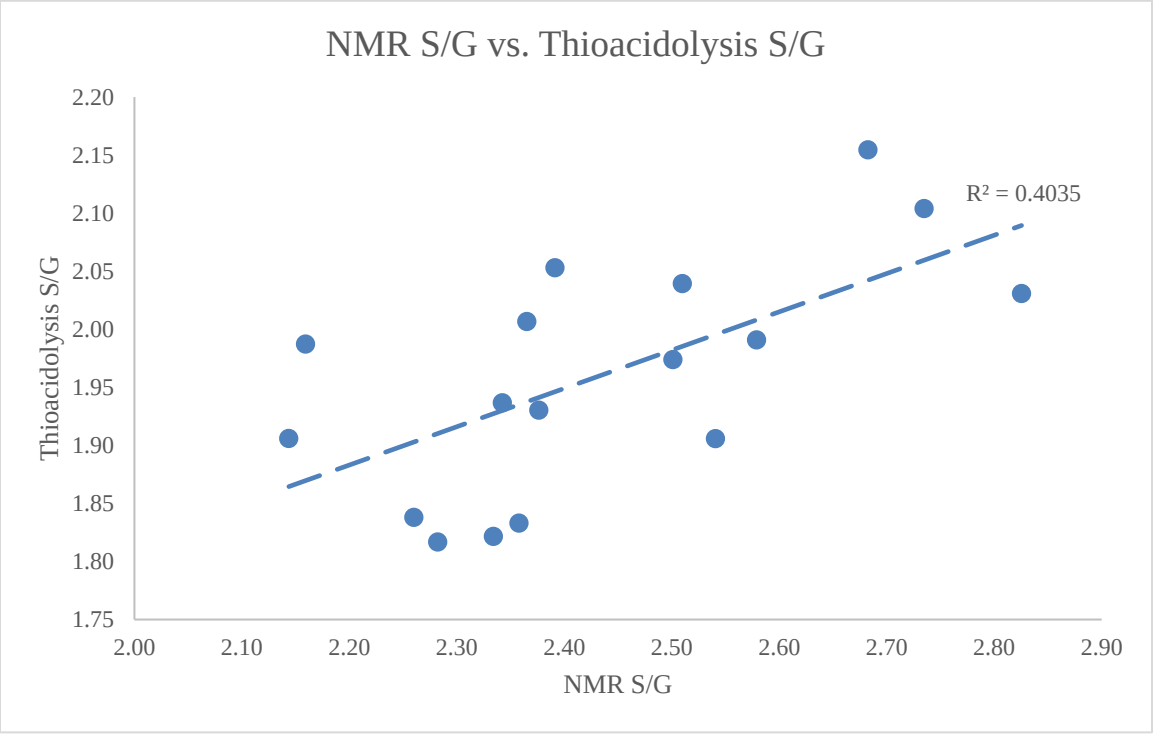


FIGURE 10-1 COMPARISON OF S/G RATIO DETERMINED BY NMR AND THIOACIDOLYSIS

All samples were analyzed by HSQC NMR. It is known that HSQC NMR tends to overestimate PB content. Therefore, samples were subjected to alkaline hydrolysis and HPLC analysis to provide a quantitative measurement of PB content and validate the HSQC NMR results. For the HPLC measurement, PB content is normalized by Klason lignin content. Samples (○) exhibited good correlation (R^2=0.712) between NMR and HPLC measurements. Due to material availability, not all samples were analyzed by HPLC. Additionally, running samples in duplicate was not possible. Therefore, a sample referred to as BESC standard poplar (▲) was utilized for replicate runs to estimate the variability for both HPLC and HSQC NMR. This reference material has been utilized in previous studies and was measured to have a Klason lignin content of 24.5%. The PB content of BESC standard poplar was measured to be 2.13% ± 0.20% by duplicate HPLC analysis, and 16.27% ± 0.36% by triplicate HSQC NMR analysis. As the subset of ten samples from this study were unable to be run in duplicate, the variability from this sample was utilized for error bars. The trendline in Figure S3 was established using only samples (i.e., excluding the BESC standard reference). The BESC standard reference is in excellent agreement with this trendline, providing confidence that this method can validate across a wide range of PB content. If the BESC standard reference is included when generating the trendline, the R^2 value increases to approximately 0.91.

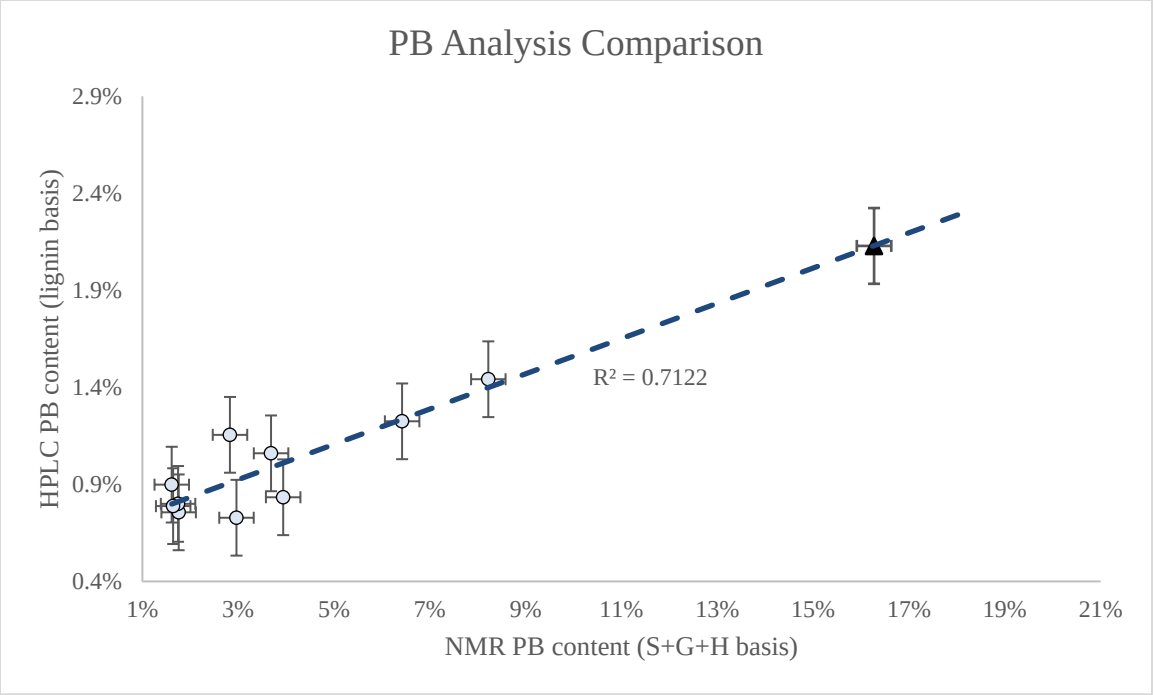


FIGURE 10-2 COMPARISON OF PB MEASUREMENT BY NMR AND HPLC

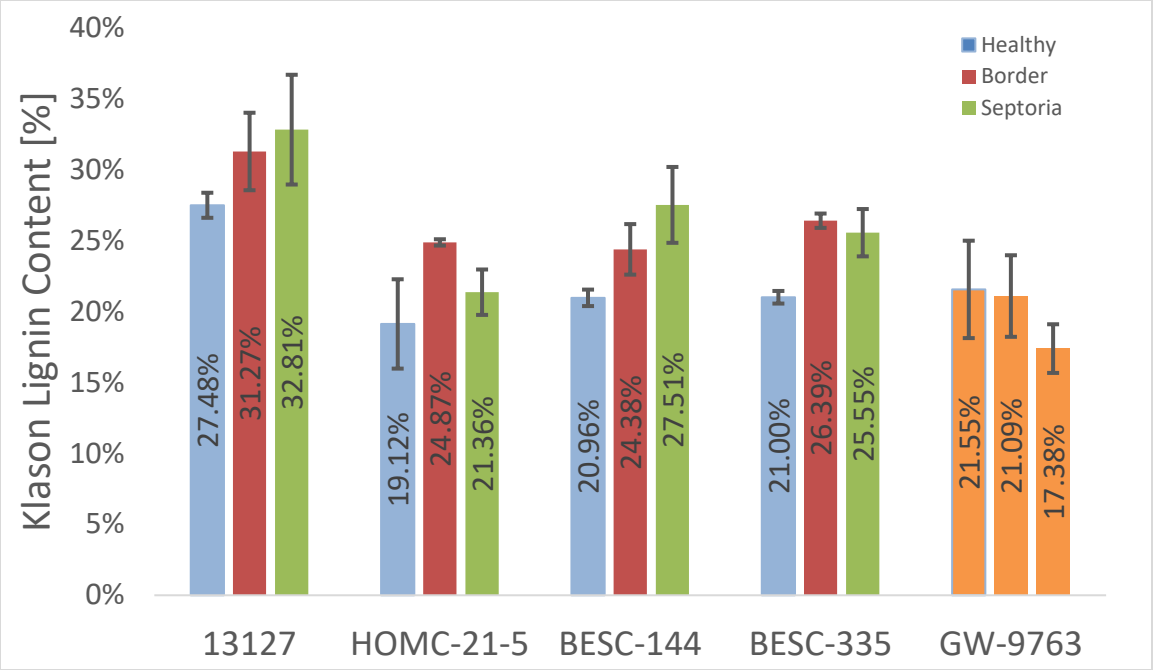


FIGURE 10-3 KLASON LIGNIN CONTENT WITH HEALTHY CONTROL SAMPLE GW-9763

Healthy, Reaction Zone, and infected region of BESC-335 was analyzed by whole cell wall (WCW) HSQC NMR by directly dissolving extractive-free, ball-milled biomass in DMSO-d_6/HMPA-d_{18} (4:1). Polysaccharide signals in the anomeric and non-anomeric regions are shown. While some minor differences were observed, polysaccharide structures are generally conserved.

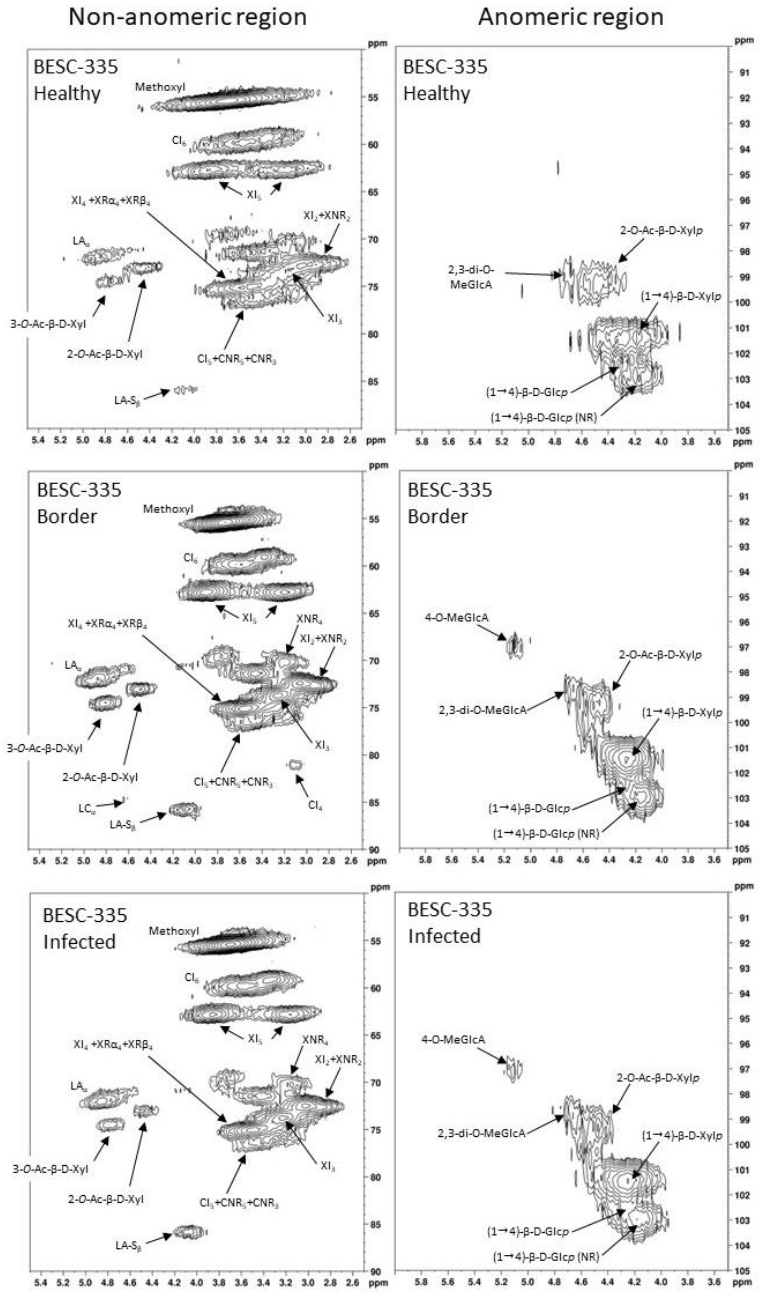


FIGURE 10-4 WHOLE CELL WALL HSQC NMR SPECTRA OF BESC-335

TABLE 10-2 ASSIGNMENT OF CROSS PEAKS FROM WHOLE CELL WALL HSQC NMR SPECTRA

Non-anomeric region			
Signal	**Description**	**δ_H**	**δ_C**
CI4	Internal cellulose unit	3.08	81.21
CI5	Internal cellulose unit	3.56	76.71
CI6	Internal cellulose unit	3.56	59.95
XI2	Internal xylose unit	3.02	72.53
XI4	Internal xylose unit	3.51	75.20
XI5	Internal xylose unit	3.17/3.87	62.81
XNR_2	Xylose non-reducing end	3.02	72.53
$XR\alpha_4$	Xylose reducing end	3.51	75.20
$XR\beta_4$	Xylose reducing end	3.51	75.20
CNR_3	Cellulose non-reducing end	3.56	76.71
CNR_5	Cellulose non-reducing end	3.56	76.71
LA-Sβ	β-O-4 linkage	4.1	85.85
3-O-Ac-β-D-Xyl	Acetylated xylan	4.78	74.51
2-*O*-Ac-β-D-Xyl	Acetylated xylan	4.46	73.07
LA_α	β-O-4 linkage	4.85	72.11
LC	β-β linkage	4.65	84.73
Anomeric region			
Signal	**Description**	**δ_H**	**δ_C**
4-O-MeGlcA	4-O-methyl-α-D-glucuronic acid	5.12	96.93
2,3-di-O-MeGlcA	2,3-di-O-methyl-a-d-glucuronic acid	4.71	98.75
(1→4)-β-D-Glc*p*	Internal cellulose unit	4.26	102.49
(1→4)-β-D-Glc*p* (NR)	Cellulose non-reducing end	4.17	102.02
2-O-Ac-β-D-Xyl*p*	acetylated xylosyl residues at C2	4.47	99.21
(1→4)-β-D-Xyl*p*	Internal xylan unit	4.26	101.53

TABLE 10-3 AVERAGE AND STANDARD DEVIATION OF PB CONTENT FROM TECHNICAL REPLIACTES OF STANDARD POPULUS BIOMASS ANALYZED BY ALKALINE HYDROLYSIS AND HPLC

BESC Standard Reference Material	
PB (HPLC)	2.13% ± 0.20% (2 replicates)

TABLE 10-4 RESULTS FROM INTERNAL PROTOCOL FOR ANALYZING TECHNICAL REPLICATES OF BIOMASS LIGNIN BY THIOACIDOLYSIS

ID#	**Sample**	**Rep.**	**H%**	**C%**	**G%**	**S%**	**S/G**
1	Genotype 1	R1	0.82	0.82	44.40	53.96	1.22
2	Genotype 1	R2	0.70	0.86	43.77	54.67	1.25
3	Genotype 1	R3	1.08	0.74	45.79	52.39	1.14
4	Genotype 2	R1	1.20	0.66	49.07	49.06	1.00
5	Genotype 2	R2	1.30	0.66	47.95	50.10	1.04
6	Genotype 2	R3	1.48	0.63	54.02	43.87	0.81
7	Genotype 3	R1	0.86	0.20	90.66	6.18	0.07
8	Genotype 3	R2	0.49	0.15	94.87	4.48	0.05
9	Genotype 3	R3	0.52	0.15	95.15	4.18	0.04
10	Genotype 4	R1	1.39	0.43	79.56	18.61	0.23
11	Genotype 4	R2	1.08	0.30	83.50	15.11	0.18
12	Genotype 4	R3	1.01	0.37	77.36	21.25	0.27

Appendix B – Supporting information for novel candidate genes for lignin structure identified through genome-wide association study of naturally varying *Populus trichocarpa*

There were 10 samples from the GWAS population that were measured to have high levels of H units and identified as outliers by JMP statistical software. It has been reported that amino acid contamination can potentially contaminate the $H_{2/6}$ signal in the HSQC spectra, resulting in the overestimation of H units. To test for potential contamination, four high H unit samples were selected for protease treatment. Protease treatment was conducted per a published procedure [180]. Briefly, enzyme lignin (EL) was combined with Proteinase K in a phosphate buffer and shaken/incubated for 48h. Afterwards, samples were washed twice with DI water and lyophilized. Samples were then prepared and analyzed via HSQC NMR per the procedure in the main text. The abundance of H units in the enzyme lignin (before protease) and the protease treated lignin (after protease) are summarized in Figure SI2 below. The protease treatment reduced H unit measurements by HSQC an average of 32%. The largest reduction was observed in sample 53, where H units were measured to be 7.2% in the enzyme lignin and 4.5% in the protease treated lignin – a reduction of 37.5%. However, despite the reduction in observed H units, all four of these samples still had H unit measurements much higher than the population average and maintained their outlier status. This indicates that these samples may indeed have a high level of authentic H units, though not quite as high as the enzyme lignin measurements would suggest. No spatial or environmental factors impacting these 10 samples were identified. These samples will be further analyzed to investigate the presence of high H units.

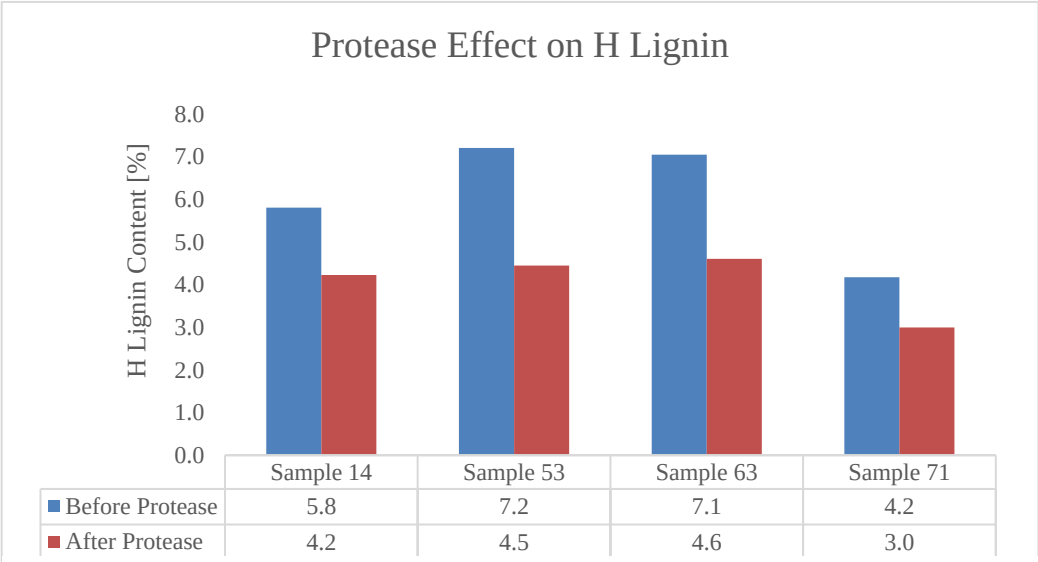


FIGURE 10-5 COMPARISON OF H UNIT LEVELS BEFORE AND AFTER PROTEASE TREATMENT

A multivariate analysis was performed in JMP to elucidate associations between the twelve phenotypes quantified by HSQC NMR. The lower triangular matrix depicts a bivariate XY plot of the two intersecting phenotypes along the diagonal. The upper triangle matrix displays the numerical Pearson correlation coefficient, along with asterisk(s) to indicate the *p*-value associated with correlation probability (** = *p*-value <0.01; * = *p*-value < 0.05; no asterisk = *p*-value >0.05).

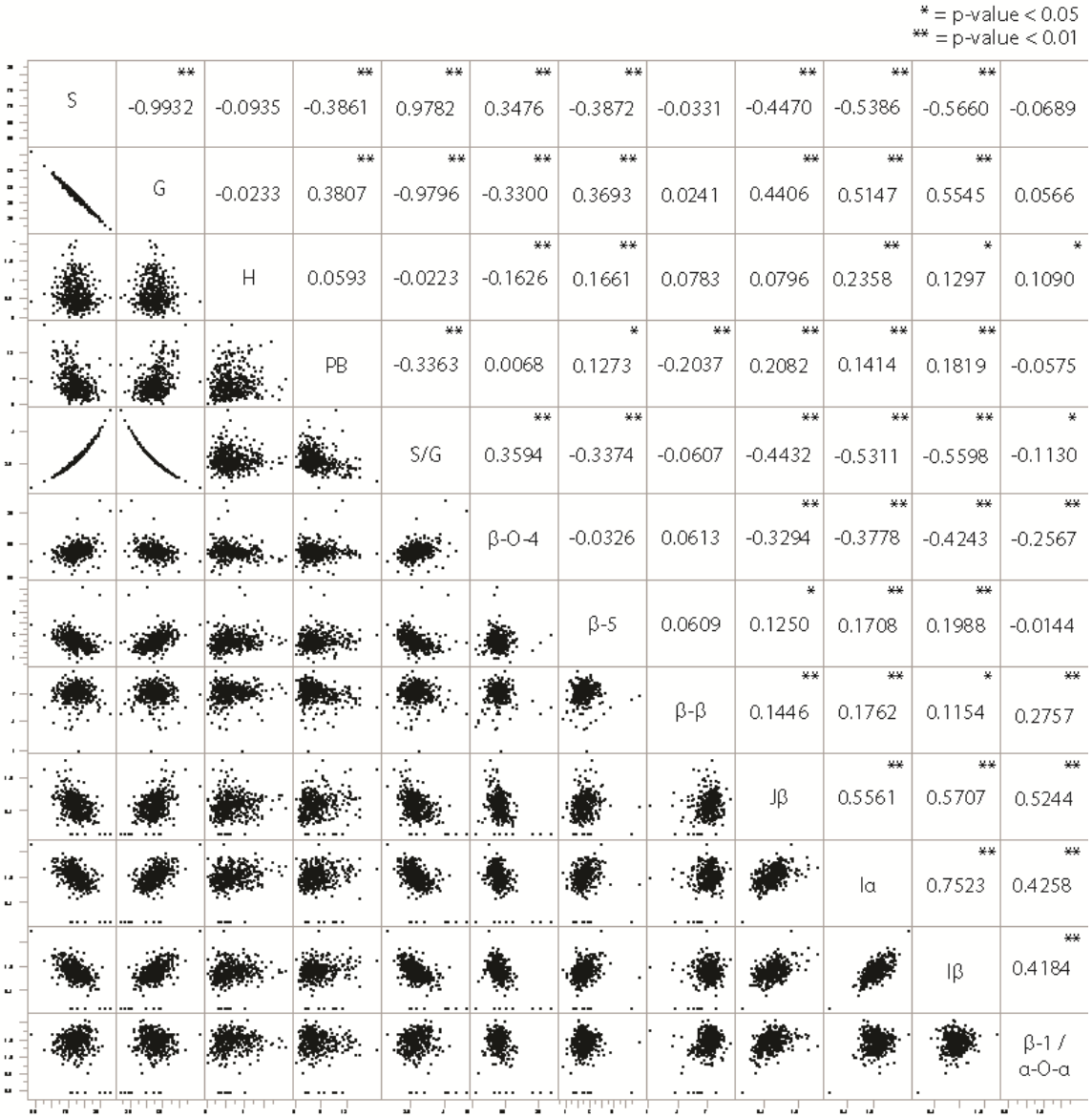
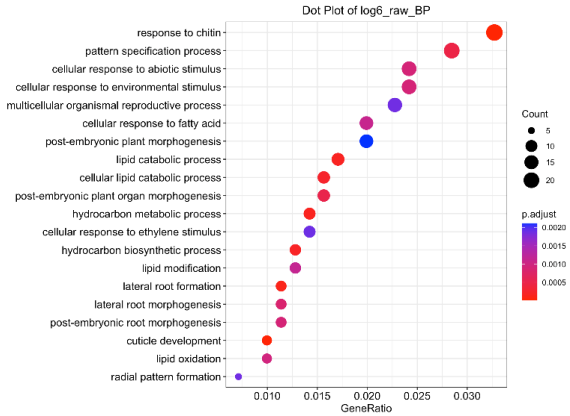
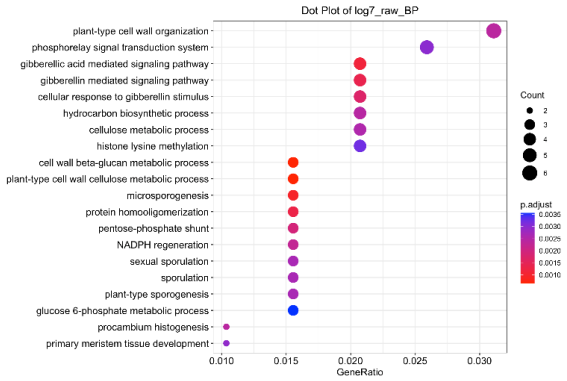
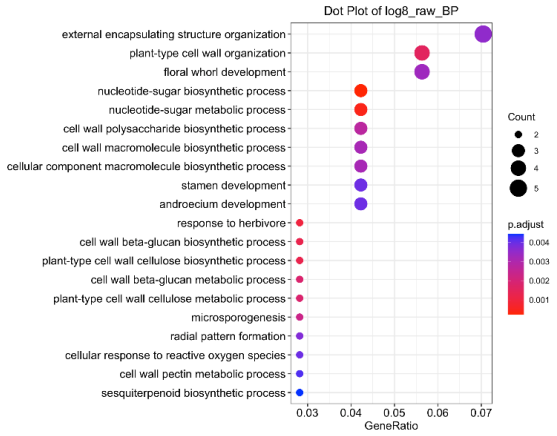


FIGURE 10-6 MULTIVARIATE ANALYSIS OF LIGNIN PHENOTYPES FROM HSQC NMR ANALYSIS

The candidate genes identified by GWAS analyses were examined for gene ontology (GO) terms. A GO enrichment analysis was conducted for each significance level of GWAS results (log6, log7, log8). The enrichment of cell wall (CW) related terms (i.e., number of CW related genes compared to total number of genes identified) increases with significance level. This trend indicates that the GWAS analyses are discriminately identifying CW related causal genes, as opposed to identifying random SNPs from the genome. Enrichment and p-values were obtained via hypergeometric test to compare the significance of the overlap between the number of CW genes detected by the analysis and the theoretical number obtained by random sampling.







Threshold	Total candidates	# of CW related genes (*Populus*)	Enrichment	*p*-value
log6	756	18	1.52	0.032
log7	204	6	1.87	0.043
log8	77	3	2.48	0.033

Threshold	Total candidates	# of CW related genes (*Arabidopsis*)	Enrichment	*p*-value
log6	756	32	2.70	1.85E-07
log7	204	13	4.06	4.91E-06
log8	77	7	5.79	2.87E-05

FIGURE 10-7 GENE ONTOLOGY AND ENRICHMENT RESULTS OF GWAS RESULTS

Appendix C – Supporting information for novel candidate genes for lignin structure revealed by NMR analysis of switchgrass GWAS panel

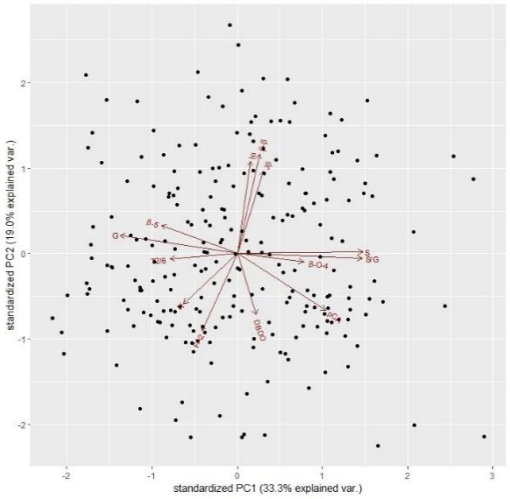


FIGURE 10-8 PRINCIPAL COMPONENT ANALYSIS OF LIGNIN PHENOTYPES

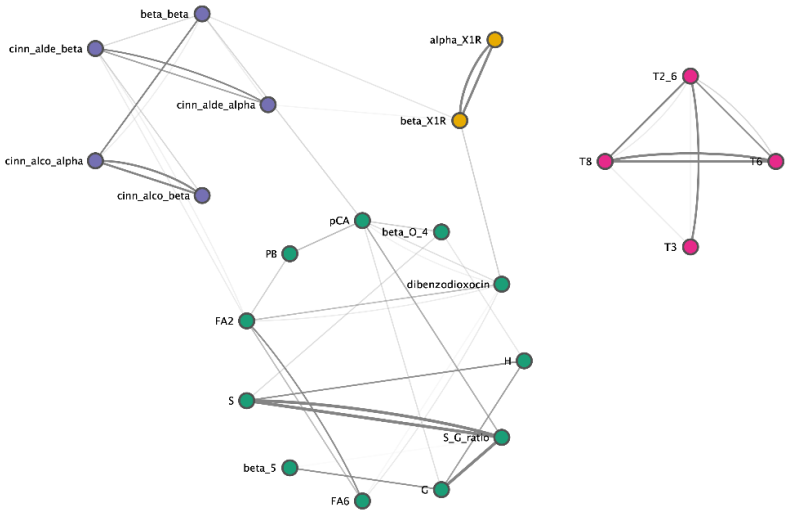


FIGURE 10-9 CLUSTERING OF RELATED LIGNIN PHENOTYPES

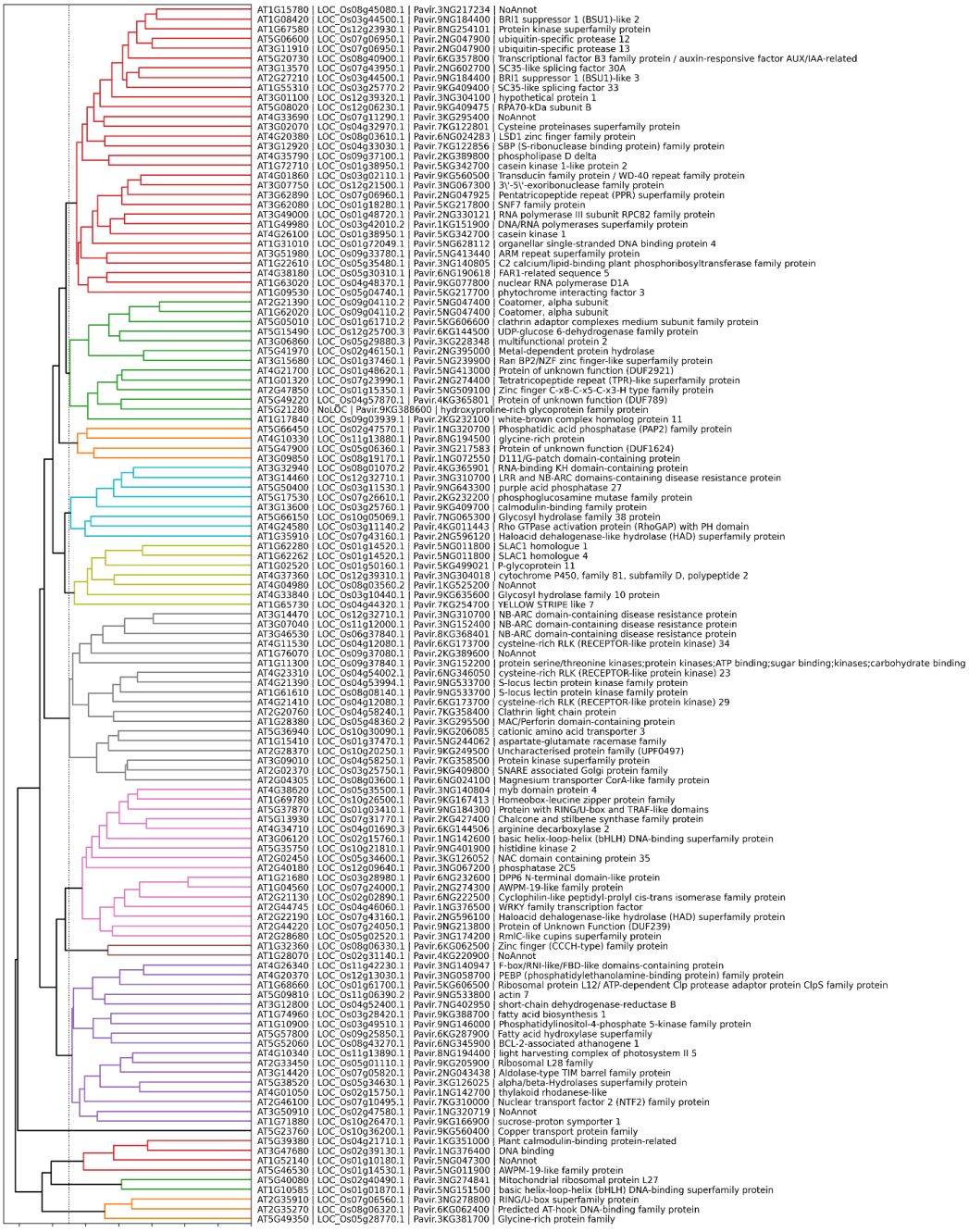


FIGURE 10-10 FUNCTIONAL PARTITIONING OF IDENTIFIED CANDIDATE GENES

Vita:

Nathan Bryant is from Bristol, TN. After completing a B.S. in chemical engineering from Tennessee Technological University in 2011, he worked as a process engineer at Nuclear Fuel Services, Inc. in Erwin, TN. In 2014 he joined a research group at Eastman Chemical Company in Kingsport, TN, where his work focused primarily on improving mechanical properties of cellulose acetate fibers. During this time, he completed a M.S. in reliability and maintainability engineering at the University of Tennessee, where his capstone project involved improving a large compressed air system to prevent process upsets. In 2017 he came to the University of Tennessee to pursue a PhD under the advisement of Dr. Arthur J. Ragauskas, and his research focuses primarily on characterizing lignin and other secondary cell wall components.