

ABSTRACT

BARKER, DAVID KENNETH. Potential of Loblolly Pine (*Pinus taeda* L.) as a Biofuel Feedstock. (Under the direction of Drs. Ross W. Whetten and Steven E. McKeand).

Woody biomass can be an alternative energy source in the Southeastern U.S. due to its wide availability and ability to grow on marginal sites. Loblolly pine, as the most productive forest tree species in the region, is of logical interest for biofuel production. Volume yield was analyzed across nine test series to assess genetic variation in profitability of young-age biomass harvesting and the effects of family on volume production in young plantations. At age eight, the top 10% of families averaged 17.2% more volume growth than the local mixture of first-generation families. Full-sib family mean heritabilities for height and volume were moderate to high at all ages (age eight broad-sense full-sib family mean heritabilities ranged from 0.61 to 0.93), reaffirming that gain in biomass yield is possible through selection. Economic analysis of the data indicated that biomass-only harvests would not be profitable at current prices. If biomass prices increased in response to the development of a bioenergy industry, young-age biomass harvesting may become profitable, but soil expectation values would not be maximized at very young ages.

Loblolly pine biomass presents a challenge for producing ethanol; enzymatic hydrolysis of polysaccharides from softwood pulp typically produces lower yields of fermentable sugars relative to hardwood pulp. Since many chemical and physical wood properties in loblolly pine are subject to genetic control, variation in some of these properties will likely affect the efficiency of ethanol production.

Wood samples were collected from a series of 8-year-old clonal trials in South Carolina and Georgia. Clonal varieties of loblolly pine were divided into groups using a cluster analysis based on near-infrared (NIR) spectra of ground wood samples from multiple individual trees, or ramets, of each of the clonal genotypes. These clusters were used to select a subset of clones for chemical analysis that would encompass most of the natural variation in the clonal population. Wood samples from three pooled ramets of each clone in this set were tested, using enzymatic hydrolysis after a dilute acid or alkaline pretreatment to produce sugar yields for each clone. The lowest yielding treatment, a dilute acid pretreatment followed by enzymatic hydrolysis using 20 filter paper units (FPU) of enzyme, produced an average of 0.21 mg sugar/mg wood. The highest yielding treatment, an alkaline pretreatment followed by enzymatic hydrolysis using 40 FPU as well as mechanical beating, produced 0.52 mg sugar/mg wood. NIR clustering and a number of wet chemistry-determined wood properties were significant predictors of sugar yield. Economic analysis of these sugar yields using a biorefinery process model indicates that ethanol from loblolly biomass is not cost-competitive with gasoline. However, if oil prices go up or bioconversion technology improves, making ethanol from biomass may become a profitable venture.

In order to characterize genetic variation in sugar yield from loblolly pine wood, 300 powdered wood samples from a clonal test series were tested using a high-throughput enzymatic hydrolysis process with a dilute acid pretreatment. The mean sugar yield was 0.21 mg sugar/mg wood, and ranged from 0.10 to 0.29 mg sugar/mg wood. Clonal genetic values indicate that yields can be improved ~4-5% over the mean through the use of the best clones. Clone mean repeatability for sugar yields ranged from 0.29 to 0.44. For NIR-predicted lignin

and cellulose contents, clone mean repeatabilities were 0.81 and 0.78, respectively. A calibration model was developed to predict sugar yields using NIR spectra, but the overall R^2 was relatively low (= 0.30). The ratio of performance to deviation (RPD) was 1.54 for the calibration model and indicates that the model may at best be accurate enough for a rough initial screening of families.

© Copyright 2014 David Kenneth Barker

All Rights Reserved

Potential of Loblolly Pine (*Pinus taeda* L.) as a Biofuel Feedstock

by
David Kenneth Barker

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

Forestry and Environmental Resources

Raleigh, North Carolina

2014

APPROVED BY:

Dr. Sunky Park

Dr. Robert C. Abt

Dr. Fikret T. Isik

Dr. Ross W. Whetten
Committee Co-Chair

Dr. Steven E. McKeand
Committee Co-Chair

DEDICATION

To my wife, Genevieve, and daughter, Audrey.

BIOGRAPHY

David Barker was born on May 1st, 1984 in Auburn, Massachusetts to Daniel and Margaret Barker. He completed high school at Northern Vance High in Henderson, NC in May of 2002, and went on to matriculate at Duke University the following fall. David completed his studies at Duke with a B.A. in Political Science in May 2006. After a brief stint in Washington, D.C. as a case assistant at the law firm Goodwin Procter LLP, he began working with the Duke Forest in Durham, NC as a management assistant in May of 2007.

David began his graduate studies at North Carolina State University in August of 2008 as a master's student. He switched over to a doctoral program after a year. As a graduate student, he worked for and was supported by the NCSU Cooperative Tree Improvement Program. In January of 2013, David took a position with Rayonier, Inc. as a Seed Orchard Specialist.

ACKNOWLEDGMENTS

I would like to thank my wife, Genevieve, and daughter, Audrey, for their support and love during my time as a graduate student. I would also like to thank my parents, Daniel and Margaret Barker, for all their contributions to my success and happiness over the years.

Thanks are also due to my committee members, Drs. Steve McKeand, Ross Whetten, Fikret Isik, Sunkyu Park, and Bob Abt. Their advice on all aspects of this endeavor, including experimental design, analysis, and writing, has been invaluable to my success as a graduate student. I'd also like to thank the NCSU Cooperative Tree Improvement Program, the Biofuels Center of North Carolina, and the Southeast Partnership for Integrated Biomass Supply Systems for funding for my graduate research and coursework. The staff and students of the Tree Improvement Program also deserve recognition for their support and especially for their assistance with field work. My thanks go as well to Plum Creek Timber Company and CellFor Inc. for providing me with wood samples for my experimental work.

TABLE OF CONTENTS

LIST OF TABLES	vii
LIST OF FIGURES	x
INTRODUCTION	1
Literature Cited	11
CHAPTER 1: BIOMASS ACCUMULATION AND HARVESTING IN YOUNG LOBLOLLY PINE PLANTATIONS	15
Abstract	16
Introduction	17
Methods	23
Results	29
Discussion	32
Conclusion	35
Literature Cited	36
Tables and Figures	42
Appendix	54
CHAPTER 2: GENETIC VARIATION IN SUGAR YIELD FROM LOBLOLLY PINE (<i>PINUS TAEDA</i> L.) WOOD: I: ALKALINE AND DILUTE ACID ENZYMATI C HYDROLYSIS OF STEMWOOD SAMPLES	58
Abstract	59
Introduction	60
Methods	69
Results	76
Discussion	79
Conclusion	82
Literature Cited	83
Tables and Figures	89
CHAPTER 3: GENETIC VARIATION IN SUGAR YIELD FROM LOBLOLLY PINE (<i>PINUS TAEDA</i> L.) WOOD: II: HIGH-THROUGHPUT DILUTE ACID ENZYMATI C HYDROLYSIS OF POWDERED CORE SAMPLES	100
Abstract	101
Introduction	102
Methods	104

Results.....	117
Discussion.....	121
Conclusion.....	124
Literature Cited.....	126
Tables and Figures.....	129
Appendix.....	139

CHAPTER 4: MODELING LOBLOLLY PINE IN A BIOREFINERY PROCESS MODEL
.....141

Abstract.....	142
Introduction.....	142
Methods.....	145
Results and Discussion.....	148
Conclusion.....	151
Literature Cited.....	152
Tables and Figures.....	154

LIST OF TABLES

CHAPTER 1:

Table 1: Biomass yields from three studies in the U.S. Southeast. Volumes expressed in dry Mg/ha.....	42
Table 2: Number of test series measured for height and diameter, by age.....	43
Table 3: Summary of series means for volume (A.) and height (B.). Significant differences between series means indicated by letter groupings. The series names starts with the state(s) within which the tests are located, followed by the TPH used for that series. Note that there are two series from Georgia with the same planting density. They have been denoted as ‘a’ and ‘b’.....	44
Table 4: Standard deviation and coefficient of variation for test means for individual tree height (m) and volume (m ³) indicate extent of differences in growth traits within test series. Statistics are for age 8 for all series except for NC_TPH1329, for which age 7 data is presented. ‘Test mean max’ and ‘test mean min’ are maximum and minimum test means for the tests within a particular series.....	45
Table 5: Average genetic parameter estimates across series for height (A.) and volume (B.). Estimates for each series can be found in Appendix A. h^2_F is the narrow-sense family mean heritability, H^2_F is the broad-sense family mean heritability, h^2_i is the narrow-sense individual tree heritability, and H^2_i is the broad-sense individual tree heritability. r_B is the type-‘B’ correlation, which is the ratio of the additive genetic variance to itself plus the variance associated with the additive-by-test interaction.....	46
Table 6: Per-hectare yields for top ~10% of families vs. local seed orchard mix. Units in m ³	47
Table 7: Per-hectare productivity for series mean volumes. Units in m ³	48
Table 8: Per-hectare SEVs for series mean volumes, using 2013 stumpage pricing (\$11/Mg pulpwood, \$17.50/Mg chip-n-saw). Prices in parentheses indicate negative values.....	49
Table 9: Per-hectare SEVs for series mean volumes, using \$17.50 (A.), \$22.00 (B.) and \$33.00 (C.) per Mg stumpage pricing. Prices in parentheses indicate negative values.....	50
Table 10: Per-hectare SEVs for top 10% of families vs. local seed orchard mix (SOM) for each series, using 2013 stumpage pricing. Prices in parentheses indicate negative values.....	51

Appendix A: Genetic parameter estimates by series for height (A.) and volume (B.). h^2_{FS} is the narrow-sense family mean heritability, H^2_{FS} is the broad-sense family mean heritability, h^2_i is the narrow-sense individual tree heritability, and H^2_i is the broad-sense individual tree heritability. r_B is the type-‘B’ correlation, which is the ratio of the additive genetic variance to itself plus the variance associated with the additive-by-test interaction.....54

CHAPTER 2:

Table 1: Sugar yield means, minimums, and maximums by treatment. Units are mg sugar/g wood. Treatment means with the same letter groupings are not significantly different. DA20 is dilute acid pretreatment with 20 FPU enzyme, ALK10 is an alkaline pretreatment with 10 FPU of enzyme, and so forth. ALK10PFI is an alkaline pretreatment with mechanical beating and 10 FPU. N = 17 for all treatments89

Table 2: Summary statistics for lignin and cellulose content predictions computed from NIR spectral data; a. For the full complement of clones (N=151) considered for harvesting. b. For the clones selected for harvest (N=23)90

Table 3: Summary of data for wet chemistry-determined properties91

Table 4: Summary of multiple linear regression models for each of seven treatments. The response variable in each model is the sugar yield produced by a given treatment. R^2 is the model coefficient of determination92

Table 5: Mean sugar yields by cluster for each treatment. DA20 treatment not shown as cluster was not included in the model for that treatment. Different letter groupings indicate significant differences between clusters93

CHAPTER 3:

Table 1: Clone mean repeatabilities (H^2) for MBMS peaks with high loadings. Bolded LRT (those < 3.841) indicates models that are not significantly better than a reduced model fit without a genetic variance component. The peaks in A. (group 1) are predominantly associated with carbohydrates, those in B. (group 2) are predominantly associated with extractives, and those in C. (group 3) are predominantly associated with lignins.....129

Table 2: Results of PLS calibration on sugar yields using NIR spectra130

Table 3: Clone mean repeatabilities and summary statistics for wood components derived from MBMS and NIR spectral data. Bolded LRT (those < 3.841) indicates models that are not significantly better than a reduced model fit without a genetic variance component.....131

Table 4: Summary statistics for measured and predicted sugar yields132

Table 5: Genetic parameters for sugar yield and NIR-predicted lignin and cellulose contents. Bolded LRT (those < 3.841) indicates models that are not significantly better than a reduced model fit without a genetic variance component133

Table 6: Percent differences of clonal BLUPs relative to mean sugar yield134

Table 7: Pearson correlations between sugar yield and growth trait GVs. P-values indicated in parentheses.....135

Appendix A: Summary of 300 samples used for sugar yield characterization with enzymatic hydrolysis. Both sample sets were harvested from the same Plum Creek Timber Company/CellFor Inc. clonal test series139

CHAPTER 4:

Table 1: Mean sugar yields by treatment. Standard deviation is in parentheses154

Table 2: Cost/liter of ethanol required to achieve a 12% IRR, by treatment. DA20 is dilute acid pretreatment with 20 FPU enzyme, ALK10 is an alkaline pretreatment with 10 FPU of enzyme, and so forth. ALK10PFI is an alkaline pretreatment with mechanical beating and 10 FPU.....155

LIST OF FIGURES

CHAPTER 1:

Figure 1: Approximate locations of the nine test series analyzed. The tests within each series were generally located in the same or contiguous counties52

Figure 2: Individual tree mean annual increment for height (A) and volume (B) for the nine study test series53

CHAPTER 2:

Figure 1: The three lignin monomer precursors. These precursors are produced via the shikimic and cinnamate metabolic pathways.....94

Figure 2: Clone mean repeatability (H^2 =broad sense heritability) for NIR spectra at wavelengths between 1100 and 2498 nm. Repeatabilities tend to increase from left to right, though there is a dip between 1900 and 2000 nm. This coincides with an increase in the error variance component95

Figure 3: Dendrogram illustrating the hierarchical structure of the clusters. The black boxes indicate the clusters identified for use in this experiment.....96

Figure 4: Predicted cellulose vs. predicted lignin with clusters indicated97

Figure 5: Predictions vs. measurements for cellulose and lignin for 17 loblolly pine samples derived from chipped bolts98

Figure 6: Lignin and cellulose contents, predicted and determined with wet chemistry, for 31 powdered loblolly pine samples.....99

CHAPTER 3:

Figure 1: Graphical representation of cluster membership for clones. The cluster memberships were based on separate NIR spectra from different sets of wood cores. The thirty boxes in each row represent the 30 clones analyzed in the new cluster analysis. The top row (New), represents the cluster analysis performed in this experiment, and the bottom row (Old), represents the cluster analysis from a previous experiment. Colors indicate different cluster groupings. Little consistency is seen between the groupings produced by the cluster analyses136

Figure 2: Principle component analysis loadings plot of MBMS data. The outliers, which are the peaks with higher loadings, can be seen in three different regions that are separate from

the central cluster of peaks: Group 1 - Peaks in the top left direction of the chart, most are associated with carbohydrates/hemicelluloses; Group 2 - those to the right, most are associated with extractives; and Group 3 - at bottom, most are associated with lignin precursors137

Figure 3: Measured (x-axis) vs. predicted (y-axis) sugar yields. Predictions are from PLS modeling of NIR spectra. R^2 describe the proportion of variation in the laboratory predictions that the PLS model was able to account for using two latent factors138

CHAPTER 4:

Figure 1: Cost/liter of ethanol with enzyme costs varying for each treatment156

Figure 2: Cost/liter of ethanol with energy cost varying for each treatment157

Figure 3: Cost/liter of ethanol with delivered biomass price varying for each treatment.....158

Figure 4: Cost/liter of ethanol comparing best- and worst-yielding clones for each treatment...
.....159

Figure 5: Cost/liter of ethanol comparing cluster three against other clusters for each treatment160

INTRODUCTION

The environmental impacts and the inevitable dwindling of supplies of fossil fuels have been increasingly relevant topics in the scientific as well as public discourse for the past decade. Several primary reasons for this are: 1. concerns about the environmental impact of gaseous emissions from engines, especially potential atmospheric warming due to carbon dioxide; 2. concerns over current prices for transportation fuel; and 3. the question of how plentiful fossil fuels will remain over the next few decades and beyond. Calls for the development of alternatives to nonrenewable fuels are not new; Henry Ford predicted that energy would be obtained from vegetation, and worked on ethanol-fueled vehicles almost 100 years ago (New York Times 1925). In the mid-to-late nineteen-thirties, the Model A included a carburetor that could accept gasoline and/or alcohol-based fuels (Mousdale 2008).

The oil crisis of the 1970's renewed concerns about oil scarcity and long-term fuel resources. The exact amount of fossil fuel remaining is a subject of debate (Brandt 2007, Watkins 2006, and Maugeri 2004). One recent estimate, based on a survey of independent authors, is that there remains approximately 903 giga-barrels (Gb) of conventional oil reserves (for comparison, 31.2 Gb of liquid fuel were consumed in 2008), and the reserves that are currently producing oil will only meet half of the demand by 2023 (Owen et al. 2010). While improvements to extraction methods can help delay the decline in oil production, the declining rate of oil discovery combined with increasing demand indicates that days of cheap and plentiful fossil fuel will not return.

Additionally, national security will benefit from an increase in biofuel production, as many of the most prominent international suppliers of crude oil are located in the often volatile and unstable Middle East. Gupta (2008), in assessing the oil vulnerability of the world's biggest net oil importers, claims that fifty of the world's oil producing countries (including eight of the top ten) have reached maximum oil production levels. As a consequence, "the ability of OPEC to control world oil supplies is likely to increase in the near term." Charles et al. (2007), while acknowledging the uncertainty of when peak oil will occur, affirm the idea that fossil fuels will not be plentiful for the long-term and stress the socio-economic disruption that will occur if alternatives are not found.

The replacement of fossil fuels with renewable ones is necessary for a wide range of energy needs (e.g., vehicle fuels, industrial uses, residential electricity), and even when restricted to vehicular fuel alternatives there are a number of approaches being tested. Electric vehicles and hybrids are a popular area of development, and hydrogen fuel cells are another prominent avenue of research (Johnston et al. 2005). Interest in the use of renewable organic feedstocks to produce liquid fuels for transportation and other combustion-powered applications has been increasing sharply over the last decade. Ethanol has already achieved commercial viability, with prices approaching that of petroleum: estimated world prices of ethanol and petroleum in 2007 were €500 (\$658) and €400 (\$526) per ton, respectively (Soetaert and Vandamme 2009). In Brazil, 15.7% of the total energy expenditure in 2011 was from sugarcane derivatives, and large quantities of ethanol are exported to other parts of the Americas, Europe, and Asia (Barros 2012). In the United States, commercial production

of biofuel is mainly for ethanol, which is primarily derived from the fermentation of sugars from corn kernels. However, there is some doubt as to the net energy efficiency of corn-based ethanol, though improvements in bioconversion plant designs may help this (Liska et al. 2009). Additionally, corn is a staple food crop in many places around the world, and diversion of corn to liquid fuel production is probably contributing to the increase of the worldwide price of corn. This potentially has an out-sized negative impact on impoverished countries, leading to ethics-based criticisms of corn-based ethanol (Pimentel 2003 and Solomon et al. 2007).

An alternative, cellulosic ethanol, has been gaining support in recent years. Indeed, the use of lignocellulosic biomass for energy has been an area of research interest in the U.S. since the 1970's oil crisis (see Smith 1974, SAF 1979, and Zobel 1980). Many potential sources for this biomass are being considered, from food crop residues (e.g., corn stover, which is not eaten by humans) to grasses and trees. Two of the primary advantages in using lignocellulosic biomass are that there are large amounts available and the use of this biomass for fuel does not compete with food markets. Production of ethanol via enzymatic hydrolysis may be competitive with corn ethanol on a production cost basis (Frederick et al. 2008), though enzyme costs are a concern. Additionally, the net energy balance for forest biomass plantations may well be better than for agricultural crops due to the lower level of inputs needed (Zobel 1980). However, the question of net energy balance is complex, as costs for inputs and transport fluctuate and bioconversion technologies improve, so there is still

considerable debate as to whether any biofuels have a positive net energy balance (e.g., Pimentel and Patzek 2005; Farrell et al. 2006).

Several studies have examined the amount of wood residues in North Carolina available per year, and have reached estimates ranging from 2.5 to 3.6 million dry Mg/year, including both softwood and hardwood sources (La Capra Associates 2006 and Galik et al. 2009). Given a conversion rate of 227 liters of ethanol per dry megagram of forest residues (Maness 2008), this is equivalent to 568 to 817 million liters of ethanol per year (equivalent to 379 to 545 million liters of gasoline, since ethanol has approximately $2/3$ the energy content of gasoline). In comparison, the NC Strategic Plan for Biofuels 2007 called for the replacement, by 2017, of approximately 2300 million liters of gasoline per year (i.e., 10% of the total used in North Carolina) with biofuels produced in and from materials grown in the state. Therefore, even following the more optimistic estimate of available residuals, the woody biomass residue available in North Carolina would fall well short of replacing 10% of the total liquid fuels consumed. Also, residual woody biomass is a target for energy generation by investor-owned utility companies, who are required by the N.C. Renewable Energy and Energy Efficiency Portfolio Standard (REPS) to produce 10% of their energy output from renewable sources by 2018 (NC Session Law 2007-397). Additional considerations also include management of harvesting and transportation costs for lower-value residuals, and existing marginal consumers of residuals (firewood, industrial cofiring). All these factors mean that the amount of residual woody biomass actually available for liquid biofuels will likely be significantly lower than the raw estimates.

Other potential feedstocks include energy grasses, municipal solid waste, sugarbeets, sweet potatoes, duckweed, algae, soybeans, sunflowers, canola, barley, flax, and camelina (BCNC 2013). However, other than municipal solid waste and possibly soybeans, none of these are available in quantities capable of producing significant amounts of biofuels. Additionally, while municipal solid waste is available, the development of this technology will likely take many years given the variability of the feedstock and environmental concerns over the burning of waste. As for soybean, it shares a disadvantage with corn in that it is also a food crop, though soybean does have a superior net energy balance to that of corn (Hill et al. 2006). Additional research into yields and ideal cultural practices for candidate crops is needed given that existing woody biomass residues alone will not be able to produce sufficient quantities to meet state- and federally-mandated biofuel targets. Energy grasses are one potential candidate for quick production of significant biomass; it should be possible to get yields of up to 22 dry Mg/hectare/year within the short term, but these energy grasses are still in the developmental phase. This study does not go into the details of feedstocks other than loblolly pine, but the work of Gonzalez et al. (2011b) established that loblolly pine is competitive with other feedstocks (including *Eucalyptus*, grasses, and agricultural biomass) on a delivered cost basis for carbohydrate and BTUs. The challenge, then, is to improve conversion of normally-recalcitrant softwood biomass.

Another economic and social challenge facing the development of a biofuels industry is convincing landowners that biomass feedstocks are worth the investment. While high oil prices will keep interest in biofuels peaked, actual investment will only be successful if the

production technology is cost-effective. Landowners and companies are unlikely to grow non-traditional bioenergy crops without assurances that they will be able to sell it. However, it is also difficult to spur development on the biorefinery side without an established feedstock supply, thus creating a “chicken-and-egg” situation. In addition, the bioconversion facility will only realistically be able to buy from growers within a given radius from the facility, which increases the challenge of growing a sufficient biomass supply. One advantage of woody biomass is that supply chains are already in existence wherever there is an active timber harvesting industry. Additionally, given that there is already incentive to grow wood for pulp and sawtimber, growing biomass for bioenergy can become a part of landowners’ investment strategies without the risk of having a worthless harvest. Provided there is a timber market for a given species, a biomass plantation can be repurposed for sawtimber should the biomass market disappear/fail to develop. A plantation system known as dual-cropping has been proposed for loblolly pine (Scott and Tiarks 2008). This system explicitly aims to produce both biomass and sawtimber though the biomass harvest of a certain portion of the stand at a young age, followed by management of the remaining rows for higher-value sawtimber. In agroforestry, dual cropping with multiple species (often referred to as alley cropping) is already practiced in the U.S. (Williams et al. 1997).

As mentioned above, investment in bioenergy facilities will be hampered by uncertainty about feedstock availability. For a species to be used as a biofuel feedstock, plantations must be able to produce sufficient quantities of biomass to supply a biorefinery. The landowner/grower needs to make a profit, and the delivered cost to the biorefinery must not

be so high that the end product (biofuel) is priced out of the market. If production facilities are established in the numbers necessary to produce significant amounts of biofuel, feedstock prices for woody biomass will likely rise (and, as another source of uncertainty, prices for many potential, but as yet undeveloped bioenergy crops are still unknown). In addition, if demand for woody residuals outstrips supply, producers will start competing (and spending more money) in the pulpwood market, further emphasizing the need to extract as much energy as possible from the feedstock.

Discounted cash flow analysis using net present value (NPV) and/or soil expectation value (SEV) can be used to evaluate the profitability of a forest plantation (Clutter et al. 1983, Gregory 1987). For evaluating the profitability of investment in biorefineries, process models can be developed (Gonzalez 2011a). These models account for the various inputs and outputs of energy and materials for a particular biorefinery and can produce an internal rate of return that indicates the profitability of the biorefinery. While these models cannot be a perfect reflection of reality, they are a useful tool for entrepreneurs to evaluate the feasibility of successfully operating a biorefinery.

Another obstacle to the development of a biofuels industry is that bioconversion techniques are still developing. Burkheisser 2010 and Mousdale 2008 provide a good picture of the current state of biorefining technologies. At the same time, the type and quantity of locally produced feedstock available will guide the selection of conversion technology.

Biochemical techniques focus on hydrolytic processes that retrieve the sugars from the

cellulose and hemicellulose. These techniques can use a variety of chemical and/or enzyme treatments to help loosen the grip of lignin on the carbohydrates and to break down those carbohydrates. Biochemical conversion is typically assisted by chemical and/or enzymatic agents, and fermentation of the sugars can be improved through specially tailored microbes (Mousdale 2008). For cellulosic ethanol, the main difficulty is extracting the valuable sugars from the biomass with as little input and as little sugar loss as possible. The structural properties of these lignocellulosic feedstocks are variable, which means that biochemical conversion techniques will need to be calibrated for specific feedstocks.

Thermochemical techniques for biofuel production involve the application of heat to the biomass. The type of biomass and the conditions of the heating will affect the production process, but typical outputs of thermochemical methods are water, oils, charcoal, and synthesis gases (Overend 2004). The removal of water is essential for producing these products, and the amount of water in the biomass at the beginning of the thermochemical process can have a big effect on the amount of energy consumed during the heating process. Bio-oils can often be used in diesel fuels, and charcoal can be burned for energy (as can raw biomass, such as firewood). The syngas produced can be converted to liquid fuel via Fischer-Tropsch reactions (Mousdale 2008). Thermochemical methods of biofuel production tend to avoid the problems caused by lignin in the biochemical methods, as lignin is an energy-rich polymer that is pyrolyzed or combusted along with the carbohydrate components. Lignin, as a byproduct of any biochemical or thermochemical process, may

also have a higher use as a renewable source for a variety of useful aromatic compounds (Pandey and Kim 2011).

There are scientific issues to be addressed both in the areas of biomass supply and biofuel production in terms of increasing the types and yields of available feedstocks and improving efficiency in fuel production. There are social issues in terms of changing historical land-use patterns by convincing individuals and companies to invest in growing new feedstocks. Most important, however, are economic factors, at least in the long term. If money can be made by growing feedstocks and producing biofuel from them, it is likely to happen. However, development is not always quick to respond to new markets. Information on biomass yields and conversion efficiency is essential to fostering industrial-scale biomass development. The difficulty of successful commercial biorefinery development is highlighted by recent failures: in 2011 North Carolina's first ethanol plant, run by Clean Burn Fuels LLC, entered Chapter 11 bankruptcy, citing an increase in corn prices that made the biorefinery unprofitable (Calhoun 2011). Another high-profile example failure in the Southeast is Range Fuels LLC, whose thermochemical-based bioenergy plant in Soperton, Georgia closed in 2011. In 2012, the site was auctioned off and is being developed by LanzaTech, which is intending to move the project toward commercial-scale production of ethanol (Lane 2013).

Literature Cited

- Barros, S. 2012. Brazil Biofuels Annual Report 2012. USDA Foreign Agricultural Service, GAIN Report Number: BR12013. 28 p. Available online: http://gain.fas.usda.gov/Recent%20GAIN%20Publications/Biofuels%20Annual_Sao%20Paulo%20ATO_Brazil_8-21-2012.pdf, Last accessed 1-9-13.
- BCNC (Biofuels Center of North Carolina). 2013. NC Biofuels Feedstocks. Available online: <http://www.biofuelscenter.org/feedstocks>. Last accessed 1-9-13.
- Brandt, A.R. 2007. Testing Hubbert. *Energy Policy* 35:3074-3088.
- Burkheisser, E.V. ed. 2010. Biological barriers to cellulosic ethanol. Nova Science Publishers Inc., New York. 251 p.
- Calhoun, J. 2011, April 4th. Clean Burn Fuels files for Chapter 11 bankruptcy. *The Fayetteville Observer*. Available online: <http://www.fayobserver.com/articles/2011/04/04/1083859>. Last accessed 8-9-11.
- Charles, M.B., Ryan, R., Ryan, N., Oloruntoba, R., 2007. Public policy and biofuels: The way forward? *Energy Policy* 35:5737–5746.
- Clutter, J.L., Fortson, J.C., Pienaar, L.V., Brister, G.H., and Bailey, R.L. 1983. Timber Management: A Quantitative Approach. John Wiley & Sons, New York. 329p.
- Farrell, A.E., Plevin, R.J., Turner, B.T., Jones, A.D., O’Hare, M., and Kammen, D.M. 2006. Ethanol can contribute to energy and environmental goals. *Science* 311:506-508.
- Frederick, W.J. Jr., S.J. Lien, C.E. Courchene, N.A. DeMartini, A.J. Ragauskas, K. Iisa. 2008. Production of ethanol from carbohydrates from loblolly pine: A technical and economic assessment. *Bioresource Technology* 99(11):5051-5057.
- Galik, C.S., R. Abt, and Y. Wu. 2009. Forest biomass supply in the southeastern United States – implications for industrial roundwood and bioenergy production. *J. For.* 107: 96-77.
- Gonzalez, R., Treasure, T., Phillips, R., Jameel, H., and Saloni, D. 2011a. Economics of cellulosic ethanol production: green liquor pretreatment for softwood and hardwood, greenfield and repurpose scenarios. *Bioresources* 6(3):2551-2567.
- Gonzalez, R., Phillips, R., Saloni, D., Jameel, H., Abt., R., Pirraglia, A., and Wright, J. 2011b. Biomass to energy in the southern United States: supply chain and delivered cost. *Bioresources* 6(3):2954-2976.

Gregory, G.R. 1987. Resource economics for foresters. John Wiley & Sons, New York. 477 p.

Gupta, E., 2008. Oil vulnerability index of oil-importing countries. *Energy Policy* 36:1195–1211.

Hill, J., E. Nelson, D. Tilman, S. Polasky, and D. Tiffany. 2006. Environmental, economic, and energetic costs and benefits of biodiesel and ethanol biofuels. *PNAS* 103:11206-11210.

Johnston, B., Mayo, M.C., Khare, A. 2005. Hydrogen: the energy source for the 21st century. *Technovation* 25:569–585.

La Capra Associates. 2006. Analysis of a renewable portfolio standard for the state of North Carolina. La Capra Associates, Boston, MA. 154 p.

Lane, J. 2013. LanzaTech: Biofuels Digest's 5-Minute Guide (Q1 2013 update). *Biofuels Digest* January 8, 2013. Available online: <http://www.biofuelsdigest.com/bdigest/2013/01/08/lanzatech-biofuels-digests-5-minute-guide-q1-2013-update/>. Last accessed 1-9-13.

Liska, A.J., H.S. Yang, V.R. Bremer, T.J. Klopfenstein, D.T. Walters, G.E. Erickson, and K.G. Cassman. 2009. Improvements in life cycle energy efficiency and greenhouse gas emissions of corn-ethanol. *Journal of Industrial Ecology* 13(1):58-74.

Maness, T. 2008. Forests as a potential feedstock for cellulosic ethanol. US Forest Service. Available online: http://www.safnet.org/fp/documents/feedstock_for_cellulosic_ethanol_08.pdf. Last accessed 8-8-11.

Maugeri, L. 2004. Oil: Never cry wolf—Why the petroleum age is far from over. *Science* 304:1114–1115.

Mousdale, D.M. 2008 *Biofuels: Biotechnology, Chemistry, and Sustainable Development*. Boca Raton : CRC Press. 424 p.

NC Session Law 2007-397. North Carolina General Assembly. Available online: <http://www.ncleg.net/Sessions/2007/Bills/Senate/HTML/S3v6.html>. Last accessed 8-8-11.

New York Times. 1925. Ford Predicts Fuel from Vegetation. Sept. 20, 1925, p. 24.

North Carolina's strategic plan for biofuels leadership. 2007. Available online: <http://www.biofuelscenter.org/>. Last accessed 8-8-11.

- Overend, R.P. 2004. Thermochemical conversion of biomass. In: Shpilrain, E.E. (ed.), Renewable Energy Sources Charged with Energy from the Sun and Originated from Earth-Moon Interaction, in Encyclopedia of Life Support Systems (EOLSS). Developed under the Auspices of the UNESCO, Eolss Publishers, Oxford ,UK.
- Owen, N.A., Inderwildi, O.R., and King., D.A. 2010. The status of conventional world oil reserves – Hype or cause for concern? *Energy Policy* 38:4743-4749.
- Pandey, M.P., and Kim, C.S. 2011. Lignin depolymerization and conversion: a review of thermochemical methods. *Chemical Engineering and Technology* 43(1):29-41.
- Pimentel, D. 2003. Ethanol fuels: Energy balance, economics, and environmental impacts are negative. *Natural Resources Research* 12(2):127-134.
- Pimentel, D., and Patzek, T.W. 2005. Ethanol production using corn, switchgrass, and wood; biodiesel production using soybean and sunflower. *Natural Resources Research* 14(1):65-76.
- SAF (Society of American Foresters). 1979. Forest Biomass As an Energy Source: Study Report of a Task Force of the Society of American Foresters. Society of American Foresters Washington, DC.
- Scott, D.A. and Tiarks, A. 2008. Dual-cropping loblolly pine for biomass energy and conventional wood products. *South. J. Appl. For.* 32(1):33-37.
- Smith, H.D. 1974. Wood as an energy source for the southern pulp and paper industry. Unpublished mimeo report. North Carolina State Cooperative Tree Improvement Program, Raleigh, NC. 18 p.
- Soetaert, W. and Vandamme, E. 2009. Biofuels in Perspective. In: Soetaert, W. and Vandamme, E. (eds.). Biofuels. John Wiley and Sons, Chichester, UK. 2009. 242p
- Solomon, B.D., Barnes, J.R., Halvorsen, K.E. 2007. Grain and cellulosic ethanol: History, economics, and energy policy. *Biomass and Bioenergy* 31:416-425.
- Watkins, G.C. 2006. Oil scarcity: what have the past three decades revealed? *Energy Policy* 34:508–514.
- Williams, P.A., Gordon, A.M., Garret, H.E., and Buck, L. 1997. Agroforestry in North America and its role in farming systems. In: Gordon, A.M. and Newman, S.M. (eds.). Temperate agroforestry systems. CAB International, Wallingford, England. 269p.

Zobel, B. J. 1980. El bosque como fuente de energia (The forest as an energy source). Seminar on Preservation and Harvesting of the Forest as a Renewable Natural Resource in Colombia, Executives Club, Bogota, September 4-6 1980. Carton de Colombia Res. Rept. #61. p. 181-195.

CHAPTER 1

BIOMASS ACCUMULATION AND HARVESTING IN YOUNG LOBLOLLY PINE PLANTATIONS

Abstract

Maximizing biomass growth is essential to the development of a biofuels industry. Loblolly pine (*Pinus taeda* L.) is the most widely planted forest tree species in the Southeast and has the capacity to meet a significant portion of future demand for woody biomass in the region. The use of genetically superior varieties for feedstock will be critical for future biomass-silvicultural systems. In this study, nine test series of full-sib families of loblolly pine were measured annually through eight years of age. Site volume productivities ranged from 66 to 181 m³ per hectare at age eight, with corresponding mean annual increments of 8.3 to 22.6 m³/year. Family volume growth, scaled up to a per-hectare basis, varied greatly within test series, with the top families often producing 1.5-2.0 times the volume of the lowest ranking families. In addition, the top 10% of families averaged 17.2% more volume growth than the local mixture of first-generation families. Full-sib family mean heritabilities for height and volume were moderate to high at all ages (age eight broad-sense full-sib family mean heritabilities ranged from 0.61 to 0.93), reaffirming that gain in volume yield is possible through selection of the best families. Economic analysis indicates that biomass-only harvests would not be profitable at current prices. If biomass prices were to increase in response to the development of a bioenergy industry, biomass-driven plantation schemes may well become profitable, but soil expectation values are not maximized at very young ages.

Introduction

The environmental impacts and the inevitable dwindling of supplies of fossil fuels have been increasingly relevant topics in the scientific as well as public discourse for the past decade. Several primary reasons for this are: 1. concerns about the environmental impact of gaseous emissions from engines, especially potential atmospheric warming due to carbon dioxide; 2. concerns over current prices for transportation fuel; and 3. the question of how plentiful fossil fuels will remain over the next few decades and beyond. The exact amount of fossil fuel remaining is a subject of debate (Brandt 2007, Watkins 2006, and Maugeri 2004). One recent estimate, based on a survey of independent authors, is that there remains approximately 903 giga-barrels (Gb) of conventional oil reserves, and the reserves that are currently producing oil will only meet half of the demand by 2023 (Owen et al. 2010).

The replacement of fossil fuels with renewable ones is necessary for a wide range of energy needs (e.g., vehicle fuel, industrial uses, residential electricity), and even when restricted to vehicular fuel alternatives, there are a number of approaches being evaluated. Electric vehicles and hybrids are popular technologies for reducing liquid fuel consumption, and hydrogen fuel cells are another prominent avenue of research (Johnston et al. 2005). Ethanol has already achieved commercial viability, with prices approaching that of petroleum: estimated world prices of ethanol and petroleum in 2007 were €500 (\$658) and €400 (\$526) per ton, respectively (Soetaert and Vandamme 2009). In Brazil, 15.7% of the total energy expenditure in 2011 was from sugarcane derivatives, and large quantities of ethanol are exported to other parts of the Americas, Europe, and Asia (Barros 2012). In the United

States, commercial production of biofuel is mainly ethanol, which is primarily derived from the fermentation of sugars from corn kernels. However, there is some doubt as to the net energy efficiency of corn-based ethanol, though improvements in bioconversion plant designs may help this (Liska et al. 2009). Additionally, corn is a staple food crop in many places around the world, and diversion of corn to liquid fuel production is probably contributing to the increase of the worldwide price of corn. This potentially has an out-sized negative impact on impoverished countries, leading to ethics-based criticisms of corn-based ethanol (Pimentel 2003 and Solomon et al. 2007).

An alternative, cellulosic ethanol, has been gaining support in recent years. Indeed, the use of lignocellulosic biomass for energy has been an area of research interest in the U.S. since the 1970's oil crisis (see Smith 1974, SAF 1979, and Zobel 1980). Many potential sources for this biomass are being considered, from food crop residues (e.g., corn stover, which is not eaten by humans) to grasses and trees. Two of the primary advantages in using lignocellulosic biomass are that there are large amounts available, and the use of this biomass for fuel does not compete with food markets. Production of ethanol via enzymatic hydrolysis may be competitive with corn ethanol on a production cost basis (Frederick et al. 2008), though enzyme costs are a concern. Additionally, the net energy balance for forest biomass plantations may well be better than for agricultural crops due to the lower level of inputs needed (Zobel 1980). However, the question of net energy balance is complex, as costs in inputs and transports are in flux as technologies improve, and there is still considerable

debate as to whether any biofuels have a positive net energy balance (e.g., Pimentel and Patzek 2005; Farrell et al. 2006).

Several studies have examined the amount of wood residues available per year for biomass feedstock. For example, in North Carolina, estimates range from 2.5 to 3.6 million dry Mg/year, including both softwood and hardwood sources (La Capra Associates 2006 and Galik et al. 2009). Given a conversion rate of 227 liters of ethanol per dry megagram of forest residues (Maness 2008), this is equivalent to 568 to 817 million liters of ethanol per year (equivalent to 379 to 545 million liters of gasoline, because ethanol has approximately 2/3 the energy content of gasoline). In comparison, the NC Strategic Plan for Biofuels 2007 called for the replacement, by 2017, of approximately 2300 million liters of gasoline per year (i.e., 10% of the total used in North Carolina) with biofuels produced in and from materials grown in the state. Therefore, even following the more optimistic estimate of available residuals, the woody biomass residue available in North Carolina would fall well short of replacing 10% of the total liquid fuels consumed. Also, residual woody biomass is a target for energy generation by investor-owned utility companies, who are required by the NC Renewable Energy and Energy Efficiency Portfolio Standard (REPS) to produce 10% of their energy output from renewable sources by 2018 (NC Session Law 2007-397). Additional considerations also include management of harvesting and transportation costs for lower-value residuals, and existing marginal consumers of residuals (firewood, industrial cofiring). All these factors mean that the amount of residual woody biomass actually available for liquid biofuels will likely be significantly lower than the raw estimates.

Another economic and social challenge facing the development of a biofuels industry is convincing landowners that biomass feedstocks are worth the investment. People and companies are unlikely to grow non-traditional bioenergy crops without some certainty that they will be able to sell it. However, it is also difficult to spur development on the production side without an established feedstock supply, thus creating a “chicken-and-egg” situation. For the producers to be interested in planting these feedstocks, they will have to feel that it is economically favorable, and that will require a bioconversion facility. In addition, the bioconversion facility will only realistically be able to buy from growers within a given radius from the facility, which increases the challenge of growing sufficient biomass. One advantage of woody biomass is that supply chains are already in existence wherever there is an active timber harvesting industry. Additionally, given that there is already incentive to grow wood for pulp and sawtimber, growing biomass for bioenergy can become a part of landowners’ investment strategies without the risk of having a worthless harvest. Provided there is a timber market for a given species, a biomass plantation can be repurposed for sawtimber should the biomass market disappear or fail to develop. Plantation systems have been proposed for loblolly pine (e.g. Scott and Tiarks 2008) that explicitly aim to produce both biomass and sawtimber though the biomass harvest of a certain portion of the stand at a young age, followed by management of the remaining rows for higher-value sawtimber.

To meet the needs of any nascent biofuel industry, it is important to outline the potential for and variation in biomass production. Cannell (1989), summarizing data from 303 coniferous

and 204 broadleaved stands from around the world, found that few stands produce more than 15 dry Mg/ha/year, and the average was closer to 5 Mg/ha/year. A few stands had productivity as high as 25 Mg/ha/year (stem and branch wood). For whole tree biomass, most stands produced 18 Mg/ha/year or less, though some stands approached or even surpassed 30 Mg/ha/yr. Cannell reported that there was no obvious difference between the yield distributions of broadleaved and coniferous stands. The stands surveyed covered a wide range of sites and conditions, but were mostly managed plantations.

In the southeastern U.S., *Pinus taeda* (loblolly pine) is the most widely-planted timber species with almost 1 billion seedlings planted annually across millions of hectares (McKeand et al. 2003, Schultz 1997). Loblolly pine's growth rates generally exceed those of other native species, hardwood and softwood alike, over a wide range of site conditions and silvicultural inputs (Cobb et al. 2008; Williams and Gresham 2006; and Samuelson et al. 2001; see Table 1 for summary of biomass yields from these studies). In these three studies, whole tree mean annual increments (MAI) range from 2.7 to 15.2 dry Mg/hectare for loblolly pine. Within each study, growth for loblolly exceeds that of the other species tested for all treatments. It is also reasonable to expect additional improvements in yield: Farnum et al. (1983) estimated that the maximum yield of loblolly pine may be as high as 30 Mg/ha/year under intensive culture and continued genetic improvement. Borders and Bailey (2001) found MAIs of 3.6 to 5.4 cords/acre/year in intensively managed treatments in five 10-12 year-old tests in Georgia. Using a conversion of 2.6 tons/cord (Dicke and Parker 2010), this is equivalent to 21 to 31 Mg/ha/year, surpassing the "maximum" benchmark set by Farnum

et al. (1983). Samuelson et al. (2008) found a total above-ground biomass in an intensively managed treatment (fertilized and irrigated) of 175.2 Mg/ha at age 10 (MAI = 17.5 Mg/ha/yr). Current annual increment in year 10 was calculated to be as high as 25.6 Mg/ha/yr for stemwood). The combination of high yields from loblolly pine and the already established industry surrounding this species makes it a potentially ideal biomass feedstock. Additionally, there is reason to expect continued genetic gains in volume yield, as loblolly pine is just entering its 4th cycle of breeding, and there is still significant variation to be exploited. In comparison, corn has had more than 120 generations of breeding for oil content, and there is evidence that significant genetic variation for this trait still remains in the breeding population (Dudley and Lambert 2004).

Genetic improvement in loblolly pine has traditionally focused on stem quality, disease resistance, and growth trait gains at rotation age (Zobel and Talbert 1984). However, if biomass becomes an important part of the timber market, landowners may desire to grow plantations for the sole purpose of biomass harvests. In this study, we estimated biomass gains from superior families of loblolly pine from a series of progeny tests conducted by the North Carolina State University Cooperative Tree Improvement Program (NCSUCTIP). Because most of the tests were measured annually for the first eight years, they are useful for studying biomass accumulation in young loblolly pine plantations. Traditional tree breeding uses these young-age measurements as proxies to indirectly select for desired traits at rotation age (McKeand 1988, Xiang et al. 2002). For this study, however, we used progeny test data to directly calculate biomass yields at young ages. Based on these biomass yields,

we evaluated biomass development in young plantations and investigated the profitability of strictly-for-biomass harvests.

Methods

The measurement data analyzed for this study is from the Early Diallel Measurement Study (EDMS) established by NCSUCTIP in the southeastern US in the late 1980's and early 1990's as a part of the program's second-generation breeding and testing effort. The original intent of the EDMS study was to examine how heritabilities and age-age genetic correlations change over time in young loblolly plantations (see Xiang et al. 2002 for this analysis). To accomplish this, yearly measurements were made in these tests from ages one to eight.

Additionally, measurements on a small number of these tests were carried out at ages 12, 16, and 20.

The mating design used for the EDMS was a disconnected six-parent half-diallel, comprised of 15 full-sib crosses (selfs and reciprocals were excluded). Two and only occasionally three of these six-parent diallels were planted in each test series, and each test series typically contained four test sites. Parents were not used in multiple diallels. Individual test sites in each series were planted in close proximity to each other, generally in the same or contiguous counties. The test series were planted across the southeastern US, covering the Atlantic and Gulf Coastal Plains and Piedmont. The number of test series for which measurement data are available varies by age (Table 2). Very few breast height diameter measurements were taken before age four.

In order to examine the development of biomass in young stands of loblolly pine, data were only analyzed for stands where planting density in trees-per-hectare (TPH) was known, so that estimates of per-hectare stemwood production could be obtained. Nine such series were identified (Table 2). Of these, eight were measured annually to age eight (one was only measured to age seven). Diameter measurements did not start until age four. A few of the series were measured at older ages: three at age 12, one at age 16, and two at age 20. These nine series are scattered across the Southeast but with little representation in the Piedmont (Figure 1).

Individual tree volumes were calculated using the equation from Shelton et al. (1984):

$$\text{VOL} = 0.00748 + (0.0000353 * \text{DBH}^2 * \text{Height}) \quad (1)$$

Where VOL = stem volume in cubic meters; DBH = diameter at breast height (1.4 meters) measured in centimeters; and Height = total stem height in meters. Summary statistics were calculated using the MEANS procedure of SAS (SAS 9.2, SAS Institute, Cary, NC). The GLM procedure was used to compare series means for height and volume using the Duncan multiple range test.

For all ages (4, 5, 6, 7, 8, 12, 16, and 20), genetic values and heritabilities for height and volume for each of the nine test series were calculated. ASReml 3.0 (Gilmour et al. 2009)

was used to estimate the variance components and make best linear unbiased predictions for volume production for each full sib cross. The following linear model was used:

$$Y_{ijklmn} = \mu + T_i + R(T)_{j(i)} + G_k + G_l + C_{kl} + TP_{ik} + TP_{il} + TC_{ikl} + P_{ijklm} + E_{ijklmn} \quad (2)$$

Where:

Y_{ijklmn} is the n th observation of the j th replication for the kl th cross in the i th test,

μ is the overall mean,

T_i is the i th fixed test effect, $i=1$ to t

$R_{j(i)}$ is the fixed effect of the j th replication within the i th test, $j=1$ to r ,

G_k, G_l is the random effect, or general combining ability (GCA), of the k th and l th parents ~ NID $(0, \sigma_g^2)$,

C_{kl} is the random effect, or specific combining ability (SCA), of the specific cross of the k th and the l th parents ($k \neq l$) ~NID $(0, \sigma_c^2)$,

TG_{ik}, TG_{il} is the random effect of the test by GCA interaction ~NID $(0, \sigma_{tg}^2)$,

TC_{ikl} is the random effect of the test by SCA interaction ~NID $(0, \sigma_{tc}^2)$,

P_{ijklm} is the random effect of plot ~NID $(0, \sigma_p^2)$,

E_{ijklmn} is the random error term ~NID $(0, \sigma_e^2)$.

Heritabilities were calculated as follows (Xiang 2001):

Narrow-sense individual tree heritability:

$$h^2_i = 4\sigma_g^2 / (2\sigma_g^2 + \sigma_c^2 + 2\sigma_{tg}^2 + \sigma_{tc}^2 + \sigma_p^2 + \sigma_e^2) \quad (3)$$

Broad-sense individual tree heritability:

$$H^2_i = 4(\sigma_g^2 + \sigma_c^2) / (2\sigma_g^2 + \sigma_c^2 + 2\sigma_{tg}^2 + \sigma_{tc}^2 + \sigma_p^2 + \sigma_e^2) \quad (4)$$

Narrow-sense full-sib family mean heritability:

$$h^2_{FS} = 2\sigma_g^2 / (2\sigma_g^2 + \sigma_c^2 + 2\sigma_{tg}^2/t + \sigma_{tc}^2/t + \sigma_p^2/tr + \sigma_e^2/trn) \quad (5)$$

Broad-sense full-sib family mean heritability:

$$H^2_{FS} = (2\sigma_g^2 + \sigma_c^2) / (2\sigma_g^2 + \sigma_c^2 + 2\sigma_{tg}^2/t + \sigma_{tc}^2/t + \sigma_p^2/tr + \sigma_e^2/trn) \quad (6)$$

Type-B genetic correlations were calculated in order to assess the stability of family performance across tests within a test series (Yamada 1962, Burdon 1977). These were calculated as:

$$r_B = \sigma_g^2 / (\sigma_g^2 + \sigma_{tg}^2) \quad (7)$$

ASReml calculates BLUPs for the GCA (additive) and SCA (dominance) effects. From these, genetic values (GVs) for volume for each full-sib family were calculated as:

$$GV = \mu + GCA_k + GCA_l + SCA_{kl} \quad (8)$$

In order to quantify biomass differences at the stand level, per-hectare biomass production was calculated using these volume GVs and test series-specific planting density and mortality data. No conclusions on spacing effects will be drawn, as replication of spacings across test series is weak to nonexistent. Additionally, the higher spacing trials tended to be in more northerly sites, confounding any attempts to look at trends in productivity caused by planting density.

A discounted-cash flow analysis was conducted using these stand-level volume estimates in order to assess the feasibility of harvesting at young ages. Stand establishment costs were set at \$740 per hectare (\$300/acre) plus \$55.00/1000 seedlings (\$0.055 per seedling). This

establishment cost would cover site prep (chemical and/or mechanical), planting, and an early fertilization treatment. Seedling costs were based on market prices for 2nd-generation seedlings (see ArborGen 2012). For estimating harvest incomes, stumpage prices follow the South-wide averages from the 2nd quarter of 2013 (Timber-Mart South 2013). The prices are as follows: \$11/green Mg for pulpwood and \$17.50/ green Mg for chip-n-saw (\$10 and \$16/ton, respectively). Additionally, it was of interest to examine revenues with a view toward the potential development of a biomass-for-energy industry, which could result in higher demand and increased prices for pulpwood. In order to examine what revenues might look like under a hypothetical biomass market where demand increases pulpwood prices, scenarios using pricing of \$17.50/green Mg (equivalent to chip-n-saw), \$22/green Mg (double current pulpwood prices) and \$33/green Mg (triple current prices) were also computed. For the purpose of estimating revenues, volume in cubic meters was converted to volume in green megagrams using a conversion factor of 0.836 (or 0.865 for the 16 & 20 year old data) Mg per cubic meter (Amateis et al. 2001, see also Burkhart et al. 1972).

The series for which there is older age data (16- and 20-year old) had minimal sawtimber, so this was not included in the merchandizing. The chip-n-saw proportions of the total volume of these older stands were determined from the diameter distributions. Ages 12 and under are treated as pure biomass (pulp) harvests, as the proportion of more valuable product classes at these ages is negligible. For the scenarios using high biomass prices, all volume was treated as pulpwood/biomass.

Soil expectation value (SEV) was calculated as the net present value of a perpetual, periodic series of forest plantations (Gregory 1987):

$$SEV = \frac{\sum_{t=0}^r R_t(1+i)^{r-t} - \sum_{t=0}^r C_t(1+i)^{r-t}}{(1+i)^{r-1}}, \quad (9)$$

where:

i = interest rate

R_t = revenue in year t (thinning or final harvest)

C_t = cost in year t

r = rotation age.

SEV determines the most profitable harvest age for each scenario. Differences in SEV reflect the differential value of this and all future rotations for a given plantation management scenario. An 8% interest rate was used because the purpose of the analysis is to consider whole-stand pulpwood harvests in young plantations. As this would be a departure from the traditional goals of plantation forestry (i.e., to maximize volume growth with view toward sawtimber production), landowners looking to invest in pine as purely a biomass feedstock would want an assurance of profitability.

To examine the difference in per-hectare biomass that can be achieved from using better genetics, the top four families (~10%) in each series were compared to the local seed orchard mix that was used in that series. These seed orchard mixes were composed of genetic material that would have been available from a typical first-generation seed orchard in that

region. The increase/decrease in SEV relative to this seed orchard mix can be compared across test series. This difference was also calculated as a difference in internal rate of return using a cash flow comprised of year zero establishment costs and biomass revenues.

Results

Height and volume summary statistics for the nine test series were produced (Table 3). The most productive series at age eight was one grown in the coastal plain of South Carolina at a spacing of 1329 trees-per-hectare, with a mean height of 11.7 m and mean per-tree volume of 0.141 m³. The lowest producing series at age eight, located in several counties near the GA-TN border, had a mean height of 7.3 m and per-tree volume of 0.044 m³, and was planted at a density of 2244 trees-per-hectare. Given the large geographic footprint of this study, it is not surprising that there are large, statistically significant site productivity differences starting at early ages. As expected, the more northern test series had a slower growth than southern ones, though there were exceptions to this trend. CVs for test mean heights are generally low, indicating similar height growth across tests within a given series (Table 4). The CVs for volume are somewhat higher, meaning that overall productivity of tests within a given series was more variable than for height. MAI for volume increased throughout the age range of the data (Figure 2). MAI for height increased through about age six or seven, then appeared to hold steady or taper off thereafter.

Heritabilities and type-B genetic correlations for height and volume were computed for each test series (see averages in table 5; genetic parameters by series are in Appendix A).

Average heritabilities for height and volume are within expected ranges: across ages, individual narrow- and broad-sense heritabilities range from 0.2 to 0.4, and family mean narrow- and broad-sense heritabilities range from 0.7 to 0.9. Heritabilities generally increased through age 12, and heritabilities for height were slightly higher than for volume. Type-B genetic correlations, which are a measure of the stability of family performance across test sites, were moderate to high for both height and volume across ages and test series. For height and volume, age eight type-B genetic correlations averaged 0.88 and 0.79, respectively.

Family GVs were scaled up to a per-hectare basis using planting density and mortality information. Volume yields from the top ~10% of families (average of four families) were compared to the local seed orchard mix (Table 6). Generally, the top genetic entries compared favorably with the SOM, averaging 21% more biomass at age eight. However, in two of the test series, the top four full-sib families did not outperform the SOM. In one of these, the difference was negligible (1%), and in the other, the SOM outperformed the top four full-sibs by 21%. The average percent difference in yield between the top 10% and the SOM increased with age, though this was not seen in all test series.

Overall series mean volume productions were scaled up to a per-hectare basis (Table 7) and subjected to discounted cash-flow analysis under several pricing scenarios. SEVs for harvesting of stands at current (2013) stumpage prices (Table 8) indicate that such harvesting is generally not profitable at an 8% discount rate. The highest yielding site, SC_TPH1329,

only becomes profitable at age eight, with an SEV of \$191/ha. This site was also the only one with 16 and 20-year data; these SEVs are much higher (\$1,634/ha and \$1,348/ha, respectively) due to continued biomass development as well as the merchandizing of part of the stand at chip-n-saw prices. The other series for which age 20 data are available, VA_TPH2244, is a northern, slower growing series, and just barely turns a profit at this age (SEV = \$3). All other series were non-profitable for all ages for which there are data, with negative SEVs ranging from hundreds to nearly three thousand dollars per hectare.

For the scenario where the pulpwood stumpage price is set to that of chip-n-saw (\$17.50/Mg), profitability is seen at age 12 for all three test series that have volume yields at this age (Table 9a). In addition, two series have positive SEVs at age eight, with one of those also being profitable at ages six and seven. As biomass prices are raised even higher, to double and triple current pulpwood prices (Tables 9b and 9c, respectively), all series become profitable by age eight or before. However, SEVs generally rise across the range of ages, and with one exception, the highest SEV for each series is at the oldest age for which there is data. The exception is the series for which 16- and 20-year old data were available; this series' SEV is highest at age 16 under all pricing scenarios.

A discounted cash-flow analysis was also done for the top four families (averaged) as well as the SOM for each series at current stumpage prices (Table 10). The pattern here is similar to that of the SEVs for the overall stands means (Table 8): Only one test series shows a profit at a young age (SC_TPH1329 at age eight). For this series, age eight SEV is \$377/ha if the top

families are used, compared to an SEV of -\$146/ha for the SOM. Across test series at age eight, the magnitude of the SEV difference between the top full-sib families and the SOM ranged from -\$329 to \$523 per hectare, with an average of \$158/hectare benefit from planting better families. This corresponds in an improvement in the internal rate of return from 0% to 2%. The average SEV improvement from using the top four full-sib families increases as the trees get older because the absolute volume differential between the top families and the SOM increased with age (Table 6).

Discussion

The best site from the progeny test data achieved a biomass production of 181 m³ by age 8 (Table 7), equivalent to an MAI of 23 m³/ha/year (18 Mg/ha/yr). The lowest yielding test had 66 m³ at year eight, and an MAI of 8 m³/ha/year (7 Mg/ha/yr). Assuming dry weight is about half of the green weight, these MAI would be 9 and 3.5 dry Mg/ha/yr for the highest and lowest yielding sites, respectively. These yields are moderate to high compared to growth-and-yield studies which applied a combination of treatments from control (no additions) to fertilization plus irrigation (Table 1). As such, these productivities represent a cross section of what might be found on an operational scale. While the range of the data is insufficient to pinpoint the age at which SEVs are maximized, the general trend indicates that this occurs beyond age eight. This was the case even in the scenario using a pulpwood stumpage price that was three times the current price.

While biomass harvesting was generally not profitable at young ages, using the top families will improve profitability (or at least decrease the amount of money lost). At age eight, using the top families resulted in an average improvement in SEV of \$158/ha. As an added benefit, the difference in absolute volume between better and average genetics increases with age. This is reflected in the increasing differential of SEVs over time between top families and the SOM (Table 10).

Two series did not produce crosses that were superior to the local 1st-generation SOM. However, this does not call into question the benefit of using better families; this study was done using progeny test data, which is typically used to screen families. Those families that do not perform well will not be selected for inclusion in seed orchards. The heritabilities for growth traits indicate that selection of the best families will lead to higher volume yields. This has been demonstrated throughout the history of the tree improvement in the Southeastern US. McKeand et al. (2003) reported that improved families averaged 10 to 30 percent higher volume over unimproved stock, at a time when most seedlings were coming from 1st- and 2nd-generation seed orchards. In comparison, this study found an average volume increase of 21% between the top four families and the 1st-cycle SOM in each test series. High type-B correlations show that there was good stability of family ranks between tests, which means that the top families tend to be good as long as they are planted within the correct deployment region.

Site productivity varied widely in these tests across the southeastern US. Economic analysis indicates that harvesting for biomass from these stands at young ages (ages four-eight) is not desirable. Current pulpwood prices resulted in taking a loss on the investment for all but one series, which turned a small profit at age eight. Even if the stumpage price for pulpwood/biomass increases two- or three-fold, early harvesting, while more profitable at these prices, will not maximize SEV for a stand. Biomass production in a young stand increases quickly enough that SEV increases beyond age eight.

Comparison of the top ~10% of full-sib families (based on stand-level yields calculated from family GVs) to the local SOM indicates that selection of the best families can make a significant difference on harvest value of a stand, and that this value differential increases as the stand ages. However, use of just top families in a stand generally did not change the general outcome of the economic analysis; early harvests are generally not profitable and do not maximize SEV. This study used data for full-sib families, but the results are applicable to open-pollinated half-sib families as well: while differences in volume yields are significant at young ages for half-sib families (Svensson et al. 1999), they will generally be out-performed by good control-pollinated full-sibs. As such, differences in volume production between half-sib families would not improve the overall economic prospects for early biomass harvesting.

Conclusion

Examination of the variation in biomass growth in the EDMS progeny tests across sites and families indicates the importance of selecting families optimized for biomass growth. While a landowner may not have the flexibility to change his/her landbase, poor site conditions can be ameliorated through good silvicultural practices. However, failure to maximize productivity due to the use of average-to-poor families cannot be corrected until the following rotation. Fortunately, genetically improved loblolly pine seedlings are available for sites throughout the U.S. Southeast. If the best genetics are planted on these best sites and are combined with optimum silvicultural practices, MAI can be pushed even higher, increasing returns for landowners.

Literature Cited

- ArborGen. 2012. Product Catalog: 2012-2013 Planting Season. Available online: <http://www.arborgen.us/uploads/catalog/Product%20catalog%206%20page%20final.pdf>. Last accessed 10-26-13.
- Amateis, R.L., and H.E. Burkhart. 1985. Site index curves for loblolly pine plantations on cut-over site-prepared lands. *Southern Journal of Applied Forestry* 9(3):166-169.
- Amateis, R.L., H.E. Burkhart, H.L. Allen, and C. Montes. 2001. FASTLOB: A Stand-Level Growth and Yield Model for Fertilized and Thinned Loblolly Pine Plantations. FASTLOB user manual available at <http://frec.vt.edu/ForestModelingResearchCooperative/documents/models/fastlob.pdf>; last accessed April 27, 2011.
- Amateis, R.L., H.L. Allen, C.R. Montes, and T.R. Fox. 2005. Overview of the Forest Nutrition Cooperative's Silvicultural Decision Support System. Forest Nutrition Cooperative, Virginia Tech and NC State Universities: Available at <http://www.forestnutrition.org/pdf/LobDSS.pdf>; last accessed April 27, 2011.
- Barros, S. 2012. Brazil Biofuels Annual Report 2012. USDA Foreign Agricultural Service, GAIN Report Number: BR12013. 28 p. Available online: http://gain.fas.usda.gov/Recent%20GAIN%20Publications/Biofuels%20Annual_Sao%20Paulo%20ATO_Brazil_8-21-2012.pdf, Last accessed 1-9-13.
- Borders, B.E., and R.L. Bailey. 2001. Loblolly pine – pushing the limits of growth. *Southern Journal of Applied Forestry* 25(2):69-74.
- Brandt, A.R. 2007. Testing Hubbert. *Energy Policy* 35:3074-3088.
- Buford, M.A. and H.E. Burkhart. 1987. Genetic improvement effects on growth and yield of loblolly pine plantations. *Forest Science* 33(3):707-724.
- Burdon, R.D. 1977. Genetic correlation as a concept for studying genotype-environment interaction in forest tree breeding. *Silvae Genetica* 26:168-175.
- Burkhart, H.E., R.C. Parker, M.R. Strub and R.G. Oderwald. 1972. Yields of old-field loblolly pine plantations. Division of Forestry and Wildlife Resources, Virginia Polytechnic and State University, Pub. FWS-3-72. 51p.
- Cannell, M.G.R. 1989. Physiological basis of wood production: a review. *Scand. J. For. Res.* 4:459-490.

- Clark, A., R.H. McAlister, J.R. Saucier, and K. Reitter. 1996. Effect of rotation age on lumber grade, yield, and strength of unthinned loblolly pine. *Forest Products Journal* 46(1):63-68.
- Cobb, W.R., Will, R.E., Daniels, R.F., Jacobson, M.A. 2008. Aboveground biomass and nitrogen in four short-rotation woody crop species growing with different water and nutrient availabilities. *Forest Ecology and Management* 255:4032-4039.
- Dicke, S. and R. Parker. 2010. Pine timber volume-to-weight conversions. Mississippi State University Extension Service, Publication 2244. Available online: <http://msucares.com/pubs/publications/p2244.pdf>. Last accessed 9-22-13.
- Dudley, J. W. and R.J. Lambert. 2004. 100 generations of selection for oil and protein in corn. *Plant Breeding Reviews* 24(1):79-110.
- EISA (Energy Independence and Security Act) of 2007. 110th Congress. H.R. 6, 2007.
- EPA. 2009. Renewable Portfolio Standards Fact Sheet. Available online at http://www.epa.gov/chp/state-policy/renewable_fs.html; last accessed April 27, 2011.
- Farnum, P., Timmis, R., and Kulp, J.L. 1983. Biotechnology of forest yield. *Science* 219(4585):694-702.
- Farrell, A.E., Plevin, R.J., Turner, B.T., Jones, A.D., O'Hare, M., and Kammen, D.M. 2006. Ethanol can contribute to energy and environmental goals. *Science* 311:506-508.
- Frederick, W.J. Jr., S.J. Lien, C.E. Courchene, N.A. DeMartini, A.J. Ragauskas, K. Iisa. 2008. Production of ethanol from carbohydrates from loblolly pine: A technical and economic assessment. *Bioresource Technology* 99(11):5051-5057.
- Galik, C.S., R. Abt, and Y. Wu. 2009. Forest biomass supply in the southeastern United States – implications for industrial roundwood and bioenergy production. *J. For.* 107: 96-77.
- Gillespie, A.R., Jose, S., Mengel, D.B., Hoover, W.L., Pope, P.E., Seifert, J.R., Biehle, D.J., Stall, T., and Benjamin, T.J. 2000. Defining competition vectors in a temperate alley cropping system in the Midwestern USA: 1. Production physiology. *Agroforestry Systems* 48(1):25-40.
- Gilmour, A.R., Gogel, B.J., Cullis, B.R., and Thompson, R. 2009. ASReml User Guide Release 3.0. VSN International Ltd, Hemel Hempstead, HP1 1ES, UK. 398p.
- Gregory, G.R. 1987. Resource economics for foresters. John Wiley & Sons, New York. 477 p.

- Harrison, W. M. and Borders, B.E. 1996. Yield prediction and growth projection for site-prepared loblolly pine plantations in the Carolinas, Georgia, Alabama and Florida. PMRC Technical Report 1996-1. School of Forest Resources, The University of Georgia, Athens, GA. 49 pp.
- Johnston, B., Mayo, M.C., Khare, A. 2005. Hydrogen: the energy source for the 21st century. *Technovation* 25:569–585.
- La Capra Associates. 2006. Analysis of a renewable portfolio standard for the state of North Carolina. La Capra Associates, Boston, MA. 154 p.
- Liska, A.J., H.S. Yang, V.R. Bremer, T.J. Klopfenstein, D.T. Walters, G.E. Erickson, and K.G. Cassman. 2009. Improvements in life cycle energy efficiency and greenhouse gas emissions of corn-ethanol. *Journal of Industrial Ecology* 13(1):58-74.
- Maness, T. 2008. Forests as a potential feedstock for cellulosic ethanol. US Forest Service. Available online:
http://www.safnet.org/fp/documents/feedstock_for_cellulosic_ethanol_08.pdf. Last accessed 8-8-11.
- Maugeri, L. 2004. Oil: Never cry wolf—Why the petroleum age is far from over. *Science* 304:1114–1115.
- McGraw, R.L. Stamps, W.T., Houx, J.H, and Linit, M.J. 2008. Yield, maturation, and forage quality of alfalfa in a black walnut alley-cropping practice. *Agroforestry Systems* 74(2):155-161.
- McKeand, S.E. 1988. Optimum age for family selection for growth in genetic tests of loblolly pine. *Forest Science*. 34(2): 400-411.
- McKeand, S.E., R.P. Crook, and H.L. Allen. 1997. Genotypic stability effects on predicted family responses to silvicultural treatments in loblolly pine. *Southern Journal of Applied Forestry* 21(2):84-89.
- McKeand, S., Mullin, T., Byram, T. and White, T. 2003. Deployment of genetically improved loblolly and slash pines in the South. *Journal of Forestry* 101(3):32-37.
- McKeand, S.E., R.C. Abt, H.L. Allen, B. Li, and G.P. Catts. 2006. What are the best loblolly pine genotypes worth to landowners? *Journal of Forestry* 104(7):352-358.
- NC Session Law 2007-397. North Carolina General Assembly. Available online:
<http://www.ncleg.net/Sessions/2007/Bills/Senate/HTML/S3v6.html>. Last accessed 8-8-11.

North Carolina's strategic plan for biofuels leadership. 2007. Available online: <http://www.biofuelscenter.org/>. Last accessed 8-8-11.

Owen, N.A., Inderwildi, O.R., and King., D.A. 2010. The status of conventional world oil reserves – Hype or cause for concern? *Energy Policy* 38:4743-4749.

Pimentel, D. 2003. Ethanol fuels: Energy balance, economics, and environmental impacts are negative. *Natural Resources Research* 12(2):127-134.

Pimentel, D., and Patzek, T.W. 2005. Ethanol production using corn, switchgrass, and wood; biodiesel production using soybean and sunflower. *Natural Resources Research* 14(1):65-76.

SAF (Society of American Foresters). 1979. Forest Biomass As an Energy Source: Study Report of a Task Force of the Society of American Foresters. Society of American Foresters Washington, DC.

Samuelson, L., Stokes, T., Cooksey, T., and McLemore, P. III. 2001. Production efficiency of loblolly pine and sweetgum in response to four years of intensive management. *Tree Physiology* 21:369-376.

Samuelson, L., Butnor, J., Maier, C., Stokes, T.A., Johnson, K., and M. Kane. 2008. Growth and physiology of loblolly pine in response to long-term resource management: defining growth potential in the southern United States. *Can. J. For. Res.* 38:721-732.

SAS Institute Inc. 2008. SAS/STAT®9.2 User's Guide. Chp. 66: The PLS Procedure. SAS Institute Inc., Cary, NC. 60p.

Schultz, P. 1997. Loblolly Pine: The Ecology and Culture of the Loblolly Pine (*Pinus taeda* L.). Agriculture Handbook 713. US Department of Agriculture, Forest Service. 514 p.

Scott, D.A. and Tiarks, A. 2008. Dual-cropping loblolly pine for biomass energy and conventional wood products. *South. J. Appl. For.* 32(1):33-37.

Shelton, M.G., L.E. Nelson, and G.L. Switzer. 1984. The weight, volume and nutrient status of plantation-grown loblolly pine trees in the interior flatwoods of Mississippi. Miss. State Tech. Bull. 121. Miss. State Agric. & For. Exp. Sta. 27 p.

Smith, H.D. 1974. Wood as an energy source for the southern pulp and paper industry. Unpublished mimeo report. North Carolina State Cooperative Tree Improvement Program, Raleigh, NC. 18 p.

- Soetaert, W. and Vandamme, E. 2009. Biofuels in Perspective. In: Soetaert, W. and Vandamme, E. (eds.). Biofuels. John Wiley and Sons, Chichester, UK. 2009. 242p
- Solomon, B.D., Barnes, J.R., Halvorsen, K.E. 2007. Grain and cellulosic ethanol: History, economics, and energy policy. *Biomass and Bioenergy* 31:416-425.
- South, D.B. 2003. “Correct” planting density for loblolly pine depends on your objectives and who you ask. *Forest Landowner Manual* 34:46-51.
- Svensson, J.C., McKeand, S.E., Allen, H.L., and Campbell, R.G. 1999. Genetic variation in height and volume of loblolly pine open-pollinated families during canopy closure. *Silvae Genetica* 48(3-4):204-208.
- Timbermart-South. 2010. Market News Quarterly - 4th Quarter, 2010. Available online at http://www.timbermart-south.com/pdf/4Q2010news_abridged.pdf. Last accessed 4-28-2011.
- Timbermart-South. 2013. The TMS News Quarterly 3Q 2013. Available online at <http://www.timbermart-south.com/prices.html> Last accessed 10-25-13.
- Wanvestraut, R.H., Jose, S., Ramachandran, P.K., and Brecke, B.J. 2004. Competition for water in a pecan (*Carya illinoensis* K. Koch) – cotton (*Gossypium hirsutum* L.) alley cropping system in the southern United States. *Agroforestry Systems* 60(2):167-179.
- Watkins, G.C. 2006. Oil scarcity: what have the past three decades revealed? *Energy Policy* 34:508–514.
- Williams, P.A., Gordon, A.M., Garret, H.E., and Buck, L. 1997. Agroforestry in North America and its role in farming systems. In: Gordon, A.M. and Newman, S.M. (eds.). Temperate agroforestry systems. CAB International, Wallingford, England. 269p.
- Williams, T.M., Gresham, C.A. 2006. Biomass accumulation in rapidly growing loblolly pine and sweetgum. *Biomass and Bioenergy* 30:370-377.
- Xiang, B. 2001. Genetic analysis of diallel tests of loblolly pine (*Pinus taeda* L.). Ph.D. Dissertation, North Carolina State University, Raleigh, NC. 143 p.
- Xiang, B., B. Li, and F. Isik. 2002. Time trend of genetic parameters in growth traits of *Pinus taeda* L. *Silvae Genetica* 52(3-4):114-121.
- Yamada, Y. 1962. Genotype by environment interaction and genetic correlation of the same trait under different environments. *Japan. J. Genetics* 37:498-509.

Zhao, D., Kane, M., and Borders, B.E. 2011. Growth responses to planting density and management intensity in loblolly pine plantations in the southeastern USA Lower Coastal Plain. *Annals of Forest Science* 68:625-635.

Zobel, B. J. 1980. El bosque como fuente de energia (The forest as an energy source). Seminar on Preservation and Harvesting of the Forest as a Renewable Natural Resource in Colombia, Executives Club, Bogota, September 4-6 1980. Carton de Colombia Res. Rept. #61. p. 181-195.

Zobel, B. and J. Talbert. 1984. Applied forest tree improvement. John Wiley & Sons, New York. 505 p.

Zobel, B. and J.P. van Buijtenen. 1989. Wood variation: its causes and control. Springer-Verlag, Berlin, Germany. 363 p.

Tables and Figures

Table 1. Biomass yields from three studies in the U.S. Southeast. Volumes expressed in dry Mg/ha.

<i>Species</i>	<i>Age</i>	<i>Mg/ha (stem)</i>	<i>Mg/ha/yr (stem)</i>	<i>Mg/ha (whole tree)</i>	<i>Mg/ha/yr (whole tree)</i>	<i>Source</i>
<i>Liquidambar styraciflua</i>	6	2-28	0.3-4.7	6-48	1.0-8.0	Cobb et al. 2008
<i>Liquidambar styraciflua</i>	6	7-39	1.2-6.5	12-63	2.0-10.5	Williams and Gresham 2006
<i>Liquidambar styraciflua</i>	4	2-13	0.3-2.2	4-24	0.7-4	Samuelson et al. 2001
<i>Pinus elliottii</i>	6	14-27	2.3-4.5	31-59	5.2-9.8	Cobb et al. 2008
<i>Pinus taeda</i>	6	24-46	4.0-7.7	44-79	7.3-13.2	Cobb et al. 2008
<i>Pinus taeda</i>	6	25-48	4.2-8.0	48-91	8.0-15.2	Williams and Gresham 2006
<i>Pinus taeda</i>	4	8-19	1.3-3.2	16-41	2.7-6.8	Samuelson et al. 2001
<i>Platanus occidentalis</i>	6	4-37	0.6-6.2	7-54	1.1-9.0	Cobb et al. 2008

Table 2. Number of test series measured for height and diameter, by age.

<i>Age</i>	<i>Height</i>	<i>Diameter</i>	<i># Test Series with Known TPA</i>
1	23	1	9
2	26	1	9
3	26	3	9
4	25	23	8
5	26	25	9
6	26	26	9
7	23	22	9
8	22	22	8
12	10	10	3
16	2	2	1
20	2	2	2

Table 3. Summary of series means for volume (A.) and height (B.). Significant differences between series means indicated by letter groupings. The series names starts with the state(s) within which the tests are located, followed by the TPH used for that series. Note that there are two series from Georgia with the same planting density. They have been denoted as ‘a’ and ‘b’.

A. Volume

Series	Age							
	4	5	6	7	8	12	16	20
SC_TPH1329	0.024 a	0.047 a	0.078 a	0.107 a	0.141 a		0.414	0.602 a
GA_TPH1347b	0.014 b	0.027 b	0.045 b	0.067 b	0.085 b			
NC_TPH1329	0.010 f	0.016 e	0.036 c	0.047 c				
GA_TPH1347a	0.013 c	0.022 c	0.031 d	0.043 d	0.060 c	0.138 a		
NC-VA_TPH1994	0.010 e	0.016 d	0.028 e	0.042 d	0.057 d			
FL-GA_TPH1604	0.011 d	0.017 d	0.028 e	0.040 e	0.056 e	0.128 b		
VA_TPH1537	0.010 f	0.014 f	0.022 f	0.033 f	0.045 f			0.239 b
GA-TN_TPH2244		0.014 f	0.020 g	0.032 g	0.044 fg			
VA_TPH1794	0.009 g	0.013 g	0.020 h	0.031 g	0.045 g	0.101 c		

B. Height

Series	Age										
	1	2	3	4	5	6	7	8	12	16	20
SC_TPH1329	0.7 a	1.9 a	3.6 a	5.2 a	6.9 a	8.8 a	10.3 a	11.7 a		20.7	24.3 a
GA_TPH1347b	0.4 d	1.3 c	2.7 c	4.0 b	5.4 b	6.8 b	7.8 b	8.7 b			
NC-VA_TPH1994	0.3 g	1.0 g	2.1 d	3.3 d	4.5 d	6.0 c	7.2 c	8.3 d			
FL-GA_TPH1604	0.4 e	0.9 h	1.8 g	3.1 e	4.5 e	5.9 e	7.2 c	8.6 c	12.9 a		
NC_TPH1329	0.3 h	0.7 i	1.6 h	2.8 h	4.2 f	6.0 d	7.0 d				
GA_TPH1347a	1.0 a	1.8 b	2.9 b	3.8 c	4.9 c	5.7 f	6.7 e	7.7 e	12.0 b		
VA_TPH1794	0.4 d	1.0 f	1.8 g	3.0 f	4.1 f	5.2 h	6.4 f	7.7 e	11.6 c		
GA-TN_TPH2244	0.5 c	1.1 d	2.0 e		4.0 g	5.1 i	6.4 f	7.3 f			
VA_TPH1537	0.3 f	1.0 e	2.0 f	2.9 g	4.1 f	5.3 g	6.3 g	7.3 f			16.6b

Table 4. Standard deviation and coefficient of variation for test means for individual tree height (m) and volume (m³) indicate extent of differences in growth traits within test series. Statistics are for age 8 for all series except for NC_TPH1329, for which age 7 data is presented. ‘Test mean max’ and ‘test mean min’ are maximum and minimum test means for the tests within a particular series.

Series	# Tests	Height				Volume			
		Std. Dev. of Test Means	Test Mean CV	Test Mean Max	Test Mean Min	Std. Dev. of Test Means	Test Mean CV	Test Mean Max	Test Mean Min
SC_TPH1329	4	0.7	6%	12.6	11.1	0.018	13%	0.16	0.12
GA_TPH1347b	4	0.8	9%	9.2	8.1	0.007	8%	0.09	0.08
GA_TPH1347a	4	0.3	4%	8.1	7.4	0.015	25%	0.08	0.05
FL-GA_TPH1604	3	1.4	16%	10.2	7.7	0.020	35%	0.08	0.04
NC-VA_TPH1994	4	0.1	2%	8.5	8.2	0.003	5%	0.06	0.05
NC_TPH1329	4	0.3	5%	7.2	6.8	0.005	11%	0.05	0.04
VA_TPH1537	4	0.6	8%	8.2	7.0	0.008	17%	0.06	0.04
VA_TPH1794	4	0.5	6%	8.2	7.2	0.006	13%	0.05	0.04
GA-TN_TPH2244	2	0.8	11%	7.8	6.7	0.012	28%	0.05	0.03

Table 5. Average genetic parameter estimates across series for height (A.) and volume (B.). Estimates for each series can be found in Appendix A. h^2_F is the narrow-sense family mean heritability, H^2_F is the broad-sense family mean heritability, h^2_i is the narrow-sense individual tree heritability, and H^2_i is the broad-sense individual tree heritability. r_B is the type-‘B’ correlation, which is the ratio of the additive genetic variance to itself plus the variance associated with the additive-by-test interaction.

A. Genetic parameter estimates for height.

Parameter	Age							
	4	5	6	7	8	12	16	20
h^2_{FS}	0.78	0.74	0.77	0.76	0.79	0.79	0.93	0.75
H^2_{FS}	0.87	0.86	0.88	0.88	0.88	0.91	0.97	0.89
h^2_i	0.22	0.22	0.27	0.27	0.29	0.36	0.42	0.17
H^2_i	0.26	0.28	0.33	0.35	0.35	0.44	0.45	0.24
r_B	0.84	0.80	0.87	0.89	0.88	0.83	0.98	0.83

B. Genetic parameter estimates for volume.

Parameter	Age							
	4	5	6	7	8	12	16	20
h^2_{FS}	0.74	0.72	0.74	0.74	0.76	0.77	0.89	0.85
H^2_{FS}	0.80	0.80	0.83	0.87	0.86	0.90	0.95	0.90
h^2_i	0.18	0.19	0.22	0.23	0.24	0.26	0.30	0.21
H^2_i	0.21	0.23	0.27	0.31	0.30	0.33	0.33	0.23
r_B	0.70	0.71	0.76	0.80	0.79	0.80	0.99	0.87

Table 6. Per-hectare yields for top ~10% of families vs. local seed orchard mix. Units in m³.

Series	Group	Age							
		4	5	6	7	8	12	16	20
SC_TPH1329	top 10%	30.6	58.5	110.3	150.3	200.1		549.2	631.4
SC_TPH1329	SOM	25.6	46.5	86.9	114.2	151.8		411.0	465.5
SC_TPH1329	diff.	19.6%	25.9%	27.0%	31.6%	31.8%		33.6%	35.6%
GA_TPH1347b	top 10%	20.5	36.0	65.3	91.7	114.0			
GA_TPH1347b	SOM	25.5	46.5	83.7	117.6	144.4			
GA_TPH1347b	diff.	-19.5%	-22.6%	-21.9%	-22.0%	-21.1%			
NC-VA_TPH1994	top 10%	20.6	33.3	58.5	89.7	113.0			
NC-VA_TPH1994	SOM	18.5	27.6	43.8	63.9	76.6			
NC-VA_TPH1994	diff.	11.0%	20.9%	33.7%	40.4%	47.5%			
GA-TN_TPH2244	top 10%		29.0	46.5	71.3	94.7			
GA-TN_TPH2244	SOM		28.7	44.9	65.5	84.6			
GA-TN_TPH2244	diff.		1.1%	3.5%	8.7%	11.9%			
NC_TPH1329	top 10%	12.4	19.6	39.6	60.3				
NC_TPH1329	SOM	12.5	18.0	34.4	57.1				
NC_TPH1329	diff.	-0.7%	8.8%	15.1%	5.7%				
VA_TPH1794	top 10%	15.6	21.4	32.9	50.4	76.2	173.1		
VA_TPH1794	SOM	14.3	17.8	25.2	37.2	56.1	135.8		
VA_TPH1794	diff.	8.9%	19.7%	30.6%	35.6%	35.8%	27.4%		
FL-GA_TPH1604	top 10%	14.7	23.7	41.1	56.6	76.7	173.3		
FL-GA_TPH1604	SOM	13.2	19.7	32.6	42.1	55.5	109.9		
FL-GA_TPH1604	diff.	11.0%	20.5%	26.1%	34.5%	38.3%	57.8%		
GA_TPH1347a	top 10%	18.4	37.7	53.9	72.5	101.8	219.1		
GA_TPH1347a	SOM	18.5	37.7	53.5	72.9	102.9	230.8		
GA_TPH1347a	diff.	-0.1%	0.2%	0.8%	-0.5%	-1.0%	-5.0%		
VA_TPH1537	top 10%	13.0	21.2	33.4	49.3	67.4			379.8
VA_TPH1537	SOM	12.7	19.2	28.8	41.0	55.2			304.3
VA_TPH1537	diff.	2.1%	10.6%	15.9%	20.4%	22.2%			24.8%
	Avg. diff.	4.0%	9.4%	14.5%	17.2%	20.7%	26.7%	33.6%	30.2%

Table 7. Per-hectare productivity for series mean volumes. Units in m³.

Series	Age							
	4	5	6	7	8	12	16	20
SC_TPH1329	31	61	101	138	181		477	590
GA_TPH1347b	19	35	60	86	106			
NC-VA_TPH1994	19	29	49	76	102			
GA-TN_TPH2244		26	39	60	81			
NC_TPH1329	12	19	44	55				
VA_TPH1794	16	23	35	54	78	174		
FL-GA_TPH1604	14	22	37	53	74	167		
GA_TPH1347a	15	25	36	49	68	154		
VA_TPH1537	14	20	33	48	66			328

Table 8. Per-hectare SEVs for series mean volumes, using 2013 stumpage pricing (\$11/Mg pulpwood, \$17.50/Mg chip-n-saw). Prices in parentheses indicate negative values.

Series	Age							
	4	5	6	7	8	12	16	20
SC_TPH1329	(\$2,276)	(\$1,357)	(\$614)	(\$173)	\$191		\$1,634	\$1,348
GA_TPH1347b	(\$2,595)	(\$1,859)	(\$1,267)	(\$848)	(\$623)			
NC-VA_TPH1994	(\$2,734)	(\$2,090)	(\$1,529)	(\$1,063)	(\$748)			
NC_TPH1329	(\$2,766)	(\$2,168)	(\$1,519)	(\$1,242)				
GA-TN_TPH2244		(\$2,194)	(\$1,725)	(\$1,300)	(\$1,002)			
FL-GA_TPH1604	(\$2,760)	(\$2,163)	(\$1,666)	(\$1,313)	(\$1,008)	(\$360)		
VA_TPH1794	(\$2,763)	(\$2,187)	(\$1,729)	(\$1,318)	(\$986)	(\$335)		
GA_TPH1347a	(\$2,681)	(\$2,054)	(\$1,643)	(\$1,325)	(\$1,033)	(\$416)		
VA_TPH1537	(\$2,758)	(\$2,185)	(\$1,717)	(\$1,357)	(\$1,079)			\$3

Table 9. Per-hectare SEVs for series mean volumes, using \$17.50 (A.), \$22.00 (B.) and \$33.00 (C.) per Mg stumpage pricing. Prices in parentheses indicate negative values.

A. Per-hectare SEVs using \$17.50/Mg stumpage pricing.

Series	Age							
	4	5	6	7	8	12	16	20
SC_TPH1329	(\$1,798)	(\$641)	\$339	\$897	\$1,369		\$1,847	\$1,422
GA_TPH1347b	(\$2,306)	(\$1,443)	(\$704)	(\$183)	\$68			
NC-VA_TPH1994	(\$2,448)	(\$1,746)	(\$1,066)	(\$475)	(\$86)			
NC_TPH1329	(\$2,581)	(\$1,939)	(\$1,109)	(\$814)				
GA-TN_TPH2244		(\$1,886)	(\$1,356)	(\$834)	(\$474)			
VA_TPH1794	(\$2,519)	(\$1,922)	(\$1,403)	(\$899)	(\$481)	\$300		
FL-GA_TPH1604	(\$2,538)	(\$1,903)	(\$1,320)	(\$906)	(\$530)	\$249		
GA_TPH1347a	(\$2,443)	(\$1,754)	(\$1,306)	(\$946)	(\$589)	\$146		
VA_TPH1537	(\$2,544)	(\$1,944)	(\$1,407)	(\$981)	(\$648)			\$316

B. Per-hectare SEVs using \$22.00/Mg stumpage pricing.

Series	Age							
	4	5	6	7	8	12	16	20
SC_TPH1329	(\$1,479)	(\$164)	\$974	\$1,610	\$2,154		\$2,596	\$2,036
GA_TPH1347b	(\$2,113)	(\$1,166)	(\$328)	\$261	\$528			
NC-VA_TPH1994	(\$2,257)	(\$1,516)	(\$757)	(\$83)	\$355			
GA-TN_TPH2244		(\$1,681)	(\$1,111)	(\$523)	(\$122)			
NC_TPH1329	(\$2,458)	(\$1,787)	(\$835)	(\$529)				
VA_TPH1794	(\$2,356)	(\$1,745)	(\$1,186)	(\$620)	(\$144)	\$723		
FL-GA_TPH1604	(\$2,389)	(\$1,730)	(\$1,090)	(\$634)	(\$211)	\$656		
GA_TPH1347a	(\$2,285)	(\$1,555)	(\$1,081)	(\$693)	(\$293)	\$521		
VA_TPH1537	(\$2,400)	(\$1,784)	(\$1,201)	(\$731)	(\$361)			\$658

C. Per-hectare SEVs using \$33.00/Mg stumpage pricing.

Series	Age							
	4	5	6	7	8	12	16	20
SC_TPH1329	(\$682)	\$1,028	\$2,562	\$3,393	\$4,117		\$4,470	\$3,573
GA_TPH1347b	(\$1,631)	(\$473)	\$610	\$1,370	\$1,680			
NC-VA_TPH1994	(\$1,780)	(\$942)	\$15	\$898	\$1,458			
GA-TN_TPH2244		(\$1,168)	(\$497)	\$253	\$757			
NC_TPH1329	(\$2,150)	(\$1,405)	(\$152)	\$184				
VA_TPH1794	(\$1,950)	(\$1,303)	(\$643)	\$79	\$698	\$1,781		
FL-GA_TPH1604	(\$2,019)	(\$1,296)	(\$513)	\$45	\$585	\$1,672		
GA_TPH1347a	(\$1,889)	(\$1,056)	(\$519)	(\$60)	\$447	\$1,458		
VA_TPH1537	(\$2,042)	(\$1,383)	(\$684)	(\$105)	\$356			\$1,512

Table 10. Per-hectare SEVs for top 10% of families vs. local seed orchard mix (SOM) for each series, using 2013 stumpage pricing. Prices in parentheses indicate negative values.

Series	Group	Age							
		4	5	6	7	8	12	16	20
SC_TPH1329	top 10%	(\$2,324)	(\$1,427)	(\$493)	(\$35)	\$377		\$2,046	\$1,504
SC_TPH1329	SOM	(\$2,452)	(\$1,664)	(\$861)	(\$501)	(\$146)		\$1,238	\$834
SC_TPH1329	diff.	\$128	\$237	\$368	\$466	\$523		\$808	\$671
GA_TPH1347b	top 10%	(\$2,585)	(\$1,873)	(\$1,203)	(\$795)	(\$558)			
GA_TPH1347b	SOM	(\$2,458)	(\$1,666)	(\$914)	(\$460)	(\$229)			
GA_TPH1347b	diff.	(\$127)	(\$206)	(\$288)	(\$334)	(\$329)			
NC-VA_TPH1994	top 10%	(\$2,718)	(\$2,037)	(\$1,405)	(\$906)	(\$646)			
NC-VA_TPH1994	SOM	(\$2,771)	(\$2,150)	(\$1,637)	(\$1,239)	(\$1,040)			
NC-VA_TPH1994	diff.	\$52	\$113	\$232	\$333	\$394			
GA-TN_TPH2244	top 10%		(\$2,164)	(\$1,632)	(\$1,177)	(\$874)			
GA-TN_TPH2244	SOM		(\$2,171)	(\$1,657)	(\$1,251)	(\$983)			
GA-TN_TPH2244	diff.		\$6	\$25	\$74	\$109			
NC_TPH1329	top 10%	(\$2,790)	(\$2,192)	(\$1,604)	(\$1,197)				
NC_TPH1329	SOM	(\$2,788)	(\$2,223)	(\$1,685)	(\$1,239)				
NC_TPH1329	diff.	(\$2)	\$31	\$82	\$42				
VA_TPH1794	top 10%	(\$2,803)	(\$2,237)	(\$1,778)	(\$1,387)	(\$1,020)	(\$357)		
VA_TPH1794	SOM	(\$2,836)	(\$2,307)	(\$1,899)	(\$1,558)	(\$1,238)	(\$583)		
VA_TPH1794	diff.	\$33	\$69	\$121	\$171	\$218	\$226		
FL-GA_TPH1604	top 10%	(\$2,788)	(\$2,159)	(\$1,620)	(\$1,281)	(\$992)	(\$338)		
FL-GA_TPH1604	SOM	(\$2,825)	(\$2,238)	(\$1,754)	(\$1,469)	(\$1,222)	(\$723)		
FL-GA_TPH1604	diff.	\$37	\$79	\$134	\$188	\$230	\$385		
GA_TPH1347a	top 10%	(\$2,639)	(\$1,839)	(\$1,381)	(\$1,042)	(\$689)	(\$37)		
GA_TPH1347a	SOM	(\$2,638)	(\$1,840)	(\$1,388)	(\$1,037)	(\$678)	\$34		
GA_TPH1347a	diff.	(\$0)	\$1	\$6	(\$5)	(\$12)	(\$71)		
VA_TPH1537	top 10%	(\$2,818)	(\$2,197)	(\$1,732)	(\$1,367)	(\$1,085)			\$158
VA_TPH1537	SOM	(\$2,825)	(\$2,237)	(\$1,804)	(\$1,475)	(\$1,218)			(\$84)
VA_TPH1537	diff.	\$7	\$40	\$72	\$108	\$133			\$243
	Avg-diff.	\$16	\$41	\$84	\$116	\$158	\$180	\$808	\$457

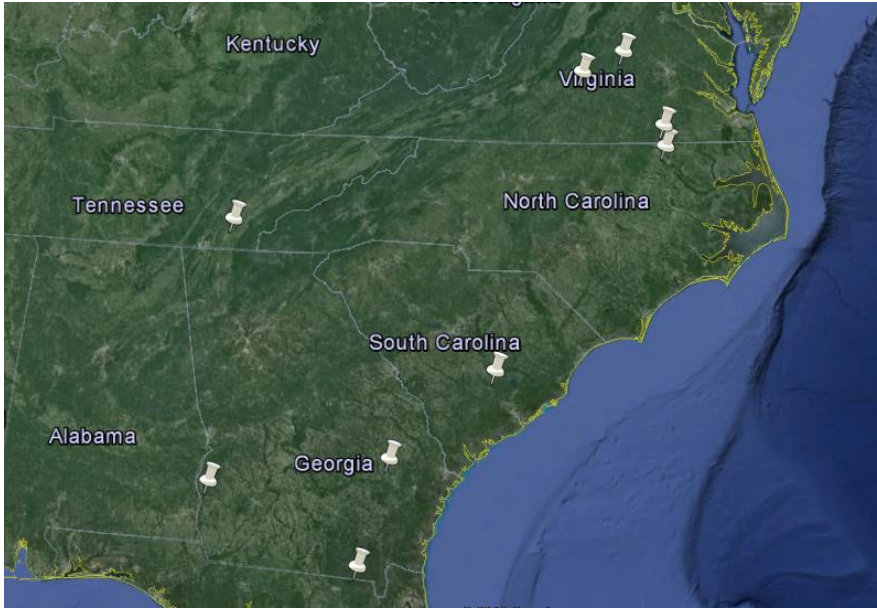
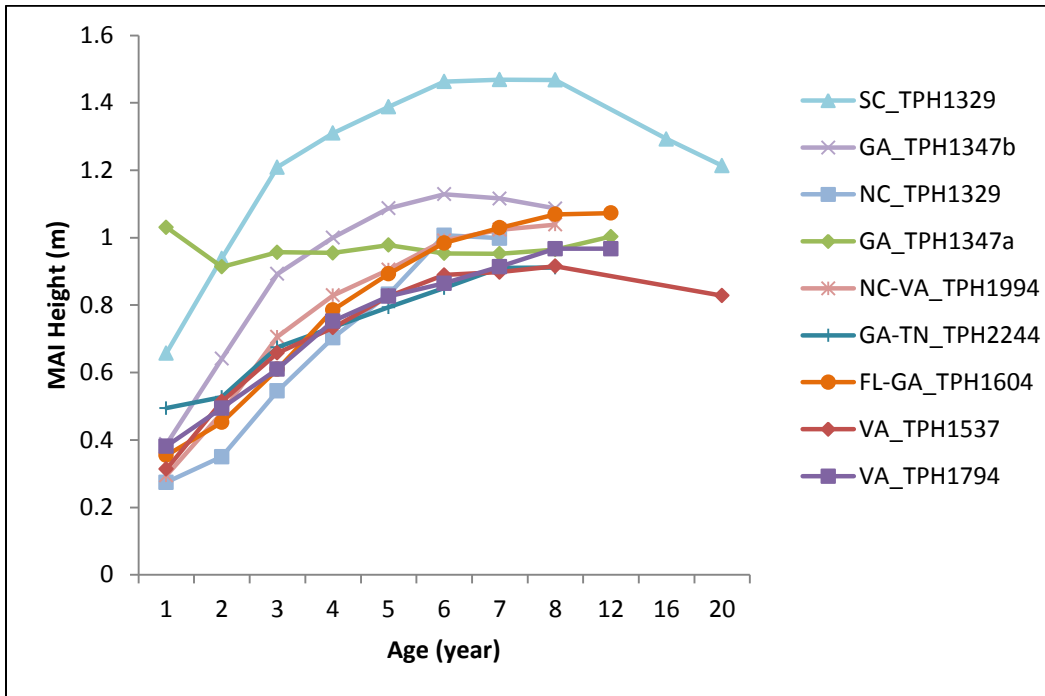


Figure 1. Approximate locations of the nine test series analyzed. The tests within each series were generally located in the same or contiguous counties

A.



B.

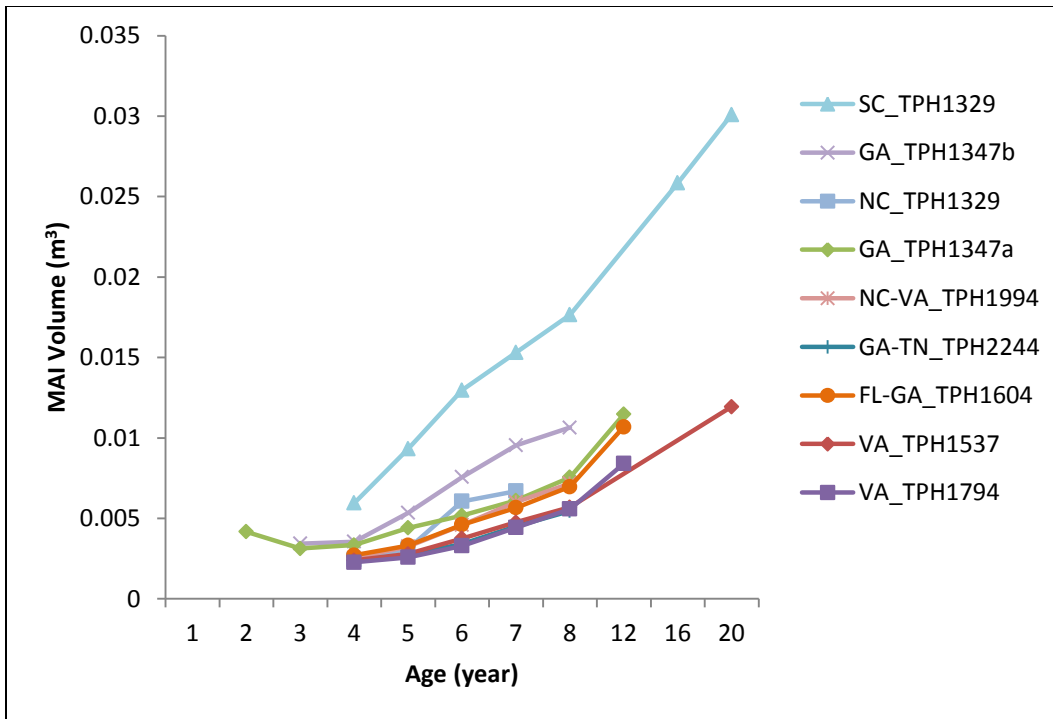


Figure 2. Individual tree mean annual increment for height (A) and volume (B) for the nine study test series.

APPENDIX

Appendix A. Genetic parameter estimates by series for height (A.) and volume (B.). h^2_{FS} is the narrow-sense family mean heritability, H^2_{FS} is the broad-sense family mean heritability, h^2_i is the narrow-sense individual tree heritability, and H^2_i is the broad-sense individual tree heritability. r_B is the type-‘B’ correlation, which is the ratio of the additive genetic variance to itself plus the variance associated with the additive-by-test interaction.

A. Genetic parameter estimates for height.

Series	Parameter	Age							
		4	5	6	7	8	12	16	20
FL-GA_TPA636	h^2_{FS}	0.84	0.89	0.89	0.92	0.93	0.94		
FL-GA_TPA636	H^2_{FS}	0.92	0.93	0.93	0.93	0.94	0.96		
FL-GA_TPA636	h^2_i	0.33	0.42	0.45	0.48	0.51	0.61		
FL-GA_TPA636	H^2_i	0.39	0.45	0.48	0.50	0.53	0.63		
FL-GA_TPA636	r_B	0.96	0.94	0.94	0.93	0.93	0.96		
GA_TPA545a	h^2_{FS}	0.79	0.76	0.79	0.8	0.77	0.65		
GA_TPA545a	H^2_{FS}	0.87	0.89	0.9	0.9	0.88	0.84		
GA_TPA545a	h^2_i	0.22	0.22	0.23	0.24	0.23	0.17		
GA_TPA545a	H^2_i	0.26	0.30	0.30	0.30	0.30	0.27		
GA_TPA545a	r_B	0.92	0.87	0.87	0.85	0.78	0.7		
VA_TPA726	h^2_{FS}	0.69	0.73	0.78	0.81	0.8	0.77		
VA_TPA726	H^2_{FS}	0.86	0.87	0.92	0.93	0.94	0.92		
VA_TPA726	h^2_i	0.14	0.18	0.28	0.31	0.32	0.3		
VA_TPA726	H^2_i	0.21	0.25	0.38	0.41	0.43	0.43		
VA_TPA726	r_B	0.74	0.73	0.82	0.86	0.87	0.84		
GA-TN_TPA908	h^2_{FS}		0.39	0.34	0.4	0.65			
GA-TN_TPA908	H^2_{FS}		0.67	0.68	0.65	0.68			
GA-TN_TPA908	h^2_i		0.08	0.07	0.10	0.19			
GA-TN_TPA908	H^2_i		0.18	0.21	0.23	0.21			
GA-TN_TPA908	r_B		0.56	0.74	>0.99	>0.99			
SC_TPA538	h^2_{FS}	0.79	0.84	0.83	0.82	0.86		0.93	0.77
SC_TPA538	H^2_{FS}	0.88	0.9	0.93	0.95	0.95		0.97	0.95
SC_TPA538	h^2_i	0.21	0.3	0.35	0.35	0.42		0.42	0.22
SC_TPA538	H^2_i	0.26	0.34	0.43	0.46	0.5		0.45	0.32
SC_TPA538	r_B	0.78	0.81	0.86	0.92	0.97		0.98	>0.99
NC-VA_TPA807	h^2_{FS}	0.83	0.79	0.7	0.72	0.77			
NC-VA_TPA807	H^2_{FS}	0.88	0.90	0.89	0.92	0.93			
NC-VA_TPA807	h^2_i	0.18	0.18	0.16	0.17	0.20			
NC-VA_TPA807	H^2_i	0.20	0.23	0.24	0.26	0.28			
NC-VA_TPA807	r_B	0.89	0.92	>0.99	0.92	0.92			
VA_TPA622	h^2_{FS}	0.57	0.68	0.79	0.79	0.74			0.72
VA_TPA622	H^2_{FS}	0.70	0.77	0.80	0.81	0.82			0.82
VA_TPA622	h^2_i	0.06	0.11	0.14	0.14	0.14			0.12
VA_TPA622	H^2_i	0.09	0.14	0.14	0.15	0.16			0.15
VA_TPA622	r_B	0.56	0.61	0.79	0.81	0.87			0.65
GA_TPA545b	h^2_{FS}	0.83	0.79	0.88	0.88	0.77			
GA_TPA545b	H^2_{FS}	0.92	0.91	0.94	0.93	0.88			
GA_TPA545b	h^2_i	0.32	0.29	0.46	0.47	0.32			
GA_TPA545b	H^2_i	0.39	0.38	0.53	0.53	0.40			
GA_TPA545b	r_B	0.89	0.80	0.87	0.83	0.70			
NC_TPA538	h^2_{FS}	0.86	0.83	0.91	0.73				
NC_TPA538	H^2_{FS}	0.92	0.90	0.93	0.93				
NC_TPA538	h^2_i	0.26	0.22	0.28	0.20				
NC_TPA538	H^2_i	0.3	0.26	0.29	0.31				
NC_TPA538	r_B	>0.99	0.93	0.96	0.93				

B. Genetic parameter estimates for volume.

Series	Parameter	Age							
		4	5	6	7	8	12	16	20
FL-GA_TPA636	h_{FS}^2	0.76	0.83	0.86	0.86	0.90	0.91		
FL-GA_TPA636	H_{FS}^2	0.81	0.88	0.89	0.89	0.91	0.91		
FL-GA_TPA636	h_i^2	0.21	0.29	0.33	0.34	0.39	0.43		
FL-GA_TPA636	H_i^2	0.24	0.32	0.35	0.37	0.40	0.43		
FL-GA_TPA636	r_B	0.73	0.89	0.88	0.87	0.89	0.89		
GA_TPA545a	h_{FS}^2	0.79	0.77	0.74	0.76	0.74	0.70		
GA_TPA545a	H_{FS}^2	0.87	0.87	0.86	0.88	0.87	0.86		
GA_TPA545a	h_i^2	0.23	0.24	0.23	0.22	0.21	0.17		
GA_TPA545a	H_i^2	0.28	0.30	0.30	0.30	0.28	0.25		
GA_TPA545a	r_B	0.83	0.78	0.73	0.77	0.71	0.68		
VA_TPA726	h_{FS}^2	0.66	0.70	0.71	0.71	0.72	0.70		
VA_TPA726	H_{FS}^2	0.84	0.86	0.91	0.91	0.92	0.92		
VA_TPA726	h_i^2	0.12	0.15	0.19	0.20	0.22	0.19		
VA_TPA726	H_i^2	0.19	0.23	0.29	0.32	0.34	0.32		
VA_TPA726	r_B	0.70	0.71	0.78	0.80	0.83	0.84		
GA-TN_TPA908	h_{FS}^2		0.17	0.24	0.38	0.51			
GA-TN_TPA908	H_{FS}^2		0.27	0.40	0.61	0.63			
GA-TN_TPA908	h_i^2		0.04	0.06	0.10	0.15			
GA-TN_TPA908	H_i^2		0.10	0.15	0.21	0.22			
GA-TN_TPA908	r_B		0.13	0.20	0.44	0.52			
SC_TPA538	h_{FS}^2	0.84	0.87	0.90	0.90	0.90		0.89	0.84
SC_TPA538	H_{FS}^2	0.84	0.87	0.91	0.93	0.94		0.95	0.91
SC_TPA538	h_i^2	0.21	0.23	0.29	0.32	0.33		0.30	0.22
SC_TPA538	H_i^2	0.21	0.24	0.30	0.34	0.36		0.33	0.25
SC_TPA538	r_B	0.71	0.76	0.86	0.90	0.90		0.99	0.88
NC-VA_TPA807	h_{FS}^2	0.80	0.82	0.76	0.79	0.84			
NC-VA_TPA807	H_{FS}^2	0.84	0.89	0.92	0.93	0.93			
NC-VA_TPA807	h_i^2	0.15	0.17	0.19	0.22	0.24			
NC-VA_TPA807	H_i^2	0.16	0.20	0.26	0.30	0.29			
NC-VA_TPA807	r_B	0.80	0.89	0.91	0.90	0.89			
VA_TPA622	h_{FS}^2	0.39	0.65	0.70	0.73	0.68			0.86
VA_TPA622	H_{FS}^2	0.48	0.74	0.78	0.81	0.82			0.89
VA_TPA622	h_i^2	0.03	0.08	0.10	0.12	0.11			0.19
VA_TPA622	H_i^2	0.04	0.11	0.13	0.15	0.16			0.20
VA_TPA622	r_B	0.27	0.58	0.70	0.78	0.88			0.85
GA_TPA545b	h_{FS}^2	0.90	0.87	0.88	0.86	0.78			
GA_TPA545b	H_{FS}^2	0.94	0.93	0.93	0.93	0.89			
GA_TPA545b	h_i^2	0.37	0.35	0.39	0.38	0.30			
GA_TPA545b	H_i^2	0.41	0.40	0.44	0.44	0.38			
GA_TPA545b	r_B	0.94	0.87	0.87	0.85	0.73			
NC_TPA538	h_{FS}^2	0.74	0.81	0.84	0.71				
NC_TPA538	H_{FS}^2	0.78	0.87	0.91	0.93				
NC_TPA538	h_i^2	0.13	0.18	0.20	0.20				
NC_TPA538	H_i^2	0.15	0.21	0.23	0.33				
NC_TPA538	r_B	0.62	0.80	0.87	0.92				

CHAPTER 2

GENETIC VARIATION IN SUGAR YIELD FROM LOBLOLLY PINE (*PINUS TAEDA L.*) WOOD: I: ALKALINE AND DILUTE ACID ENZYMATIC HYDROLYSIS OF STEMWOOD SAMPLES.

Abstract

Loblolly pine (*Pinus taeda* L.) as the most productive forest tree species in the Southeastern U.S. is a logical species of interest for biofuel production in this region. However, loblolly pine biomass currently presents a challenge for producing ethanol; enzymatic hydrolysis of polysaccharides from softwood pulp typically produces lower yields of fermentable sugars than similar treatment of hardwood pulp. Many chemical and physical wood properties in loblolly pine are subject to genetic control, and variation in some of these properties will likely have an effect on the efficiency of ethanol production.

Wood samples were collected from a series of 8-year-old clonal trials in South Carolina and Georgia. Clonal varieties of loblolly pine were divided into groups using a cluster analysis based on near-infrared (NIR) spectra of ground wood samples from multiple individual trees, or ramets, of each of the clonal genotypes. NIR spectra reflect chemical and physical wood properties, so the clustering should have produced groups of clones that are similar for some combination of these properties. The cluster analysis was performed on 178 clones, and a subset of these clones was selected for enzymatic hydrolysis. Wood samples from three pooled ramets of each clone in this subset were tested, using enzymatic hydrolysis after a dilute acid or alkaline pretreatment to produce data on sugar yields for each clone. Sugar yield is directly correlated with ethanol yield since the sugar is fermented to produce ethanol. The lowest yielding treatment was a dilute acid pretreatment followed by enzymatic hydrolysis using 20 filter paper units (FPU) of enzyme, which produced an average of 210 mg sugar/g wood. The highest yielding treatment was an alkaline pretreatment followed by

enzymatic hydrolysis using 40 FPU as well as mechanical beating, which produced 520 mg sugar/g wood. The NIR clustering as well as a number of wet chemistry-determined wood properties were significant predictors of sugar yield for most of the treatments. Given the high heritabilities of most chemical and physical wood properties, these results suggest that it should be possible to identify superior genotypes for biofuel production using NIR analysis.

Introduction

An obstacle to the development of a biofuels industry is that bioconversion techniques are still developing; Burkheisser (2010) and Mousdale (2008) provide a good picture of the current state of biorefining technologies. At the same time, the type and quantity of locally produced feedstock available will guide the selection of conversion technology.

Biochemical techniques focus on hydrolytic processes that retrieve the sugars from the cellulose and hemicellulose. These techniques can use a variety of chemical and/or enzyme treatments to help loosen the grip of lignin on the carbohydrates and to break down those carbohydrates. Biochemical conversion is typically assisted by chemical and/or enzymatic agents, and fermentation of the sugars can be improved through specially tailored microbes (Mousdale 2008). For cellulosic ethanol, the main difficulty is extracting the valuable sugars from the biomass with as little input and as little sugar loss as possible. The structural properties of these lignocellulosic feedstocks are variable, which means that biochemical conversion techniques will need to be calibrated for specific feedstocks.

Thermochemical techniques for biofuel production involve the application of heat to the biomass. The type of biomass and the conditions of the heating will affect the production process, but typical outputs of thermochemical methods are water, oils, charcoal, and synthesis gases (Overend 2004). The removal of water is essential for producing these products, and the amount of water in the biomass at the beginning of the thermochemical process can have a big effect on the amount of energy consumed during the heating process. Bio-oils can often be used in diesel fuels, and charcoal can be burned for energy as can raw biomass, such as firewood. The syngas produced can be converted to liquid fuel via Fischer-Tropsch reactions (Mousdale 2008). Thermochemical methods of biofuel production tend to avoid the problems caused by lignin in the biochemical methods, as lignin is an energy-rich polymer than is pyrolyzed or combusted along with the carbohydrate components. Lignin, as a byproduct of any biochemical or thermochemical process, may also have a higher use as a renewable source for a variety of useful aromatic compounds (Pandey and Kim 2011).

In making cellulosic ethanol from woody biomass via enzymatic hydrolysis, lignin is the major obstacle to overcome. Lignin, an irregularly branching aromatic phenylpropanoid polymer, serves as the “glue” that binds the cellulose and hemicellulose components of wood together and gives woody biomass its strength. The initial phase of the hydrolysis process is generally a pretreatment method designed to make the lignin “release” the cellulose, making it available to undergo hydrolysis (Hu et al. 2008). Lignin fills in the spaces around the cellulose and hemicelluloses in the secondary cell wall and is thought to form covalent bonds with the hemicellulose. Lignin is a very difficult polymer to break up, and the loss of

cellulose and hemicellulose can be an issue due to the harsh chemical processes needed to degrade the lignin. Hardwoods have been thought to be preferable feedstocks for enzymatic hydrolysis, as the lignin composition of hardwoods tends to lead to easier hydrolysis and superior sugar yields (Grethlein et al. 1984; Ramos et al. 1992; and Yu et al. 2011).

Softwood has been shown to be especially recalcitrant compared to hardwood in terms of percent sugar yield relative to total available carbohydrate (Grethlein et al. 1984, Ramos et al. 1992). The major causes of lower sugar yields in softwood are related to the characteristics of softwood lignin. The composition of cellulose is the same in both hardwood and softwood, and is not the cause of softwood recalcitrance. Grethlein et al. (1984) found very different glucose yields for white pine and mixed hardwood chips (maximums of 65 and 95%, respectively), whereas potential glucose was the same for both samples (42%). While average cellulose contents across species may be lower for softwoods (Pettersen 1984), this does not account for the differences in sugar yields. Additionally, most softwoods have larger hemicellulose fractions than hardwoods (Mabee et al. 2006).

Softwood hemicelluloses tend to contain more mannans than xylans, and as such should actually be easier to digest than hardwood hemicelluloses, which generally contain more xylans (Grethlein et al. 1984, Gregg and Saddler 1996, Saha 2003).

Much of the recalcitrance of softwood relative to hardwood sugar yields is related to the differential chemical composition of softwood and hardwood lignin. Lignin is produced via the shikimic acid metabolic pathway. The shikimic acid pathway is also responsible for

production tannins and other phenolic extractives in the wood. The shikimic pathway produces the amino acids tyrosine and phenylalanine, which in turn produce the lignin monomer precursors via the cinnamate metabolic pathway. The three lignin monomer precursors *p*-coumaryl, coniferyl, and sinapyl alcohols (Figure 1) become the lignin *p*-hydroxyphenol, guaiacyl, and syringyl subunits, respectively. Lignin monomers are combined, through a process of radicalization and coupling, to create the lignin polymer (Vanholme et al. 2010).

Softwood lignin varies from hardwood lignin in that it contains a preponderance of guaiacyl units, whereas hardwood has significant amounts of both guaiacyl and syringyl units (Sjöström 1993, Stenius 2000). Both softwoods and hardwoods contain small fractions of *p*-hydroxyphenol units. Syringyl units (which are not present in softwoods) tend to form fewer interlinking bonds during the biosynthesis of lignin. This is due to the absence of a free position on the phenyl group which is available in guaiacyl units. That is, the carbon-3 and carbon-5 positions on the phenyl group in syringyl units are occupied by methoxy groups, whereas in guaiacyl units only the carbon-3 position is methoxylated. This means that the carbon-5 position is available for crosslinking during biosynthesis of the lignin polymer in softwood. Under acid cooking conditions, guaiacyl ring carbons have been found to be less reactive than syringyl ring carbons, and the condensation products of guaiacyl-based lignin are more stable (Shimada et al. 1997). This means that under pulping conditions, softwood is less amenable to delignification, which will result in less sugar yield compared to hardwood.

The stability of softwood lignin is probably due to the unmodified phenyl ring position on the guaiacyl monomer.

An additional drawback to softwood lignin is its effect on the interaction of enzymes with the cellulose and hemicellulose. Pan et al. (2005) found that softwood lignin reduces the efficacy of cellulase enzymes in two ways, physically blocking the access of the cellulase to the cellulose, and binding to cellulases and rendering them inactive. Removal of a portion of the lignin improved hydrolysis results, indicating that the lignin formed a physical barrier, and experiments with adding extra proteins to the reaction showed that a portion of the lignin likely does have an affinity for protein, which affirms that cellulase is captured by the lignin. In comparison, cellulase was not shown to have affinity for adsorption for a variety of hardwood substrates (Yu et al. 1995).

Research has been done on the recycling of enzymes in the hope of improving the efficiency and cost of enzymatic hydrolysis. Hydrolysis of the entire substrate is essential to effective enzyme recovery. Substrate residues with higher lignin contents will not hydrolyze as well as substrates with lower lignin contents, and lower levels of cellulase enzyme recovery have been observed from these high-lignin samples (Lee et al. 1994). Since softwood lignin is more likely to build up in the substrate, enzyme recovery will be inhibited for softwoods (Boussaid and Saddler 1999). The fact that softwood also starts with slightly higher lignin content than hardwood does not help (Pettersen 1984). The buildup of softwood lignin in the substrate is likely due to its relative stability in reaction, compared to that of hardwood.

However, work in the area of enzyme recovery suggests that it may be possible to achieve better results in softwood using an ultrafiltration cell with a high molecular weight membrane (Lu et al. 2002).

The term ‘lignin-carbohydrate complex’ refers to the interlinked, covalently-bonded system of lignin, hemicellulose, and cellulose. In softwood, these complexes seem to have more stability than in hardwood (Karlsson and Westermark 1998). In addition, softwood lignin seems to restrict the amount of surface area accessible to the enzyme in two ways: Firstly, softwood lignin may reduce the porosity of the substrate, meaning that the enzyme has little ability to penetrate the lignin-carbohydrate complex (Mooney et al. 1998, Yu et al. 2011). Secondly, guaiacyl-based lignin restricts the swelling of the cellulose, further interfering with enzyme access (Mooney et al. 1998, Ramos et al. 1992).

In summary, the recalcitrance of softwood is likely due to the particular chemical composition of guaiacyl-type lignin. The presence of the unmodified phenyl ring position in coniferyl alcohol causes softwood lignin to be much more stable relative to hardwood lignin. The strength of the crosslinking between guaiacyl-units makes softwood more resistant than hardwood to delignification and makes the wood carbohydrates less accessible to enzymes. Additionally, softwood lignin also absorbs cellulase and hinders enzyme recovery to a greater extent than hardwood lignin.

Genetic Control of Wood Properties

Wood properties are largely defined by the properties of the cell wall, and given that lignin is a major component of the cell wall, it is likely that genetic manipulation of wood properties will have some bearing on the characteristics of the lignin present in the wood and thereby the recalcitrance of the wood. Wood qualities tend to be under strong genetic control (have high heritability, $h^2 = \sigma^2_{\text{genetic}} / \sigma^2_{\text{phenotypic}}$), which should enable the identification of particular varieties well suited to bioconversion. For example, modifications to cellulose and lignin content in softwoods should be possible; in loblolly pine, Sykes et al. (2006) found half-sib family mean heritabilities of 0.46 and .56 for transition wood cellulose and lignin content, respectively.

Wood specific gravity, generally considered to be the most important single wood property, has h^2 values between 0.4 and 0.7 in conifers (Zobel and van Buijtenen 1989). Higher specific gravity is related to lower moisture content, meaning families with higher specific gravity will produce more biomass per green ton than families with lower specific gravity. Specific gravity is closely related to other cell properties, such as latewood percentage, cell wall thickness, cell (tracheid/vessel) length (h^2 of .44), and cell diameter.

Microfibril angle (the physical orientation of the cellulose bundles to the axis of the cell) is closely correlated with cell length and wall thickness (Zobel and van Buijtenen 1989). This may be of significance in the delignification process as microfibril angle likely has an impact on the disposition of the lignin in the cell (Chaffey 2000), and more importantly, on the

pulping process itself. Microfibril angle has been found to be under some genetic control in *Eucalyptus* ($h^2=0.29$ and 0.43 in Lima et al. 2005 and Hein et al. 2012, respectively), *Pinus radiata* ($h^2 = 0.37$ to 0.62 ; Baltunis et al. 2007) and *Pinus taeda* ($h^2 = 0.17$ to 0.51 , Myszewski 2004; H^2 (clone mean repeatability) = 0.79 , Isik et al. 2008).

Compression wood is also under genetic control; Shelbourne et al. (1969) found a very high heritability for compression wood in loblolly pine ($h^2 = 0.95$). This genetic effect on compression wood makes sense, as compression wood is highly related to stem straightness (Zobel and Haught 1962), which has moderate heritability ($h^2 = .38$; Shelbourne and Stonecypher 1971). The relevance of compression wood for pulp and biomass is related to the fact that compression wood in conifers contains an overall higher level of lignin (Zobel and van Buijtenen 1989) as well as an elevated level of p-hydroxyphenyl monolignols (the precursor of which is coumaryl alcohol). p-Hydroxyphenyl monolignols lack methoxy groups at both the carbon-3 and carbon-5 phenyl groups, resulting in lignin with more crosslinks that may be even more difficult to process than guaiacyl units (Whetten and Sederoff 1991).

Juvenile wood is also a concern for biofuel production. The area of juvenile wood formation encompasses the first 5 to 20 growth rings, depending on the species (Zobel and van Buijtenen 1989), and has different wood properties than that of mature wood. In pine, this can include lower specific gravity, higher moisture content, thinner cell walls, short cell lengths, higher levels of compression wood, and higher levels of lignin (Zobel and Blair

1976). Given that the two major sources of pine biomass that will likely be available for biorefineries are thinnings and young biomass plantations, juvenile wood will likely make up a major proportion of the biomass being processed at a biorefinery. In loblolly and slash pine, there is genetic variation for transition age when juvenile wood formation gives way to the development of mature wood (Loo et al. 1985; and Hodge and Purnell 1993).

Evaluation of wood properties and components through wet chemistry can be time-consuming and expensive. However, there are spectroscopic methods that can predict many of these properties. Near infrared reflectance (NIR) spectroscopy is one such technique, wherein light reflection or absorption of a sample is measured for some portion of the near infrared light spectrum (typically wavelengths from 700 up to 2500 nm). Early uses of this technique in forestry include analysis of foliage for nitrogen, lignin, and cellulose content (Wessman et al. 1988, McClellan et al. 1991, Martin and Aber 1994). These studies correlated NIR spectra to wet-lab measurements of these wood properties using multiple linear regression methods. Another statistical method for correlating laboratory measurements with spectral data is partial least squares (PLS; also referred to as projection to latent structures) (Martens and Jensen 1983). This technique can be better than stepwise multiple regression, providing predictions with lower standard errors (Shenk and Westerhaus 1991). PLS models can take more wavelengths into account than can linear regression models and are able to deal with the correlations between nearby wavelengths.

NIR spectroscopy has been used for prediction of wood and pulp properties (Wright et al.

1990, Antti et al. 1996, Kelley et al. 2004a & b, Hodge and Woodbridge 2010). Kelley et al. (2004a) successfully used NIR spectroscopy to make predictions of modulus of elasticity and modulus of rupture in six softwood species. For most models, the correlations between predicted and actual measurements exceeded 0.80. A study on predicting lignin and sugar content in loblolly pine using NIR spectra also found correlations that were generally higher than 0.80 (Kelley et al. 2004b). More recently, Hodge and Woodbridge (2010) have combined data from seven pine species (tropical and temperate) to develop global prediction models for lignin and cellulose content in pines. Correlation coefficients from these global models were high (0.97 for lignin and 0.82 for cellulose). Predicting other wood properties in loblolly pine such as cell length and wall thickness/lumen diameter using near-infrared spectra has also been demonstrated (Sykes et al. 2005; Aspinwall et al. 2010).

This experiment looks to improve outcomes for biofuel development in the Southeast through investigation of variation in sugar yields and optimization of chemical pretreatments. To foster the development of a biofuel industry, it will be useful to determine what traits (if any) have the potential to improve the efficiency of the conversion of loblolly pine to ethanol via enzymatic hydrolysis.

Methods

Wood samples were collected from a series of 8-year-old clonal trials conducted by Plum Creek Timber Company and CellFor Inc. The trees for this test were produced using somatic embryogenesis, a technique wherein plant embryos are cultured and bulked up to produce

clonal replicates. The test site was laid out in an alpha-lattice, randomized incomplete-block design, with single-tree plots. In all, approximately 200 clones from 13 full-sib and two half-sib families were tested. Wood samples were collected from a group of 178. Three entire replications were sampled at each of three separate trials (one treatment was in Holly Hill, SC and two were in Oliver, GA). The replications contained between 216 and 220 ramets. Approximately 1700 cores were taken. The purpose of sampling from a large number of clones was to provide a diverse group from which to select a subset for harvest and subsequent enzymatic hydrolysis. The criteria used to select this subset will be described below.

Each wood sample consisted of a 12 mm wood core taken at approximately breast height (1.4 m). Avoidance of compression wood and branch scars was prioritized over the taking of each core at precisely breast height. Entire cores were oven-dried for a minimum of two days. Half of each core (minus the bark and first year of juvenile wood) was ground in a Wiley mill using a 20-mesh screen. Prior to being ground, the half-cores were cut into pieces using hand clippers in order to facilitate the loading of the samples into the mill. The powdered samples were subjected to near-infrared spectroscopy for reflectance wavelengths (using a FOSS NIRSystems model 6500). A spinning sample cup for reduced sample sizes was used. Data were collected at two nanometer intervals in the near-infrared range from 1100 to 2498 nm. The powdered samples were stored in small scintillation vials in boxes of 100. The vials were uncapped and placed in the laboratory where the NIR spectrometer was located for a minimum of 48 hours before scanning. This was done to allow the samples to

reach equilibrium with the humidity and temperature conditions present inside the spectrometer.

Best linear unbiased predictions (BLUPs) for each wavelength for each clone were compiled to produce a predicted spectrum for each clone. This analysis was done using ASReml 2.0 (Gilmour et al. 2006) to fit the following mixed effects model for each wavelength for all clones:

$$Y_{ijklmn} = \mu + S_i + R(S)_{ij} + B_k + M_l + C_m + \varepsilon_{ijklmn} \quad (1)$$

where:

Y_{ijklmn} is the n th observation of the m th clone in the j th rep,

μ is the overall mean,

S_i is the effect,

$R(S)_{ij}$ is the j th fixed replication effect within the i th site,

B_k is the k th fixed box effect,

M_l is the l th fixed mill effect,

C_m is the m th random clone effect $\sim \text{NID}(0, \sigma_g^2)$,

ε_{ijklmn} is the random error term $\sim \text{NID}(0, \sigma_e^2)$.

The samples were processed by the box, so the box effect was included to account for drift in the laboratory equipment over time. The mill effect was included because two Wiley Mills were used for processing the samples. ASReml incorporates pedigree into the analysis in order to construct a genetic relationship matrix. This serves to adjust clonal BLUPs based on the performance of relatives. Using the variance components produced by this analysis, broad-sense heritability (clone mean repeatability) for each wavelength was calculated as:

$$H^2 = \sigma_g / (\sigma_g + \sigma_e/n) \quad (2)$$

Where H^2 is the broad-sense heritability, σ_g is the genetic variance, σ_e is the error variance, and n is the mean number of ramets per clone. σ_g is the variation due to clone, and combines both heritable and non-heritable genetic effects. As such, the clone mean repeatability is informative about the performance of clonal genotypes, but will not be informative about how much of that variance is heritable. σ_e is the portion of the variance which is unexplained by the model.

These predicted clonal spectra were then subjected to a hierarchical cluster analysis. The software package R was used for this analysis (R Core Team 2012). First, a matrix of Spearman correlations between predicted clonal spectra was created using the *cor()* function. Then, using the *dist()* function, a distance matrix was created which centered these correlations around zero. The dimensions of the correlation and distance matrices were 178 by 178, comparing every clone to all the other clones. The hierarchical clustering was done

using the distance matrix and the *hclust()* function. This agglomerative clustering process starts with all the clones being in their own clusters. At each step of the clustering process, the two clusters with the smallest distance between them are merged. The complete linkage method was used, meaning that when the distance matrix is recalculated after clusters are merged, the distance between a new cluster and others is dependent on the maximum difference between that cluster's members and the other clusters (El-Hamdouchi and Willett 1989).

The resulting clusters were used to inform the design of a harvesting strategy to select clones for enzymatic hydrolysis. The goal was to choose clones that encompass the variation available in the sampled clonal trials. Since NIR spectra reflect chemical and physical wood properties, the clustering should have produced groups of clones that are similar for some combination of these properties. If a cluster can be identified that is superior for sugar yield, this may provide a quick way of identifying superior genotypes for sugar production. To have been considered for harvesting and testing for conversion efficiency, a clone must have had samples from greater than at least three ramets; 151 clones met this criterion.

Lignin and cellulose contents were determined using a previously derived calibration curve based on a comparison of NIR spectra to the results of a wet chemistry analysis of actual lignin and cellulose contents (see Hodge et al. 2004 for details on the development of this calibration curve). Hodge and Woodbridge (2010) discuss the development of this model to predict wood properties across pine species.

Using the cluster analysis and cellulose and lignin content predictions, 23 clones were selected to be harvested and converted using enzymatic hydrolysis. Selections were spread across six clusters (three clones for cluster one, four each for clusters two through six). The 23 selections also captured a cross section of variation in lignin and cellulose content (Table 1; Figure 4). Three ramets of each clone selected were sampled in order to obtain a better representation of the clone. However, due to time and labor constraints, the ramets for each clone were bulked into a single sample for the enzymatic hydrolysis experiment. A 2.4 m bolt from the lower bole of each ramet was collected from the field. These bolts were harvested from a single site in Holly Hill, SC. The bolts were debarked and chipped, and these chips were then mixed to obtain the clonal representative sample for use in the enzymatic hydrolysis experiment. Unfortunately, several samples were lost during the processing phase, so data was obtained for only 17 out of the original 23 clones.

Prior to enzymatic hydrolysis, three different pretreatments were tested: dilute acid, alkaline, and alkaline with mechanical beating (10,000 rotations in a PFI-type beater). The purpose of pretreatment is to release the cellulose from the lignin, thereby making the cellulose available for enzymatic hydrolysis. Additionally, multiple levels of enzyme loading were tested: One level of enzyme (20 FPU, or filter paper units) was used for the dilute acid pretreatment and three levels of enzyme (10, 20, and 40 FPU) were used for the alkaline and alkaline with mechanical beating pretreatments, for a total of seven treatments. Each treatment was applied to the bulked mixture of wood chips representing each clone. Wood properties for these samples, including lignin %, cellulose %, total carbohydrate %, extractive %, ash %, and

coarseness (mg/m), and fiber length (mm), were measured using wet chemistry. The bolt processing, wet chemistry, and enzymatic hydrolysis were performed at the Department of Forest Biomaterials at NC State University.

In addition, 31 of the original powdered wood cores, selected for divergent (high and low) lignin contents were also subjected to wet chemistry analysis of wood properties, in order to test the accuracy of the NIR predictions and to see how much variation in lignin content existed in the population. The TTEST procedure in SAS (SAS 9.2, SAS Institute, Cary, NC) was used to test the significance of the difference between the high- and low-lignin content cores, and the CORR procedure was used to check correlations between predicted and measure lignin and cellulose contents, as well as lignin content and sugar yields.

A statistical model was fit for each treatment with sugar yield as the dependent variable. Independent variables tested for significance included NIR cluster, lignin %, cellulose %, total % carbohydrate, extractive %, coarseness (mg/m), and fiber length (mm). Insignificant effects were removed from the models, unless their inclusion helped the significance of other model parameters. Due to the loss of the several of the samples, one of the clusters was left with only one observation. This data point was removed so that the clusters in the mean separation tests would all have more than one observation. The GLM procedure in SAS (SAS 9.2, SAS Institute, Cary, NC) was used to fit the linear models:

$$Y_{ij} = \mu + C_i + \beta_1 L_{j,1} + \beta_2 U_{j,2} + \beta_3 O_{j,3} + \beta_4 E_{j,4} + \beta_5 R_{j,5} + \beta_6 F_{j,6} + \varepsilon_{ij} \quad (3)$$

where:

Y_{ij} is the j th observation in the i th cluster,

C_i is the i th fixed cluster effect,

$L_{j,1}$ is the percent lignin covariate,

$U_{j,2}$ is the percent cellulose covariate,

$O_{j,3}$ is the percent total carbohydrate covariate,

$E_{j,4}$ is the percent extractive covariate,

$R_{j,5}$ is the fiber coarseness (mg/m) covariate,

$F_{j,6}$ is the fiber length (mm) covariate,

ε_{ij} is the random error term $\sim \text{NID}(0, \sigma_e^2)$.

The Waller multiple range test was used to test cluster means for significant differences.

Results

Mean sugar yields were calculated for each treatment and ranged from 210 to 521 mg sugar/mg wood (Table 1). The lowest-yield treatment was the one using a dilute-acid pretreatment, and the highest yielding treatment was that with the alkaline pretreatment, mechanical beating, and the highest level of enzyme (40 FPU). Analysis of the treatment means indicates a significant difference between the dilute acid treatment and the green liquor treatment that used the same amount of enzyme (DA20 and ALK20). Increasing enzyme levels consistently produced greater yields of sugars for both the mechanical beating and non-mechanical beating pretreatments. Use of mechanical beating increased sugar yields

by more than 100 mg/g wood compared to treatments with equivalent levels of enzyme that did not use mechanical beating.

Predicted lignin content for all samples ranged from 25.7 to 27.9%, with a mean of 27% (Table 2a). For predicted cellulose contents, the range was 40.1 to 42%, with a mean of 41.1%. Lignin and cellulose contents for the stemwood samples, as determined through benchtop wet chemistry measurements, had wider ranges and higher standard deviations (Table 3) than were predicted from the NIR spectra (Table 2b). There was no correlation between measured and NIR-predicted lignin content for these samples ($r = 0.23$, $p = 0.37$) (Figure 5). This was also the case for the correlation between measured and NIR-predicted cellulose content ($r = -0.04$, $p=0.89$).

However, for a group of 31 powdered wood cores that also underwent enzymatic hydrolysis, the correlations between predicted and measured contents were better: For lignin content, the correlations between measured and predicted were $r = 0.60$ ($p=.067$) and $r = -0.48$ ($p=0.027$) for the high- and low-lignin samples, respectively (Figure 6). A t-test indicated that the high- and low-lignin groups were significantly different for predicted and measured lignin content ($p<0.0001$ in both cases). For cellulose content, the correlation was $r = 0.63$ ($p = 0.0001$) (Figure 6). These powdered samples were the same samples from which the NIR spectra were obtained, so the stronger relationships between the measured and predicted contents for these samples compared to the stemwood samples are expected.

The correlations between lignin content and sugar yields for each treatment were calculated to test if lignin had an obvious impact on sugar yields. However, the resulting correlations were moderately- to highly-non-significant (p-values ranged from 0.092 to 0.587). Only one treatment, ALK40PFI, had a negative correlation between lignin content and sugar yield ($r = -0.40$, $p = 0.108$).

The analysis of the NIR spectra indicates a range of 0.23 to 0.60 for clone mean repeatability for each individual wavelength across all clones (Figure 2). This suggests that some wavelengths are likely subject to greater genetic determination than others. This information was not used in the hierarchical clustering, however, as the BLUP-generated spectra already include the genetic component of variation. Six clusters were used for grouping the clones (Figure 3). This allowed for a sample size of 3-4 clones per cluster to be harvested for the conversion phase of the experiment, which kept the overall size of the experiment at a reasonable level. Predicted lignin and cellulose contents were also used to pre-screen the selected clones to maximize variation for the conversion experiment. The selected clones represent a good cross section of the available variation in cellulose and lignin content (Table 2a & b, Figure 4a & b).

Models with overall p-values < 0.05 were found for three of the seven treatments (Table 4). Two of the models for other treatments just missed this significance threshold. In general, r^2 increased and significance of individual factors increased as the treatment level became more intense (i.e., increasing enzyme level and adding mechanical beating pretreatment). Wood

composition and wood property factors were significant for many of the models. Cluster was also significant for all but one treatment: DA20, which was the lowest yielding treatment.

Mean separation tests indicate statistically significant pairwise differences between clusters within five of the treatments (Table 5). Cluster differences were not assessed for dilute acid as the term was removed from the model for that treatment. The ALK20PFI and ALK40PFI treatments are the only treatments for which one cluster is significantly different from all others. For all treatments, cluster three is the top ranking cluster. There is no obvious pattern to the rankings of the other clusters, 1, 2, 4, & 5 (6 was omitted as it had only one observation).

Discussion

As expected, increasing enzyme dose and applying mechanical beating to the samples increases sugar yields. It is notable, however, that for alkaline pretreatments both with and without mechanical beating, the increase of 10 FPU (i.e., going from 10 to 20FPU) causes an equal or slightly greater increase in sugar yield compared to the increase of 20FPU (i.e., going from 20 to 40 FPU). The samples averaged 62% total carbohydrate content (Table 3). The treatments with no mechanical beating - DA20, ALK10, ALK20, and ALK40 - averaged 210, 228, 332, and 414 mg sugar/mg wood, respectively (Table 4), which equals 34%, 37%, 54%, and 67% total conversion of the carbohydrate fraction, respectively. For the treatments with mechanical beating – ALK10PFI, ALK20PFI, and ALK40PFI – the average sugar yields were 341, 445, and 521 mg sugar/g wood, with corresponding total

carbohydrate conversions equal to 55%, 72%, and 84%, respectively. This indicates diminishing returns from increasing enzyme use, which is expected as there is a limited pool of sugars on which the enzymes can act. Additionally, those sugars not removed by the lower enzyme doses are likely to be more difficult to extract from the lignin structure.

NIR cluster as well as a number of wood property measures were significant for predicting sugar yields. It is not surprising that the wood properties would have a significant impact on sugar yields: cellulose and total carbohydrate indicate the amount of sugars present within the wood, and lignin content, coarseness, and fiber length would tend to impact the availability of these carbohydrates. It is interesting that the significance of the models as well as of individual predictors generally increases as more intensive treatments are used. One interpretation of this trend is that as sugar yields increase, factors that affect these yields tend to become easier to detect. For several of the models, a relatively non-significant parameter was included in order to improve the significance of other model parameters (Table 4). This occurred with the fiber coarseness parameter for the ALK20 and ALK40 models. Removal of this parameter caused previously significant parameters to become non-significant, and decreased overall model significance. Removing additional parameters did not improve the model. The implication of this is that sugar yield is a complex trait that is affected by a number of different wood characteristics. It is encouraging that this experiment, with a relatively small number of data points for each treatment, was able to detect some of these effects.

The significance of cluster in most of the models, which also contain these wood property measures, indicates that cluster represents some other trait or combination of traits that are not present in the models. Since these NIR clusters were derived from BLUP predicted spectra, which in turn are a representation of physical and chemical wood properties, there can be some expectation that the trait or traits represented by cluster have a heritable component. This is reflected by the moderate broad-sense heritabilities found for wavelengths across the NIR spectrum (Figure 2). In addition, many of the wood properties that are significant predictors for sugar yield are known to be heritable. This has positive implications for hopes to increase sugar yields through genetic improvement.

The wet chemistry measurements for lignin and cellulose are, as expected, significant predictors of sugar yields. However, for the 17 clonal bulked samples used for the conversion experiment, the correlations between the NIR-predicted and the wet chemistry-determined lignin and cellulose contents are surprisingly low. However, these correlations are much better for a set of 31 powdered samples from the initial NIR screening phase of the experiment. Several factors may be involved in this. First, the 31 powdered samples were from the set of powders that were used to produce the NIR spectra. In contrast, while the set of 17 clonal bulks may have included some of the same trees as were included in the NIR analysis, the samples themselves were chips taken from the lower eight feet of stemwood. As such, it is not surprising that the wet chemistry results would be better correlated to the 31 powdered samples. Also, great care was taken to select cores composed of clear wood, and only the better half of each core was used for the powdered sample. In comparison, wood

defects (e.g. branch scars, compression wood) were not eliminated from the chipped stemwood samples, which would increase the differences between these solid-wood samples and powdered samples. Another possibility is that the small number of data points (17) contributed to the poor relationship between the predictions and the wet chemistry measurements.

Conclusion

There is significant improvement in extracting sugars for fermentation to ethanol from using an alkaline pretreatment instead of a dilute acid pretreatment. Increasing enzyme loads also improves sugar yields, but there appears to be a point of diminishing returns as more of the easily extracted sugars are released and additional enzyme additions are less effective.

Clusters derived from NIR spectra were, along with several other wood properties, significant for explaining variation in sugar yield. This is encouraging, in that it implies that there are heritable traits in loblolly pine that can likely be manipulated to improve conversion efficiency.

Literature Cited

- Antti, H., Sjoström, M., and L. Wallback. 1996. Multivariate calibration models using NIR spectroscopy on pulp and paper industrial applications. *Journal of Chemometrics* 10(5-6):591-603.
- Aspinwall, M.J., Li, B., McKeand, S.E., Isik, F., and Gumpertz, M.L. 2010. Prediction of whole-stem α -cellulose yield, lignin content, and wood density in juvenile and mature loblolly pine. *South. J. Appl. For.* 34(2):84-90.
- Baltunis, B.S., Wu, H.X., and Powell, M.B., 2007. Inheritance of density, microfibril angle, and modulus of elasticity in juvenile wood of *Pinus radiata* at two locations in Australia. *Can. J. For. Res.* 37:2164-2174.
- Boussaid, A. and J.N. Saddler. 1999. Adsorption and activity profiles of cellulases during the hydrolysis of two Douglas fir pulps. *Enzyme and Microbial Technology* 24:138-143.
- Burkheisser, E.V. ed. 2010. Biological barriers to cellulosic ethanol. Nova Science Publishers Inc., New York. 251 p.
- Chaffey, N. 2000. Microfibril orientation in wood cells: new angles on an old topic. *Trends in Plant Science* 5:360-362.
- Chiang, V.L. 2006. Monolignol biosynthesis and genetic engineering of lignin in trees, a review. *Environ Chem Lett* 4:143-146.
- Courchene, C.E., Peter, G.F., Litvay, J. 2006. Cellulose microfibril angle as a determinant of paper strength and hygroexpansivity in *Pinus taeda* L. *Wood and Fiber Science* 38:112-120.
- Dimmel, D.R., MacKay, J.J., Courchene, C.E., Kadla, J.F., Scott, J.T., O'Malley, D.M., McKeand, S.E. 2002. Pulping and bleaching of partially CAD-deficient wood. *Journal of Wood Chemistry and Technology* 22:235-248.
- Gilmour, A.R., Gogel, B.J., Cullis, B.R., and Thompson, R. 2006. ASReml User Guide Release 2.0. VSN International Ltd, Hemel Hempstead, HP1 1ES, UK. 342p.
- Gregg, D.J. and Saddler, J.N. 1996. Factors affecting cellulose hydrolysis and the potential of enzyme recycle to enhance the efficiency of an integrated wood to ethanol process. *Biotechnology and bioengineering* 51(4):375-383.
- Gressel, J. 2008. Transgenics are imperative for biofuel crops. *Plant Science* 174:246-263.

- Grethlein, H.E., D.C. Allen, and A.O. Converse. 1984. A comparative study of the enzymatic hydrolysis of acid-pretreated white pine and mixed hardwood. *Biotechnology and Bioengineering* 26:1498-1505.
- Hein, P.R.G., Bouvet, J.M., Mandrou, E., Vigneron, P., Clair, B., and Chaix, G. 2012. Age trends of microfibril angle inheritance and their genetic and environmental correlations with growth, density and chemical properties in *Eucalyptus urophylla* S.T. Blake wood. *Annals of Forest Science* 69:681-691.
- Hidayat, B.J., Felby, C., Johansen, K.S., and Thygesen, L.G. 2012. Cellulose is not just cellulose: a review of dislocations as reactive sites in the enzymatic hydrolysis of cellulose microfibrils. *Cellulose* 19:1481-1493.
- Hodge, G.R., and Purnell, R.C. 1993. Genetic parameter estimates for wood density, transition age, and radial growth in slash pine. *Can. J. For. Res.* 23:1881-18921
- Hodge, G.R., Woodbridge, W., and Dvorak, W.S. 2004. *Use of NIR to predict cellulose content of pine wood* In: Camcore 2004 Annual Report. Dept. of Forestry and Environmental Resources, NCSU, Raleigh, NC. pp 28-30.
- Hodge, G.R., and Woodbridge, W.C. 2010. Global near infrared models to predict lignin and cellulose content of pine wood. *J. Near Infrared Spectrosc.* 18:367-380.
- Hu, G., Heitmann, J. A., and Rojas, O. J. 2008. Feedstock pretreatment strategies for producing ethanol from wood, bark, and forest residues. *BioResources* 3(1):270-294.
- Isik, F., Gumpertz, M., Li, B., Goldfarb, B., and Sun, X. 2008. Analysis of cellulose microfibril angle using a linear mixed model in *Pinus taeda* clones. *Can. J. For. Res.* 38:1676-1689.
- Karlsson, O. and U. Westermarck. 1996. Evidence for chemical bonds between lignin and cellulose in kraft pulps. *J. Pulp Paper Sci.* 22:397-401.
- Kelley, S.S., Rials, T.G., Groom, L.R., and So, C. 2004a. Use of near infrared spectroscopy to predict the mechanical properties of six softwoods. *Holzforschung* 58:252-260.
- Kelley, S.S., Rials, T.G., Snell, R., Groom, L.H., and Sluiter, A. 2004b. Use of near infrared spectroscopy to measure the chemical and mechanical properties of solid wood. *Wood Science and Technology* 38(4):257-276.
- Lee, D., A.H.C. Yu, and J.N. Saddler. 1994. Evaluation of cellulase recycling strategies for the hydrolysis of lignocellulosic substrates. *Biotechnol. Bioeng.* 45:328-336.

- Lima, J.T., Breese, M.C., and Cahalan, C.M. 2004. Variation in microfibril angle in *Eucalyptus* clones. *Holzforschung* 58:160-166.
- Loo, J.A., Tauer, C.G., and McNew, R.W. 1985. Genetic variation in the time of transition from juvenile to mature wood in loblolly pine (*Pinus taeda* L.). *Silvae genetica* 34(1):14-19.
- Lu, Y., B. Yang, D. Gregg, J.N. Saddler, and S.D. Mansfield. 2002. Cellulase adsorption and an evaluation of enzyme recycle during hydrolysis of steam-exploded softwood residues. *Appl. Biochem. Biotechnol.* 98-100:641-54.
- Mabee, W.E., D.J. Gregg, C. Arato, A. Berlin, R. Bura, N. Gilkes, O. Mirochnik, X. Pan, K. Pye and J.N. Saddler. 2006 Updates on softwood-to-ethanol process development. *Applied Biochemistry and Biotechnology* 129–132:55-70.
- Martens, H., and Jensen, S.A. 1983. Partial least squares regression: A new two-stage NIR calibration method. In *Proc. 7th World Cereal and Bread Congress*, Prague, June 1982. Eds. Holas and Kratochvíl. Elsevier Publ., Amsterdam. pp. 607-647.
- Martin, M., and J. Aber. 1994. Analyses of forest foliage III: determining nitrogen, lignin, and cellulose in fresh leaves using near infrared reflectance data. *J Near Infrared Spec* 2:25-32.
- McLellan, T., Aber, J., Martin, M., Melillo, J., and Nadelhoffer, K. 1991. Determination of nitrogen, lignin, and cellulose content of decomposing leaf material by near infrared reflectance spectroscopy. *Can. J. For. Res.* 21:1684-1688.
- Mooney, C.A., S.D. Mansfield, M.G. Touhy, and J.N. Saddler. 1998. The effect of initial pore volume and lignin content on the enzymatic hydrolysis of softwoods. *Bioresource Technology* 64:113–119.
- Mousdale, D.M. 2008 *Biofuels: Biotechnology, Chemistry, and Sustainable Development*. Boca Raton : CRC Press. 424 p.
- Myszewski, J.H., Bridgwater, F.E., Lowe, W.J., Byram, T.D., and Megraw, R.A. 2004. Genetic variation in the microfibril angle of loblolly pine from two test sites. *South. J. Appl. For.* 28(4):196-204.
- Overend, R.P. 2004. Thermochemical conversion of biomass. In: Shpilrain, E.E. (ed.), *Renewable Energy Sources Charged with Energy from the Sun and Originated from Earth-Moon Interaction*, in *Encyclopedia of Life Support Systems (EOLSS)*. Developed under the Auspices of the UNESCO, Eolss Publishers, Oxford ,UK.

Pan, X.J., Xie, D., Gilkes, N., Gregg, D.J., and Saddler, J.N. 2005. Strategies to enhance the enzymatic hydrolysis of pretreated softwood with high residual lignin content. *Applied Biochemistry and Biotechnology* 121-124:1069-1079.

Pandey, M.P., and Kim, C.S. 2011. Lignin depolymerization and conversion: a review of thermochemical methods. *Chemical Engineering and Technology* 43(1):29-41.

Pauly, M., & Keegstra, K. 2008. Cell-wall carbohydrates and their modification as a resource for biofuels. *The Plant Journal* 54:559–568.

Pettersen, R.C. 1984. The Chemical Composition of Wood. In: *The Chemistry of Solid Wood* ed. R.M. Rowell. The American Chemical Society, Washington DC. 614p.

Ramos, L.P., C. Breuil, and J.N. Saddler. 1992. Comparison of steam pretreatment of eucalyptus, aspen, and spruce wood chips and their enzymatic hydrolysis. *Applied Biochemistry and Biotechnology* 34-35:37-48.

Saha, B.C. 2003. Hemicellulose bioconversion. *J. Ind. Microbiol. Biotechnol.* 30:279-291.

Shelbourne, C.J.A., and Stonecypher, R.W. 1971. The inheritance of bole straightness in young loblolly-pine. *Silvae genetica* 20:151-156.

Shelbourne, C.J., Zobel, B.J., and Stonecypher, R.W. 1969. The inheritance of compression wood and its genetic and phenotypic correlations with six other traits in 5-year-old loblolly pine. *Silvae Genetica* 18:43-47.

Shenk, J.S., and M.O. Westerhaus. 1991. Population definition, sample selection and calibration procedures for near-infrared reflectance spectroscopy. *Crop Sci.* 31:469-474.

Shimada, K., S. Hosoya, and T. Ikeda. 1997. Condensation reactions of softwood and hardwood lignin model compounds under organic acid cooking conditions. *Journal of Wood Chemistry and Technology* 17(1-2):57-72.

Sjöström, E. 1993. *Wood Chemistry*. Academic Press, San Diego. 293 p.

Stenius, P., ed. 2000. *Forest Products Chemistry*. Fapet Oy, Helsinki, Finland. 350p.

Sykes, R., Li, B., Hodge, G., Goldfarb, B., Kadla, J., and Chang, H.-m. 2005. Prediction of loblolly pine wood properties using transmittance near-infrared spectroscopy. *Can. J. For. Res.* 35:2423-2431.

- Sykes, R., Li, B., Isik, F., Kadla, J., and Chang, H.M. 2006. Genetic variation and genotype by environment interactions of juvenile wood chemical properties in *Pinus taeda*. *Ann. For. Sci.* 63: 897–904.
- Thygesen, L.G., Hidayat, B.J., Johansen, K.S., and Felby, C. 2011. Role of supramolecular cellulose structure in enzymatic hydrolysis of plant cell walls. *J. Ind. Microbiol. Biotechnol.* 38:975-983.
- Vanholme, R., Demedts, B., Morreel, K., Ralph, J., and Boerjan, W. 2010. Lignin biosynthesis and structure. *Plant Physiology* 153:895-905.
- Wessman, C.A., Aber, J.D., Peterson, D.L., and J.M. Melillo. 1988. Foliar analysis using near infrared reflectance spectroscopy. *Can. J. For. Res.* 18:6-11.
- Whetten, R., and Sederoff, R. 1991. Genetic engineering of wood. *Forest Ecology and Management* 43:301-316.
- Wright, J.A., Birkett, M.D., and M.J.T. Gambino. 1990. Prediction of pulp yield and cellulose content from wood samples using near infrared reflectance spectroscopy. *Tappi J.* 73:164-166.
- Wu, R.L., Remington, D.L., MacKay, J.J., McKeand, S.E., O'Malley, D.M. 1999. Average effect of a mutation in lignin biosynthesis in loblolly pine. *Theor Appl Genet* 99:705-710.
- Yu, A.H.C., D. Lee, and J.N. Saddler. 1995. Adsorption and desorption of cellulase components during the hydrolysis of a birch substrate. *Biotechnol. Appl. Biochem.* 21:203–216.
- Yu, Q., Li, B., Nelson, C.D., McKeand, S.E., Batista, V.B., Mullin, T.J. 2006. Association of the cad-n1 allele with increased stem growth and wood density in full-sib families of loblolly pine. *Tree Genetics & Genomes* 2:98–108.
- Yu, Q., McKeand, S.E., Nelson, C.D., Li, B., Sherrill, J.R., Mullin, T.J. 2005. Differences in wood density and growth of fertilized and nonfertilized loblolly pine associated with a mutant gene, cad-n1. *Can. J. For. Res.* 35:1723-1730.
- Yu, Z., H. Jameel, H. Chang, and S. Park. 2011. The effect of delignification of forest biomass on enzymatic hydrolysis. *Bioresource Technology* 102:9083-9089.
- Zobel, B.J., and Blair, R. 1976. Wood and pulp properties of juvenile wood and topwood of the southern pines. *Applied Polymer Symposium* 28:421-433.
- Zobel, B.J., and Haught, E.A. 1962. Effect of bole straightness on compression wood of loblolly pine. Tech. Rept. 15, School of Forestry, N.C. State University, Raleigh, NC. 5p.

Zobel, B.J., and van Buijtenen J.P. Wood Variation: Its Causes and Control. Springer-Verlag, Berlin. 1989.

Tables and Figures

Table 1. Sugar yield means, minimums, and maximums by treatment. Units are mg sugar/g wood. Treatment means with the same letter groupings are not significantly different. DA20 is dilute acid pretreatment with 20 FPU enzyme, ALK10 is an alkaline pretreatment with 10 FPU of enzyme, and so forth. ALK10PFI is an alkaline pretreatment with mechanical beating and 10 FPU. N = 17 for all treatments.

<i>Treatment</i>	<i>DA20</i>	<i>ALK10</i>	<i>ALK20</i>	<i>ALK40</i>	<i>ALK10PFI</i>	<i>ALK20PFI</i>	<i>ALK40PFI</i>
Mean	210 ^a	228 ^a	332 ^b	414 ^c	341 ^b	445 ^d	521 ^e
Std. dev.	20	53	63	60	61	66	35
Max	248	295	455	545	466	582	585
Min	173	136	234	310	258	339	446

Table 2. Summary statistics for lignin and cellulose content predictions computed from NIR spectral data;

a. For the full complement of clones (N=151) considered for harvesting.

<i>Predicted Value (%)</i>	<i>Mean</i>	<i>Std. Dev.</i>	<i>Minimum Value</i>	<i>Maximum Value</i>
Lignin	27.0	0.4	25.7	27.9
Cellulose	41.1	0.4	40.1	42.0

b. For the clones selected for harvest (N=23).

<i>Predicted Value (%)</i>	<i>Mean</i>	<i>Std. Dev.</i>	<i>Minimum Value</i>	<i>Maximum Value</i>
Lignin	27.0	0.5	26.0	27.9
Cellulose	41.0	0.3	40.4	41.6

Table 3. Summary of data for wet chemistry-determined properties.

	<i>Total lignin %</i>	<i>Cellulose %</i>	<i>Total Carbohydrates %</i>	<i>Extractives %</i>	<i>Ash %</i>	<i>Coarseness, mg/m</i>	<i>Fiber length, mm</i>
Mean	30.4	39.7	62.5	3.3	0.3	5.9	2.2
Std. dev.	1.4	1.0	1.0	0.5	0.1	1.9	0.3
Max	32.1	41.6	63.9	4.5	0.5	8.5	2.8
Min	28.1	37.8	60.7	2.5	0.2	2.6	1.9

Table 4. Summary of multiple linear regression models for each of seven treatments. The response variable in each model is the sugar yield produced by a given treatment. R^2 is the model coefficient of determination.

Treatment	Overall model significance (p-value)	R^2	Predictor significance level (p-value)						
			cluster	lignin %	cellulose %	total carbohydrate %	extractive %	coarseness (mg/m)	fiber length (mm)
DA20	0.11	0.29	.	.	.	0.12	0.07	.	.
ALK10	0.06	0.68	0.06	.	.	0.05	0.04	.	.
ALK20	0.04	0.92	0.02	0.04	0.03	0.01	0.01	0.12	0.06
AKL40	0.11	0.86	0.05	0.04	0.03	0.02	0.07	0.21	0.07
ALK10PFI	0.08	0.83	0.03	0.03	0.03	0.02	.	0.02	0.05
ALK20PFI	0.03	0.92	0.01	0.005	0.007	0.004	0.09	0.006	0.02
ALK40PFI	0.01	0.92	0.005	0.001	0.01	0.02	.	0.04	0.09

Table 5. Mean sugar yields by cluster for each treatment. DA20 treatment not shown as cluster was not included in the model for that treatment. Different letter groupings indicate significant differences between clusters.

ALK10		ALK20		ALK40		ALK10PFI		ALK20PFI		ALK40PFI	
Clust.	Mean	Clust.	Mean	Clust.	Mean	Clust.	Mean	Clust.	Mean	Clust.	Mean
3	268 ^a	3	381 ^a	3	458 ^a	3	399 ^a	3	507 ^a	3	559 ^a
5	238 ^{ab}	5	338 ^{ab}	1	395 ^a	5	337 ^{ab}	1	425 ^b	4	517 ^b
4	225 ^{ab}	4	324 ^{ab}	4	394 ^a	1	324 ^{ab}	2	425 ^b	2	513 ^b
1	213 ^{ab}	1	313 ^b	5	393 ^a	2	316 ^b	5	423 ^b	1	503 ^b
2	184 ^b	2	281 ^b	2	386 ^a	4	304 ^b	4	409 ^b	5	496 ^b

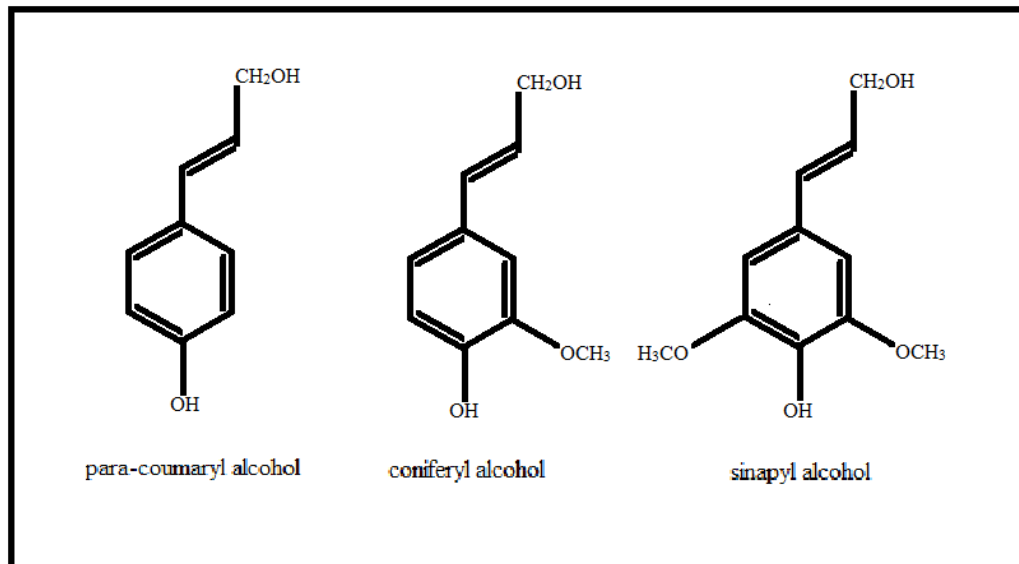


Figure 1. The three lignin monomer precursors. These precursors are produced via the shikimic and cinnamate metabolic pathways.

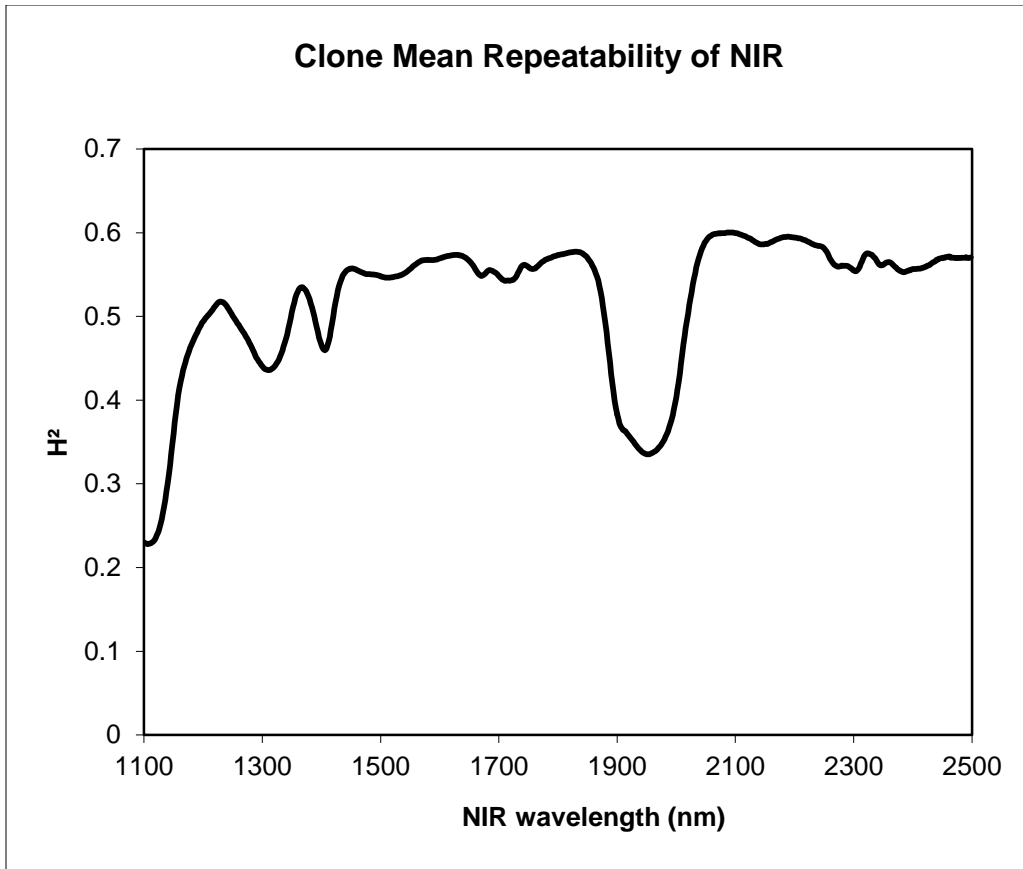


Figure 2. Clone mean repeatability (H^2 =broad sense heritability) for NIR spectra at wavelengths between 1100 and 2498 nm. Repeatabilities tend to increase from left to right, though there is a dip between 1900 and 2000 nm. This coincides with an increase in the error variance component.

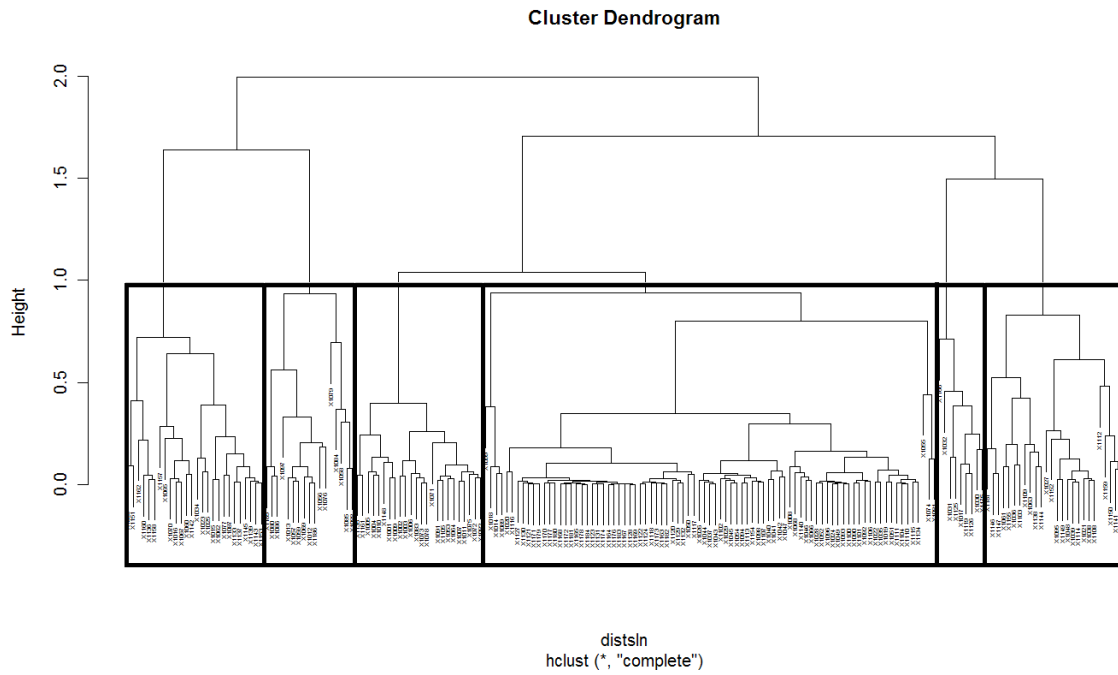
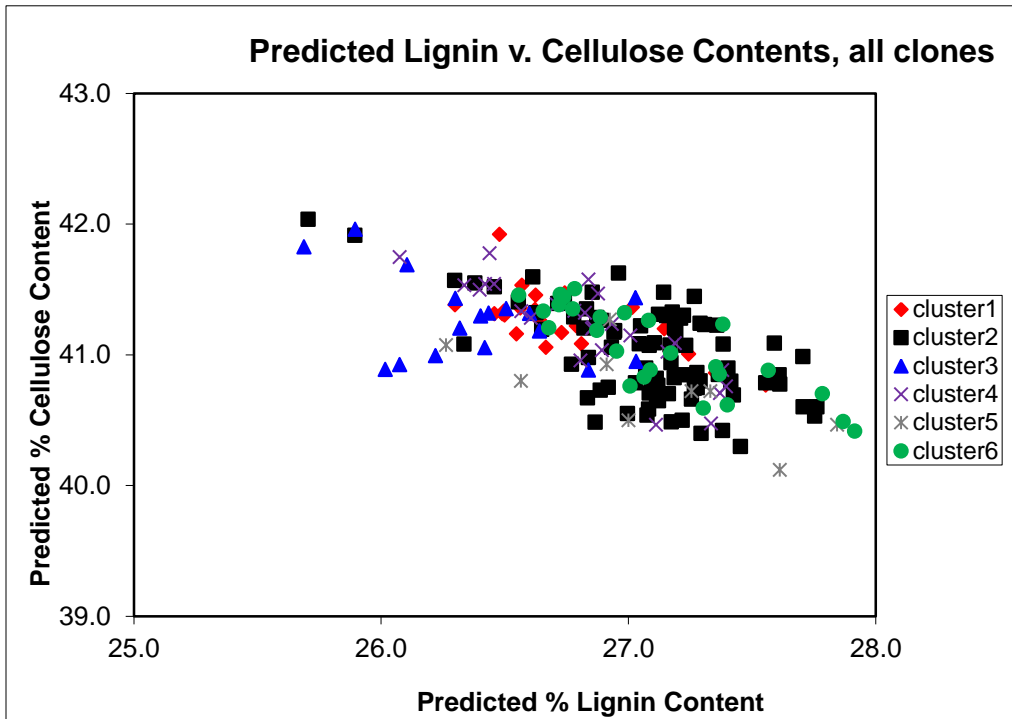


Figure 3. Dendrogram illustrating the hierarchical structure of the clusters. The black boxes indicate the clusters identified for use in this experiment.

A. All clones



B. Clones selected for harvest and conversion

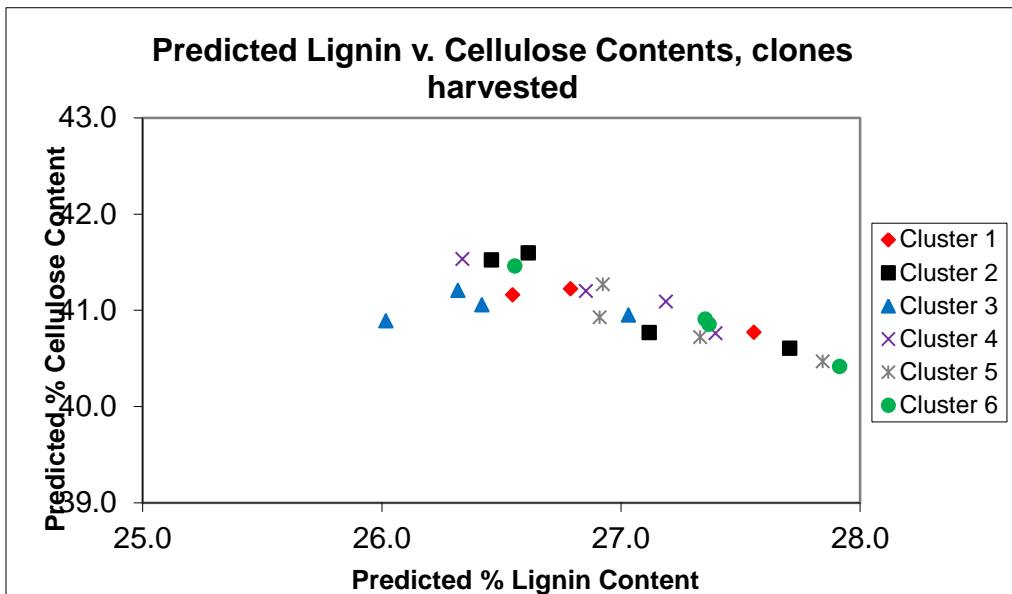
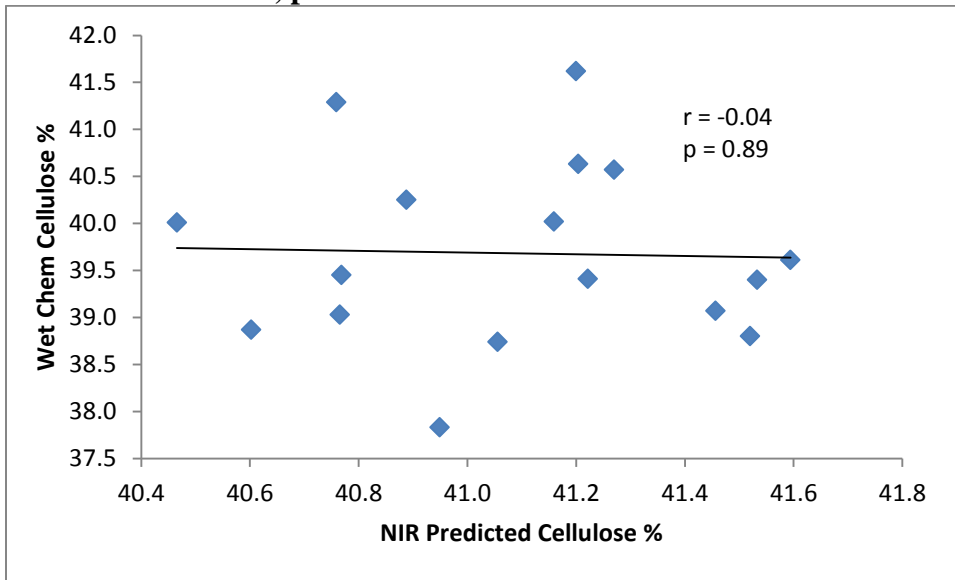


Figure 4. Predicted cellulose vs. predicted lignin with clusters indicated.

A. Cellulose, predicted vs. measured.



B. Lignin, predicted vs. measured.

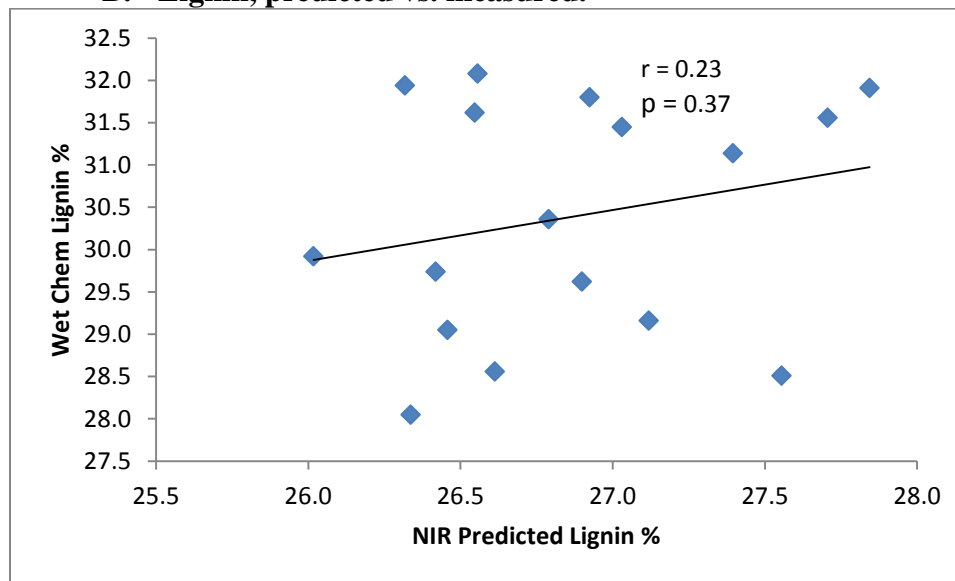
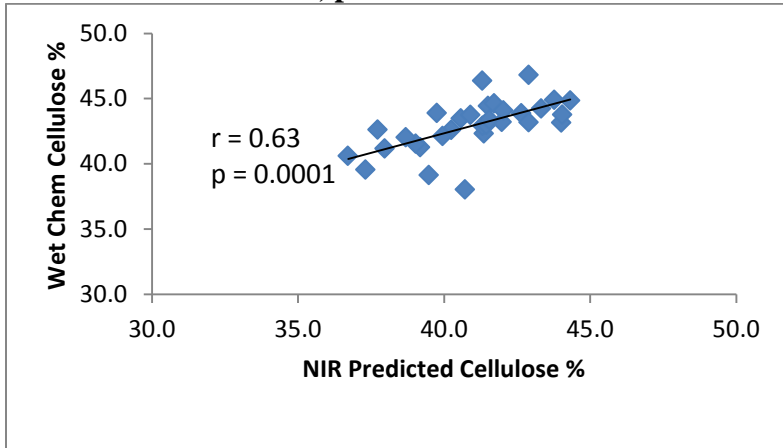
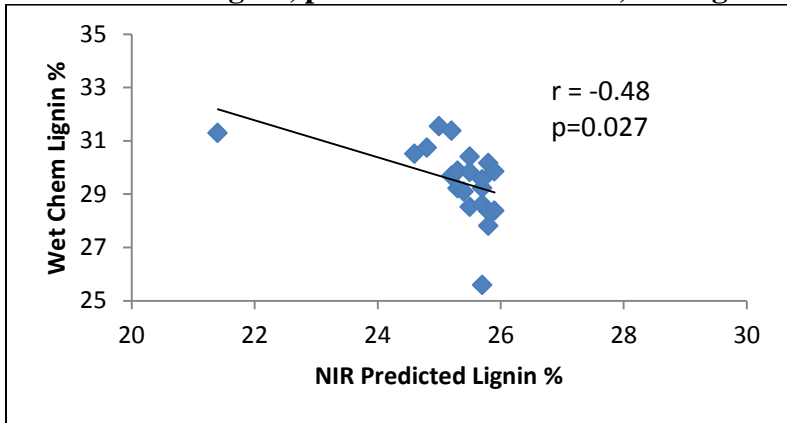


Figure 5. Predictions vs. measurements for cellulose and lignin for 17 loblolly pine samples derived from chipped bolts.

A. Cellulose, predicted vs. measured.



B. Lignin, predicted vs. measured, low-lignin samples.



C. Lignin, predicted vs. measured, high-lignin samples.

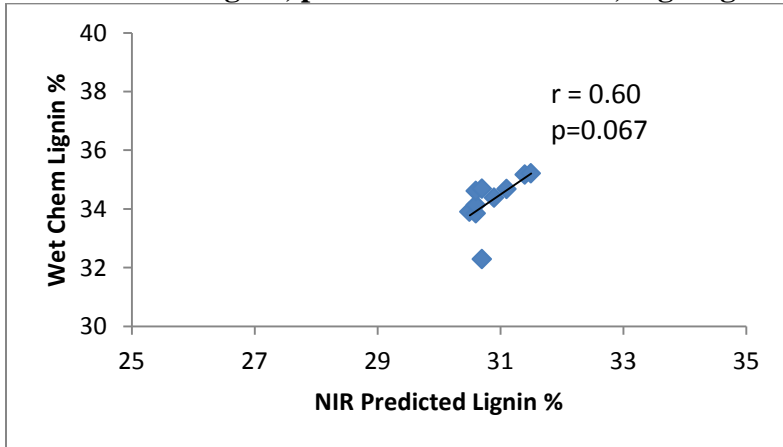


Figure 6. Lignin and cellulose contents, predicted and determined with wet chemistry, for 31 powdered loblolly pine samples.

CHAPTER 3

GENETIC VARIATION IN SUGAR YIELD FROM LOBLOLLY PINE (*PINUS TAEDA* L.) WOOD: II: HIGH-THROUGHPUT DILUTE ACID ENZYMATIC HYDROLYSIS OF POWDERED CORE SAMPLES.

Abstract

Woody biomass may be a good alternative energy source in the Southeastern U.S. due to its wide availability and its ability to grow on marginal sites. Loblolly pine (*Pinus taeda* L.), as the most productive forest tree species in the Southeastern U.S., is a logical species of interest for biofuel production in this region. However, loblolly pine biomass currently presents a challenge for producing ethanol; enzymatic hydrolysis of polysaccharides from softwood pulp typically produces lower yields of fermentable sugars than similar treatment of hardwood pulp. Many chemical and physical wood properties in loblolly pine are subject to genetic control, and variation in some of these properties will likely have an effect on the efficiency of ethanol production.

In order to characterize genetic variation in sugar yield from loblolly pine wood, 300 powdered wood samples from a clonal test series were tested using a high-throughput enzymatic hydrolysis process with a dilute acid pretreatment. The mean sugar yield was 0.21 mg sugar/mg wood, and ranged from 0.10 to 0.29 mg sugar/mg wood. Genetic values for the clones indicate that yields can be improved ~4-5% over the mean through the use of the best clones. Clone mean repeatability for sugar yields ranged from 0.29 to 0.44. For NIR-predicted lignin and cellulose contents, clone mean repeatabilities were 0.81 and 0.78, respectively. A calibration model was developed to predict sugar yields using NIR spectra. The overall R^2 was relatively low ($= 0.30$), and the RPD, or the ratio of performance to deviation, was 1.54 for the calibration model. This RPD is low, and indicates that the model may at best be accurate enough for a rough initial screening of new families.

Introduction

Biochemical techniques for the conversion of biomass to ethanol focus on hydrolytic processes that retrieve the sugars from the cellulose and hemicellulose. These techniques can use a variety of chemical and/or enzyme treatments to help loosen the grip of lignin on the carbohydrates and to break down those carbohydrates. Biochemical conversion is typically assisted by chemical and/or enzymatic agents, and fermentation of the sugars can be improved through specially tailored microbes (Mousdale 2008). For cellulosic ethanol, the main difficulty is extracting the valuable sugars from the biomass with as little input and as little sugar loss as possible. The structural properties of these lignocellulosic feedstocks are variable, which means that biochemical conversion techniques will need to be calibrated for specific feedstocks. Loblolly pine as a feedstock may well be improved through genetic screening and/or breeding, and would be a good choice as a feedstock given the sizable supply of pine biomass that already exists in the Southeastern U.S.

Our work on sugar extraction in loblolly pine using enzymatic hydrolysis with dilute acid or alkaline pretreatments identified variation of sugar yields that was related to NIR-based groupings of genetic entries (Chapter 2). These groupings were based on a hierarchical clustering of predicted NIR spectra for a set of clones. Additionally, linear modeling indicated that NIR cluster and a number of wood properties were significant for predicting sugar yields. The significance of cluster in most of the models, which also contain these wood property measures, indicates that cluster represents some other trait or combination of traits that are not already present in the models. Since these clusters were derived from NIR

spectra, which are a representation of physical and chemical wood properties, there can be some expectation that the trait or traits represented by cluster have a heritable component. In addition, many of the wood properties that are significant predictors for sugar yield are known to be heritable.

This study is an extension of the work in Chapter 2, which, in addition to the above-described findings concerning cluster and wood property effects on sugar yield, found that there is significant improvement to be made in extracting sugars for fermentation to ethanol depending on pretreatment and enzyme level used. The identification of a cluster that has higher sugar yields than others, along with the fact that wood properties (e.g. specific gravity, cellulose content, lignin content) are generally heritable, suggests that sugar yield may be under some degree of genetic control. However, given the logistics of processing and applying treatments to the stemwood samples in the underlying experiment, multiple measurements of clones were not possible. While a mixture of three ramets from each clone was used in order to have an “average” clonal mixture, the lack of replication meant that no conclusions were possible with regard to clonal effects on sugar yield. This study describes the use of a high-throughput method for enzymatic hydrolysis. This method was used to obtain multiple measurements of sugar yields for clones in order to characterize genetic variation in that trait.

Methods

Samples were collected in winter 2012 from a test in Oliver, GA that is part of a series of clonal trials conducted by Plum Creek Timber Company and CellFor Inc. The trees for this test were produced using somatic embryogenesis, a technique wherein plant embryos are cultured and bulked up to produce clonal replicates. The test site was laid out in an alpha-lattice, randomized incomplete-block design, with single-tree plots. Samples from five ramets each of a group of 30 clones were collected, for a total of 150 trees. The clones sampled were spread equally among six groups that were based on a previous hierarchical cluster analysis of NIR spectra from clones in the same test series (see Chapter 2). An additional 150 samples that were a part of that previous cluster analysis were also used for the enzymatic hydrolysis step in this experiment (as will be described below). Details on the harvesting and processing of those samples are available in Chapter 2.

Each wood sample consisted of a 12 mm wood core taken at approximately breast height. Avoidance of compression wood and branch scars was prioritized over the taking of each core at precisely breast height. Two cores were taken per ramet in order to get a better pooled sample to represent the tree. The pooled sample was created by mixing equal amounts of wood powder from the two cores samples from each tree. The cores were sampled from five reps, out of eight total. The three other reps had been sampled for a previous experiment, and so were excluded from this sample set. Entire cores were oven-dried for a minimum of two days. The entire core (minus the bark and first year of juvenile

wood) was ground in a Wiley Mini-Mill using a 20-mesh screen. Prior to being ground, the cores were cut into small discs using hand clippers in order to facilitate grinding.

Sample Characterization

The powdered samples were subjected to near-infrared reflectance spectroscopy using a FOSS NIRSystems model 6500. The primary purpose of this was to provide spectral data that could be used to develop a predictive model for sugar yield. A secondary purpose was to repeat the hierarchical cluster analysis from previous work on this clonal population, to see if cluster membership is consistent.

A spinning sample cup for reduced sample sizes was used. Data were collected at two nanometer intervals in the near-infrared range from 1100 to 2498 nm. The day before scanning, the samples were placed into borosilicate glass test tubes and re-dried to ensure consistent moisture content.

Best linear unbiased predictions (BLUPs) for each wavelength for each clone were compiled to produce a predicted spectrum for each clone. This analysis was done using ASReml 3.0 (Gilmour et al. 2009) to fit the following mixed effects model for each wavelength:

$$Y_{ijkl} = \mu + R_i + G_j + \beta_1 P_{i,1} + \beta_2 L_{i,2} + C_k + \varepsilon_{ijkl}, \quad (1)$$

where:

Y_{ijkl} is the l th observation of the k th clone in the i th rep,
 μ is the overall mean,
 R_i is the i th fixed replication effect,
 G_j is the j th fixed grind date effect,
 $P_{l,1}$ is the pith offset covariate,
 $L_{l,2}$ is the core length covariate,
 C_k is the k th random clone effect $\sim \text{NID}(0, \sigma_g^2)$,
 E_{ijkl} is the random error term $\sim \text{NID}(0, \sigma_e^2)$.

“Grind date” is included to account for changes in the Wiley Mill (e.g., blade sharpness, distension of the mesh screen over time) and “pith offset” (distance of the pith from the tree center) is included as it may help to account for compression wood from less straight stems.

These predicted clonal spectra were then subjected to a hierarchical cluster analysis in order to compare the consistency of cluster assignment with those from a previous cluster analysis (Chapter 2). The software package R was used for this analysis (R Core Team 2012). First, a matrix of Spearman correlations between predicted clonal spectra was created using the *cor()* function. Then, using the *dist()* function, a distance matrix was created which centered these correlations around zero. The dimensions of the correlation and distance matrices were 30 by 30, comparing every clone to all the other clones. The hierarchical clustering was done using the distance matrix and the *hclust()* function. This agglomerative clustering process starts with all the clones being in their own clusters. At each step of the clustering process,

the two clusters with the smallest distance between them are merged. The complete linkage method was used, meaning that when the distance matrix is recalculated after clusters are merged, the distance between a new cluster and others is dependent on the maximum difference between that cluster's members and the other clusters (El-Hamdouchi and Willett 1989).

Lignin and cellulose contents were determined using a previously derived calibration curve based on a comparison of NIR spectra to the results of a wet chemistry analysis of lignin and cellulose contents as described in Hodge et al. (2004). Hodge and Woodbridge (2010) discuss the development of this model to predict wood properties across pine species.

Pyrolysis-molecular beam mass-spectrometry (py-MBMS) was carried out on the 150 ground samples in order to get compositional data and to assess the variation within the sample set. This spectroscopic technique for evaluation of biomass was developed by Evans and Milne (1987) at the US Department of Energy's National Renewable Energy Lab (then called the Solar Energy Research Institute). Py-MBMS has been used for a number of applications, such as characterizing soil organic matter (Magrini et al. 2002, Plante et al. 2009), exploring lignin composition and structure in switchgrass (Mann et al. 2009, Hu et al. 2010), characterizing chemical changes in spruce wood due to brown rot biodegradation (Kelley et al. 2002), and exploring hydrocarbon formation during pyrolysis (Shin et al. 2001). Prediction of cell wall composition in loblolly pine and hybrid poplar was demonstrated by Davis et al. (1999). Kelley et al. (2004) compared PLS calibration predictions for lignin and

sugar content derived from both NIR and py-MBMS spectra.

For this analytical technique, a small amount of powdered sample is pyrolyzed. The pyrolyzed sample (now in a gaseous state) is carried through an ionizer and then the “beam” of pyrolyzed sample is focused onto an ion detector. The data produced are values for a range of mass-to-charge ratios (or atomic mass units). These values indicate the relative proportion of the sample accounted for by a particular mass-to-charge ratio. Many of the mass-to-charge ratios (or “peaks”) are associated with one or more particular molecules, so py-MBMS can be used to directly assess biomass composition (Evans and Milne 1987; Sykes et al. 2009).

The dried and powdered core samples were sent to the National Renewable Energy Laboratory in Golden, CO for py-MBMS analysis. This analysis resulted in spectra consisting of mean-normalized ion counts at mass/charge (m/z) ratios from 50 to 450. Principle component analysis (PCA) was used to help identify peaks that were responsible for variation within the MBMS spectra (i.e., those that with loadings). The PCA was conducted on the mean of two replications of each of the 150 trees that were sampled. The R function ‘prcomp’ was used to carry out the PCA (R Core Team 2012). ASReml 3.0 (Gilmour et al. 2009) was used to estimate variance components for the outlier peaks using the following linear model:

$$Y_{ijk} = \mu + C_i + R_j + E_{ijk}, \quad (2)$$

where:

Y_{ijk} is the k th observation of the i th clone in the j th rep,

μ is the overall mean,

C_i is the i th random clone effect \sim NID $(0, \sigma_g^2)$,

R_j is the j th fixed rep effect,

E_{ijk} is the random error term \sim NID $(0, \sigma_e^2)$.

The model used the pedigree structure to construct a genetic relationship matrix to adjust clonal predictions based on the performance of relatives. Clone mean repeatabilities were calculated as:

$$H^2 = \sigma_g^2 / (\sigma_g^2 + \sigma_e^2/n) \quad (3)$$

Where H^2 is the clone mean repeatability, σ_g^2 is the genetic variance, σ_e^2 is the error variance, and n is the mean number of ramets per clone.

Extractive and lignin content predictions were calculated from the MBMS spectra by colleagues at NREL. The predictions are based on the summation of peaks known to be associated with those components and a correction factor determined from wet-lab analysis (Sykes et al. 2009). These predictions, as well as the NIR-based lignin and cellulose content predictions, were also analyzed using the above model. Each model was fit with and without

the clone effect. A likelihood ratio test (LRT) was used to compare the full and reduced model in order to test the significance of the genetic variance component. LRT is computed as $2 * (\log\text{-likelihood}_{\text{fullmodel}} - \log\text{-likelihood}_{\text{reducedmodel}})$ (Gilmour et al. 2009). This statistic has a χ^2 distribution, with one degree of freedom since the models differ by only one parameter. At a confidence level = 0.05, the critical value for the statistic is 3.481. LRTs above this value indicate that the full model (i.e., with the genetic variance component) is significantly better than the reduced model. The CORR procedure in SAS (SAS 9.2, SAS Institute, Cary, NC) was used to produce a simple correlation between the MBMS- and NIR-predicted lignin contents.

Sugar yield assessment and prediction

In order to characterize the variation in sugar yield among the clones, 300 samples were subjected to enzymatic hydrolysis with a dilute acid pretreatment. These 300 samples are composed of the 150 samples collected in winter 2012 (representing 30 clones), as well as 150 additional samples from the work described in Chapter 2 that were collected in fall 2008 (both sample sets are summarized in Appendix A). The additional 150 samples represent 24 additional clones, with an average of 6.25 ramets sampled per clone (# ramets/clone ranged from three to seven). NIR spectra were also available for these additional samples. Three clones were present in both the newer and older sets of samples. Between the two sets of samples, a total of 51 clones were evaluated for sugar yield.

The high-throughput system used to obtain sugar yield data was developed at NREL in Golden, CO. This system can handle multiple 96-well plates and utilizes auto-sampling technology which reduces the time-intensive labor usually associated with getting wet-chemistry data (Decker et al. 2009, Selig et al. 2010). A significant time savings is also seen due to the use of enzyme-based assays to measure sugar release, instead of traditional high-pressure liquid chromatograph (HPLC) methods (Selig et al. 2011). Using these enzyme assays, a single 96-well plate can be processed in 5-6 hours; in comparison, 48 hours would be necessary to do an equal number of assays with an HPLC.

Prior to sending the samples to NREL for enzymatic hydrolysis, the samples were subjected to a 24-hour ethanol extraction using a Soxhlet extractor. 200 to 250 mg per sample were loaded into teabags, labeled in pencil, and rolled up (see Decker et al. 2012). Only a small fraction of the total extracted portion (~20 mg) was used for the enzymatic hydrolysis sugar extraction step, but a larger quantity was needed in order for the automated sampling system to function properly. Once removed from the extractor, the samples were placed under a fume hood to dry overnight. After this, the samples were sent to NREL, where they underwent a dilute sulfuric acid pretreatment and enzyme hydrolysis using the high throughput system discussed above. The pretreatment and hydrolysis steps were done sequentially, without washing or separation of sugars in between steps, so the sugar released from both steps was assessed collectively with one measurement at the end of the process.

Partial least squares (PLS; also referred to as projection to latent structures) was used to investigate the prediction of sugar yield using both the NIR and the py-MBMS spectra that are available for the samples. PLS is a statistical method for correlating laboratory measurements with spectral data (Martens and Jensen 1982). This technique can be better than stepwise multiple regression, providing predictions with lower standard errors (Shenk and Westerhaus 1991). PLS models, by identifying latent structures in the predictors to explain variation in the response variable, can take more wavelengths into account than can linear regression models and are better able to deal with correlations between neighboring wavelengths (i.e., it handles collinearity better).

Prior to the PLS analysis, the NIR data were smoothed using multiplicative scatter correction and a Savitzky-Golay second-derivative transformation, using 7 data points (Hodge and Woodbridge 2010). This smoothing serves to reduce the noise in the dataset. A 20% sample of the data (N= 60) was selected at random to serve as a test data set, and the remaining observations (N=239) were used as a training/calibration data set. Leave-one-out cross-validation was used on the training data set to develop the calibration model. With this method of cross-validation, a separate PLS model is calculated for each observation (which act as one-member test sets), so each observation in the training set is used for training and validation. The remaining 20% of the data is not used to develop the model, but is held in reserve to test the calibration model. The PLS procedure in SAS (SAS 9.2, SAS Institute, Cary, NC) was used to conduct the analysis.

The optimum number of factors to use in the PLS analysis is driven by minimizing the predicted residual sum of squares (PRESS). However, models that use fewer factors and have similar but slightly higher PRESS scores may be preferable, as this can reduce the chances of over-fitting the model. SAS provides a significance test of the differences between residuals of the different models. Using this statistic, the model chosen is the one with the fewest factors that also has residuals that are not significantly larger than the model with the lowest PRESS (SAS Institute Inc. 2008, van der Voet 1994).

Coefficients of determination (R^2) for the training and test sets as well as standard errors of calibration and prediction (SEC and SEP, respectively) are common ways to measure and compare prediction models (Kelley et al. 2002, Schimleck et al. 2003a, Hodge and Woodbridge 2010). SEC and SEP are calculated from the residuals of the training and test data sets, respectively. RPD (ratio of performance to deviation) is another statistic useful for evaluating prediction models, and is calculated as the ratio of the standard deviation of the measurements to the standard error of the model predictions (i.e., the SEC or SEP). An RPD = 2.5 is considered to be the threshold for a good calibration model; below this, the performance of the prediction model for a breeding program is questionable (Williams and Sobering 1993). If RPD = 1, the standard deviation of the measurement data is equal to the standard error of the predictions and the predictions will be unreliable. An additional 1107 NIR spectra were available from the same previous experiment that supplied the 150 additional samples for the sugar analysis. These spectral data were used to generate

additional predicted sugar values for testing genetic effects on sugar yields. These additional spectra provide information on 136 additional clones.

Clone and cluster effects on sugar yields (both measured as well as predicted) were analyzed.

The following linear model was used:

$$Y_{ijklm} = \mu + C_i + L_j + S_k + R_{l(k)} + E_{ijklm}, \quad (4)$$

where:

Y_{ijklm} is the m th observation of the i th clone in the j th cluster in the l th rep at the k th site,

μ is the overall mean,

C_i is the i th random clone effect $\sim \text{NID}(0, \sigma_g^2)$,

L_j is the j th fixed cluster effect,

S_k is the k th fixed site effect,

$R_{l(k)}$ is the l th fixed rep effect within the k th site,

E_{ijklm} is the random error term $\sim \text{NID}(0, \sigma_e^2)$.

The model used the pedigree structure to construct a genetic relationship matrix to adjust clonal BLUPs based on the performance of relatives. Clone mean repeatabilities were calculated as described in equation three.

For each of the response variables, predicted and measured sugar yield, three models were fit; one using old cluster groupings, one using the new cluster groupings, and one that omitted the cluster effect. These models were also re-fit without the clonal effect in order to test the significance of the genetic variance component. LRTs were used to compare the models with and without clonal effect. The models that differ in terms of their fixed effects (i.e., by the cluster term) cannot properly be compared by the LRT (Gilmour et al. 2009). However, comparison of the clone mean repeatabilities and their standard errors produced by these models may be informative about the models' usefulness at partitioning genetic and environmental variation. Genetic values (GVs) for each clone were produced by these models. GVs were used to examine how much clonal performance varies relative to mean sugar yields. Additionally, sugar yield GVs were compared to GVs for growth traits (height and volume). These growth trait GVs are from a genetic analysis of this clonal series done by Zapata-Valenzuela et al. (2013). The CORR procedure in SAS (SAS 9.2, SAS Institute, Cary, NC) was used to produce Pearson correlations between sugar yield and growth trait GVs.

Sugar yields for the three clones that were present in both the younger and older sets of wood samples were compared in order to see if sugar yield was affected by time between collection and hydrolysis. The samples from the previous experiment were collected about three years before those for this experiment. The GLM procedure in SAS (SAS 9.2, SAS Institute, Cary, NC) was used to test the effect of sample age. The following linear model was used:

$$Y_{ijkl} = \mu + C_i + S_j + T_k + E_{ijkl}, \quad (5)$$

where:

Y_{ijkl} is the l th observation of the i th clone at the j th site at the k th time of sampling,

μ is the overall mean,

C_i is the i th fixed clone effect,

S_j is the j th fixed site effect,

T_k is the k th fixed sample age effect,

E_{ijk} is the random error term $\sim \text{NID}(0, \sigma_e^2)$.

For the set of clones sampled for this experiment, two cores were taken per ramet to provide a ‘pooled’ sample that would be better representative of that particular tree. These 150 pooled samples were the basis for the spectral and laboratory analyses. However, NIR spectra also were obtained for each of the cores separately, meaning that there are spectral data for two different samples from each tree. The calibration model was used to predict sugar content using these spectra. Evaluating the correlation between predictions for the two core samples taken from each tree may help indicate if going to the extra effort of taking two samples is worthwhile. The CORR procedure in SAS (SAS 9.2, SAS Institute, Cary, NC) was used to evaluate the correlation between sugar yield predictions of cores taken from the same tree.

Results

Analysis of spectral data

Based on a hierarchical cluster analysis, 30 clones were assigned to six groups (Figure 1). Six groups were used as it corresponds to the number of groups utilized in the previous cluster analysis of this clonal test series (Chapter 2). There is very little consistency apparent between the “old” and “new” cluster memberships. The two large clusters in the new grouping structure both have some commonality of membership with clusters in the older analysis, but generally cluster assignment was different.

PCA was conducted on the py-MBMS spectra for the 150 trees that were sampled. Principal components 1 (x-axis) and 2 (y-axis) account for 41% and 12% of the variation in the population, respectively (Figure 2). The loadings plot shows outliers in three regions of the plot. The higher loadings of these outliers indicate that they are responsible for a relatively higher proportion of sample variation than those m/z peaks clustered near the middle of the plot. These outliers fall into three general groupings: Most of the peaks in the top left direction of the chart are associated with carbohydrates, most of those to the right with extractives, and those at the bottom with lignins (Evans and Milne 1987, Schulten 1996, Sykes et al. 2009).

The py-MBMS ion counts for these influential peaks were subjected to a genetic analysis to determine whether these peaks are heritable (Table 1). Clone mean repeatabilities for peaks associated with carbohydrates/hemicellulose wood components were low-to-moderate (H^2

ranged from 0.13 to 0.62), with relatively large standard errors. Peaks related to extractive components also had low-to-moderate (0.24 to 0.44) clone mean repeatabilities with higher standard errors. For peaks associated with lignin content, clone mean repeatabilities were higher (0.51 to 0.81), with somewhat lower standard errors. LRTs indicate that models for some of the peaks were not significantly different from the reduced form of the model (i.e., without a genetic variance component). Those models that are not significantly different from the reduced model also tend to be those with a higher standard error for the estimate of clone mean repeatability.

PLS modeling of the sugar yield data with the MBMS spectra failed to produce a calibration model for predicting sugar yields. PRESS was minimized in the model fitting zero factors, indicating that no latent factor was found that produces a better model than the intercept-only model. However, using the NIR data to create a calibration equation was successful (Figure 3). Using the NIR data, PRESS was maximized with four latent factors in the model, but the final model ended up using two factors, as the PRESS for that model was insignificantly different from that for the four-factor model.

The coefficient of determination for the training set was low ($R^2 = 0.30$), indicating that a good amount of variation in the response variable is not explained (Table 2). The SEP (0.019) is slightly higher than the SEC (0.016), and likewise the test set R^2 is somewhat lower than for the training set. This is normal since the calibration model would not be expected to perform as well on a test data set compared to the data set it was trained on. The

RPD were low (1.54 and 1.37 for the calibration and prediction models, respectively). This is below the RPD threshold of 2.5 that indicates whether a model is reliable for breeding purposes.

Genetic analysis of wood content predictions and sugar yields

Wood component content predictions derived from py-MBMS and NIR spectra were analyzed to investigate genetic variation. Models for the wood components predicted using MBMS were not significantly different from reduced models which did not include a genetic parameter (Table 3). This was not the case for the components predicted with NIR spectra. For those, the model with the clonal effect was significantly better than the one without. Lignin content was predicted using both NIR and MBMS spectra. The clone mean repeatabilities produced from these predictions are fairly different, (0.53 for the MBMS-based predictions vs. 0.81 for the NIR-based predictions), though the means and ranges of those predictions are similar. However, the Pearson correlation between NIR and MBMS lignin predictions is low ($r = 0.14$, $p = 0.08$).

Mean sugar yield from the high-throughput screening was 0.207 mg sugar/mg wood (Table 4). This is comparable to the mean sugar yield from a larger-scale bench top analysis (0.210 mg sugar/mg wood) that used a dilute acid pretreatment (Chapter 2). Mean predicted sugar content across all available spectra (0.209 mg sugar/mg wood) was comparable to the laboratory data. Clone mean repeatabilities for sugar yield were moderate to high depending on the model used and the response variable (Table 5). Only one of the models was not

significantly different than the reduced model that omitted the genetic variance component. Predicted sugar yields had much higher clone mean repeatabilities than the laboratory-based sugar yields. Additionally, using the old cluster analysis groupings produced models with higher clone mean repeatabilities than those that used the new cluster groupings. However, the highest clone mean repeatability ($H^2 = 0.82$) was estimated using the predicted sugar data as a response variable and a model with no cluster effect fit.

Genetic values (GVs) from the analysis of measured sugar yields showed that, depending on the model used, the best clone is predicted to produce 3.7 to 4.7% more sugar relative to the mean (Table 6). The worst clone is predicted to produce 2.9 to 4.9% less sugar than the mean. The means of the top five clones from each analysis range from 2.7 to 3.5% more sugar produced. The Pearson correlations between the sugar yield GVs and growth trait GVs range from 0.11 to 0.26 and -0.02 to 0.18 for height and volume, respectively (Table 7).

Only one correlation approached significance: the Pearson correlation between height GVs and sugar yield GVs that were predicted using the model that used the “new” cluster variable had a p-value = 0.097. All other correlations were highly non-significant.

Three clones for which sugar yield data was obtained also had samples that were collected both during the course of this experiment and during a previous experiment. The analysis of sugar yields for these clones indicated that the age of the core sample (i.e., the length of time between collection and enzymatic hydrolysis) did not affect measured sugar yield. Both sets

of samples were dried and ground after harvesting, so changes to the structure of the wood while in storage would not be expected to be extreme.

The comparison of predicted sugar values for the two cores taken from each tree found that the average difference between these predictions was 0.006 mg sugar/mg of wood. Given the overall mean yield of 0.21 mg sugar/mg wood (Table 4), the average difference on a percentage basis was 3%. The Pearson correlation coefficient for the two different predictions from each tree was 0.74 ($p < 0.0001$).

Discussion

The genetic analysis of the outlier peaks from the MBMS data indicated the presence of clonal variation for some peaks, including some of those associated with carbohydrates. However, py-MBMS spectra were not used for prediction of sugar yields since the PLS analysis did not find any model with a lower PRESS than the model with just an intercept. This may be related to the depression in levoglucosan yields in biomass pyrolysis (Evans and Milne 1987, Sykes et al. 2009). The formation of levoglucosan, which is a primary product of cellulose pyrolysis, is typically decreased by trace amounts of alkali metals in the biomass. The diversion of the pyrolyzed cellulose into other compounds will have an effect on the spectral data, possibly obscuring information that could well be relevant to the prediction of sugar yield.

For the NIR-based calibration model, the coefficient of determination ($R^2 = 0.30$) was

relatively low. In comparison, one study that incorporated data from a number of different pine species found NIR calibration models with $R^2 = 0.97$ and 0.84 for lignin and cellulose content, respectively (Hodge and Woodbridge 2010). In green wood samples of loblolly pine, NIR calibrations with high R^2 (ranging from 0.79 to 0.88) have been reported for density, microfibril angle and stiffness samples (Schimleck 2003a). R^2 for dry wood samples were slightly higher (0.87 to 0.95). However, high R^2 does not guarantee a high RPD: For the lignin model above, RPD was good (5.91), whereas the RPD for cellulose (2.53) was just above the level for what is considered a good calibration model. The study on mechanical properties (Schimleck 2003a) found RPD below 2.5 for green wood samples and RPD ranging from 1.9 to 3.2 for dry wood samples. A study on predicting a number of traits in agricultural grain and seed crops found R^2 ranging from 0.65 to 0.99 and RPD ranging from 1.02 to 9.08 (Williams and Sobering 1993). The RPD for sugar yield in this study (1.54 and 1.37), while low, are not unusually so, and models with RPD around 1.5 may be useful for a rough initial screening (Schimleck et al. 2003b). In addition, R^2 may be somewhat low given the nature of the trait: sugar yield is affected by a number of other wood properties. Lignin, cellulose, total sugar, and extractive contents as well as coarseness and fiber length can influence sugar yield (Chapter 2). However, for the dilute acid treatment in Chapter 2, no significant model could be fit. It was observed, however, that as sugar yields increased, detection of significant models and model parameters increased.

The genetic analysis of the measured and predicted sugar yields indicates that there is a genetic component to variation in sugar yield. Clone mean repeatability for the best model

for the measured sugar yields was 0.44. For the modeling of the predicted sugar yields, the best clone mean repeatability was 0.82. Part of the higher heritability for the predicted sugar yields may be due to the much larger data set: predictions for 1,109 additional spectra were included in that analysis, in addition to the 299 spectra for which there are measured sugar yields. It is encouraging that the analysis of the predicted data was able to identify a significant genetic component in the variation, given that the low RPD for the model indicates that the predictions may not be reliable.

The model that had the highest clone mean repeatability used predicted sugar yield as the response variable and did not include either cluster grouping as an effect. For the models using measured sugar yield as the response, the one with the highest clone mean repeatability used the old cluster grouping as an effect. However, the p-value was non-significant, so fitting the model without any cluster effect would be preferable. Interestingly, the cluster effect for both the new and old groupings was significant in the models with predicted sugar yield as a response variable. Using the old cluster groupings produced higher clone mean repeatabilities than using the new cluster groupings with either the measured or predicted sugar yield as the response variable. Cluster membership was inconsistent between the new and old groupings (Figure 1), which may be related to the fact that the old groupings were based on a cluster analysis of a much larger set of clones (178) than were the groupings for the new set of trees sampled for this experiment (representing 30 clones). This may also help explain why the old cluster groups were more useful in the genetic analysis than the new cluster groups; given the larger data set behind it, the old cluster analysis may better

represent real differences in chemical/physical wood properties.

Variation in sugar yield from enzymatic hydrolysis is present in this clonal population of loblolly pine, and genetic values for top clones were ~4-5% higher than the mean.. In comparison, the best clone from the dilute acid treatment from the bench-top experiment (Chapter 2) had a sugar yield that was 18% higher than the mean. This increase in sugar yield resulted in an improvement of \$0.42/liter in the cost of ethanol compared to the mean (Chapter 4). Since the yield from this phenotype was approximately four times greater than that for the top GV predicted in this study, we may plausibly expect to capture up to one quarter of that improvement in ethanol price (i.e., \$0.10/liter). Successive generations of breeding would likely be able to increase gains even more, given the heritability of sugar yields and other wood properties. In general, the correlations between sugar yield and growth traits were highly non-significant. The only correlation that was close to significance ($p = 0.097$) involved sugar yields that were predicted using the model that used the “new” cluster variable. However, the model that predicted those sugar yields was not significantly different from a model excluding the genetic variance component altogether. For those models with more reliable predictions of sugar yield, there seems to be no correlation with growth traits.

Conclusion

Testing and selection in a larger population should lead to greater improvements in yield, given the heritable nature of most wood properties. Sugar yields were not high, but different

conversion techniques are already available that will increase the proportion of carbohydrates in the xylem that can be made available for fermentation to ethanol. It is encouraging that genetic differences were detectable even with relatively low sugar yields; higher sugar yields may well lead to even bigger differences between genetic groups.

Literature Cited

- Davis, M., Elam, C., Wiselogel, A., Evans, R., Megraw, R., Wheeler, N., Jech, K., Sewell, M., Neale, D., Tuskan, G., West, D., and Dinus, R. 1999. Application of pyrolysis molecular beam mass spectrometry for the determination of loblolly pine and hybrid poplar cell wall composition. In: Proc. 1999 Tappi Pulping Conference, Oct. 31st to Nov. 4th, 1999, Orlando, FL. 1077-1081
- Decker, S., Brunecky, R., Tucker, M., Himmel, M., and Selig, M. 2009. High-throughput screening techniques for biomass conversion. *BioEnergy Res* 2(4):179–192.
- Decker, S.R., Carlile, M., Selig, M.J., Doepcke, C., Davis, M., Sykes, R., Turner, G., Ziebell, A. 2012 Reducing the effect of variable starch levels in biomass recalcitrance screening. *Methods in Molecular Biology* 908:181-195.
- El-Hamdouchi, A., and Willett, P. 1989. Comparison of hierarchic agglomerative clustering methods for document retrieval. *The Computer Journal* 32(3):220-227.
- Evans, R.J., and Milne, T.A. 1987. Molecular characterization of the pyrolysis of biomass. 1. Fundamentals. *Energy & Fuels* 1(2):123-137.
- Gilmour, A.R., Gogel, B.J., Cullis, B.R., and Thompson, R. 2009. ASReml User Guide Release 3.0. VSN International Ltd, Hemel Hempstead, HP1 1ES, UK. 398p.
- Hodge, G.R., Woodbridge, W., and Dvorak, W.S. 2004. *Use of NIR to predict cellulose content of pine wood* In: Camcore 2004 Annual Report. Dept. of Forestry and Environmental Resources, NCSU, Raleigh, NC. pp 28-30.
- Hodge, G.R., and Woodbridge, W.C. 2010. Global near infrared models to predict lignin and cellulose content of pine wood. *J. Near Infrared Spectrosc.* 18:367-380.
- Hu, Z., Sykes, R., Davis, M.F., Brummer, E.C., and A.J. Ragauskas. 2010. Chemical profiles of switchgrass. *Bioresource Technology* 101(9):3253-3257.
- Kelley, S.S., Jellison, J., and B. Goodell. 2002. Use of NIR and pyrolysis-MBMS coupled with multivariate analysis for detecting the chemical changes associated with brown-rot biodegradation of spruce wood. *FEMS Microbiology Letters* 209:107-111.
- Kelley, S.S. Rowell, R.M., Davis, M., Jurich, C.K., and R. Ibach. 2004. Rapid analysis of the chemical composition of agricultural fibers using near infrared spectroscopy and pyrolysis molecular beam mass spectrometry. *Biomass and Bioenergy* 27:77-88.

Mangrini, K.A., Evans, R.J., Hoover, C.M., Elam, C.C., and Davis, M.F. 2002. Use of pyrolysis molecular beam mass spectrometry (py-MBMS) to characterize forest soil carbon: method and preliminary results. *Environmental Pollution* 116:255-268.

Mann, D.G.J., Labbe, N., Sykes, R.W., Gracon, K., Kline, L., Swamidoss, I.M., Burris, J.N., Davis, M., and C.N. Stewart Jr. 2009. Rapid assessment of lignin content and structure in switchgrass (*Panicum virgatum* L.) grown under different environmental conditions. *Bioenergy Resources* 2:246-256.

Martens, H., and Jensen, S.A. 1983. Partial least squares regression: A new two-stage NIR calibration method. In *Proc. 7th World Cereal and Bread Congress*, Prague, June 1982. Eds. Holas and Kratochvíl. Elsevier Publ., Amsterdam. pp. 607-647.

Mousdale, D.M. 2008 *Biofuels: Biotechnology, Chemistry, and Sustainable Development*. Boca Raton : CRC Press. 424 p.

Plante, A.F., Magrini-Bair, K., Vigil, M.F., and Eldor, P. 2009. Pyrolysis-molecular beam mass spectrometry to characterize soil organic matter composition in chemically isolated fractions from differing land uses. *Biogeochemistry* 92(1-2):145-161.

R Core Team. 2012. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.

SAS Institute Inc. 2008. SAS/STAT®9.2 User's Guide. Chp. 66: The PLS Procedure. SAS Institute Inc., Cary, NC. 60p.

Schimleck, L.R., Mora, C., and Daniels, R.F. 2003a. Estimation of the physical wood properties of green *Pinus taeda* radial samples by near infrared spectroscopy. *Can. J. For. Res.* 33:2297–2305.

Schimleck, L.R., Doran, J.C., and Rimbawanto, A. 2003b. Near infrared spectroscopy for cost effective screening of foliar oil characteristics in a *Melaleuca cajuputi* breeding program. *J. Agric. Food Chem.* 51:2433-2437.

Schulten, H.R. 1996. Direct pyrolysis-mass spectrometry of soils: a novel tool in agriculture, ecology, forestry and soil science. In: Bouton, T.W., and Yamasaki, Y., eds., *Mass Spectrometry of Soils*. Marcel-Dekker, New York. pp. 373-436.

Selig, M.J., Tucker, M.P., Sykes, R.W., Reichel, K.L., Brunecky, R., Himmel, M.E., Davis, M.F., and Decker, S.R. 2010. Lignocellulose recalcitrance screening by integrated high-throughput hydrothermal pretreatment and enzymatic saccharification. *Ind Biotechnol* 6(2):104–111.

Selig, M.J., Tucker, M.P., Law, C., Doeppke, C., Himmel, M.E and Decker, S.R. 2011. High throughput determination of glucan and xylan fractions in lignocelluloses. *Biotechnol Lett* 33:961–967.

Shenk, J.S., and M.O. Westerhaus. 1991. Population definition, sample selection and calibration procedures for near-infrared reflectance spectroscopy. *Crop Sci.* 31:469-474.

Shin, E.J., Nimlos, M.R., and R.J. Evans. 2001. A study of the mechanisms of vanillin pyrolysis by mass spectrometry and multivariate analysis. *Fuel* 80:1689-1696.

Sykes, R., Yung, M., Novaes, E., Kirst, M., Peter, G., and Davis, M. 2009. High-throughput screening of plant cell-wall composition using pyrolysis molecular beam mass spectroscopy. In: Mielenz, J.R. (ed.), *Biofuels, Methods and Protocols, Methods in Molecular Biology*, Humana Press, New York. 308p.

van der Voet, H. 1994. Comparing the Predictive Accuracy of Models Using a Simple Randomization Test. *Chemometrics and Intelligent Laboratory Systems* 25(2):313–323.

Williams, P.C., and Sobering, D.C. 1993. Comparison of commercial near infrared transmittance and reflectance instruments for the analysis of whole grains and seeds. *Journal of Near Infrared Spectroscopy* 1:25–33.

Zapata-Valenzuela, J., Whetten, R.W., Neale, D., McKeand, S., and Isik., F. 2013. Genomic estimated breeding values using genomic relationship matrices in a cloned population of loblolly pine. *G3* 3(5):909-916.

Tables and Figures

Table 1. Clone mean repeatabilities (H^2) for MBMS peaks with high loadings. Bolded LRT (those < 3.841) indicates models that are not significantly better than a reduced model fit without a genetic variance component. The peaks in A. (group 1) are predominantly associated with carbohydrates, those in B. (group 2) are predominantly associated with extractives, and those in C. (group 3) are predominantly associated with lignins.

A. Group 1

<i>Mass Peak (m/z)</i>	<i>H²</i>	<i>Std. Err. (H²)</i>	<i>LRT (vs. reduced)</i>
56	0.13	0.30	0.06
57	0.47	0.18	5.34
60	0.41	0.19	3.92
69	0.42	0.18	5.21
70	0.57	0.15	7.80
71	0.35	0.21	1.80
73	0.35	0.19	3.49
97	0.47	0.17	4.56
114	0.62	0.12	12.81

B. Group 2

<i>Mass Peak (m/z)</i>	<i>H²</i>	<i>Std. Err. (H²)</i>	<i>LRT (vs. reduced)</i>
197	0.32	0.21	2.12
237	0.29	0.21	2.09
239	0.43	0.19	4.19
240	0.41	0.20	3.14
285	0.44	0.18	4.56
300	0.24	0.22	1.37

C. Group 3

<i>Mass Peak (m/z)</i>	<i>H²</i>	<i>Std. Err. (H²)</i>	<i>LRT (vs. reduced)</i>
110	.	.	0
124	0.74	0.09	10.85
137	0.59	0.14	5.57
164	0.54	0.16	5.73
180	0.81	0.06	21.05

Table 2. Results of PLS calibration on sugar yields using NIR spectra.

<i>Data</i>	<i>R</i> ²	<i>SEC/SEP</i>	<i>RPD</i>
Training Set	.30	.016	1.54
Test Set	.24	.019	1.37

Table 3. Clone mean repeatabilities and summary statistics for wood components derived from MBMS and NIR spectral data. Bolded LRT (those < 3.841) indicates models that are not significantly better than a reduced model fit without a genetic variance component.

<i>Component</i>	<i>Mean Content (%)</i>	<i>St. Dev.</i>	<i>Min Content (%)</i>	<i>Max Content (%)</i>	<i>H²</i>	<i>Std. Err. (H²)</i>	<i>LRT (vs. reduced)</i>
MBMS_Lignin %	29.5	0.49	28.0	30.8	0.53	0.18	2.09
MBMS_Extractive %	4.9	0.86	3.5	8.9	0.37	0.20	3.06
NIR_Lignin %	28.7	0.66	27.2	30.5	0.81	0.06	31.66
NIR_Cellulose %	42.3	0.79	40.3	44.7	0.78	0.07	21.91

Table 4. Summary statistics for measured and predicted sugar yields.

<i>Estimation Method</i>	<i>Mean Yield (mg sugar/mg wood)</i>	<i>StDev.</i>	<i>Min Yield (mg sugar/mg wood)</i>	<i>Max Yield (mg sugar/mg wood)</i>
Lab	0.207	0.025	0.103	0.288
Prediction (all)	0.209	0.015	0.154	0.328

Table 5. Genetic parameters for sugar yield and NIR-predicted lignin and cellulose contents. Bolded LRT (those < 3.841) indicates models that are not significantly better than a reduced model fit without a genetic variance component.

<i>Response Variable</i>	<i>H²</i>	<i>Std. Err. (H²)</i>	<i>Cluster grouping (“New” or “Old”)</i>	<i>Significance level of cluster effect</i>	<i>LRT (vs. reduced)</i>
Sugar yield (lab determined)	0.27	0.25	New	0.29	1.25
	0.44	0.16	Old	0.58	5.52
	0.39	0.17	.	.	4.90
Sugar yield (predicted)	0.73	0.08	New	0.02	57.48
	0.81	0.03	Old	0.03	202.28
	0.82	0.03	.	.	262.26

Table 6. Percent differences of clonal BLUPs relative to mean sugar yield.

<i>Response Variable</i>	<i>Cluster grouping ("New" or "Old")</i>	<i>% difference compared to mean</i>		
		<i>Min</i>	<i>Max</i>	<i>Avg. of Top 5</i>
Sugar yield (lab determined)	New	-2.9	3.7	2.7
	Old	-4.9	4.7	3.5
	.	-4.0	4.3	3.5

Table 7. Pearson correlations between sugar yield and growth trait GVs. P-values indicated in parentheses.

<i>Response Variable</i>	<i>Cluster grouping (“New” or “Old”)</i>	<i>Pearson correlation vs. growth GVs</i>	
		<i>Height</i>	<i>Volume</i>
Sugar yield (lab determined)	New	0.26 (0.097)	0.18 (0.246)
	Old	0.11 (0.496)	-0.02 (0.903)
	.	0.18 (0.258)	0.04 (0.777)

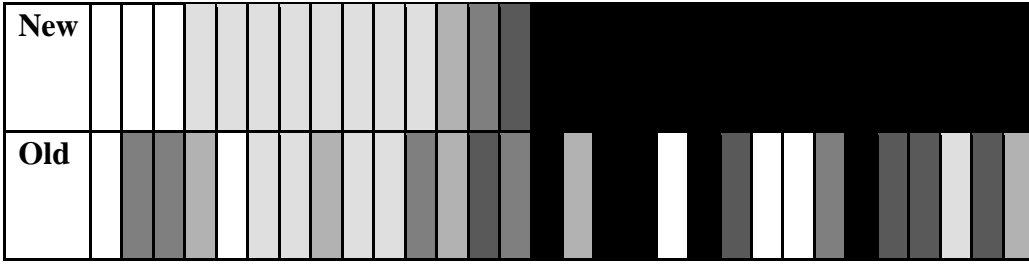


Figure 1. Graphical representation of cluster membership for clones. The cluster memberships were based on separate NIR spectra from different sets of wood cores. The thirty boxes in each row represent the 30 clones analyzed in the new cluster analysis. The top row (New), represents the cluster analysis performed in this experiment, and the bottom row (Old), represents the cluster analysis from a previous experiment. Colors indicate different cluster groupings. Little consistency is seen between the groupings produced by the cluster analyses.

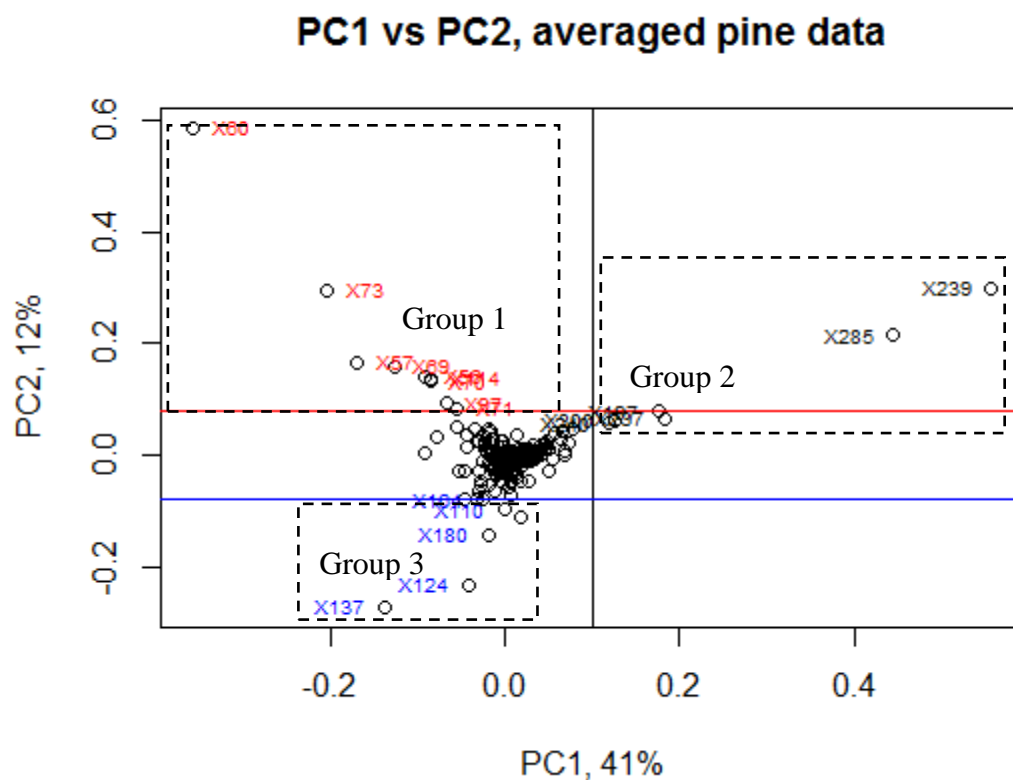
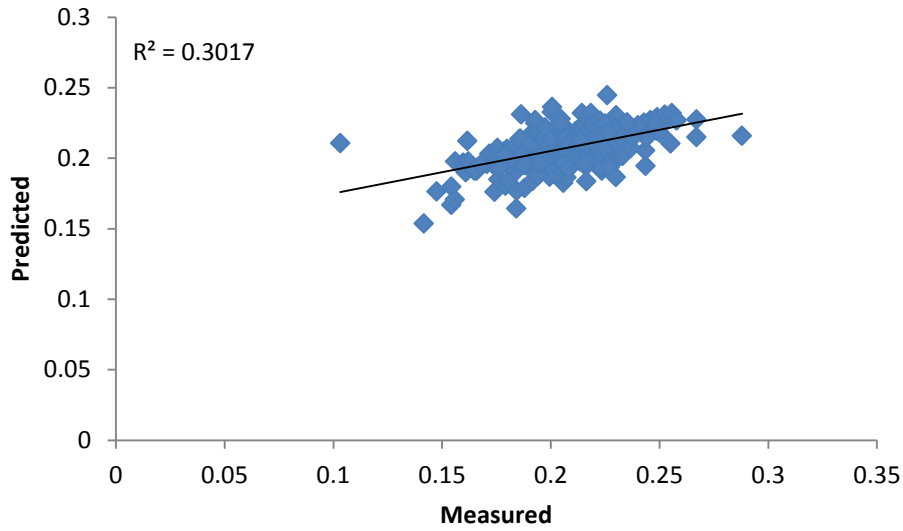


Figure 2. Principle component analysis loadings plot of MBMS data. The outliers, which are the peaks with higher loadings, can be seen in three different regions that are separate from the central cluster of peaks: Group 1 - Peaks in the top left direction of the chart, most are associated with carbohydrates/hemicelluloses; Group 2 - those to the right, most are associated with extractives; and Group 3 - at bottom, most are associated with lignin precursors.

A. Training data set



B. Test data set

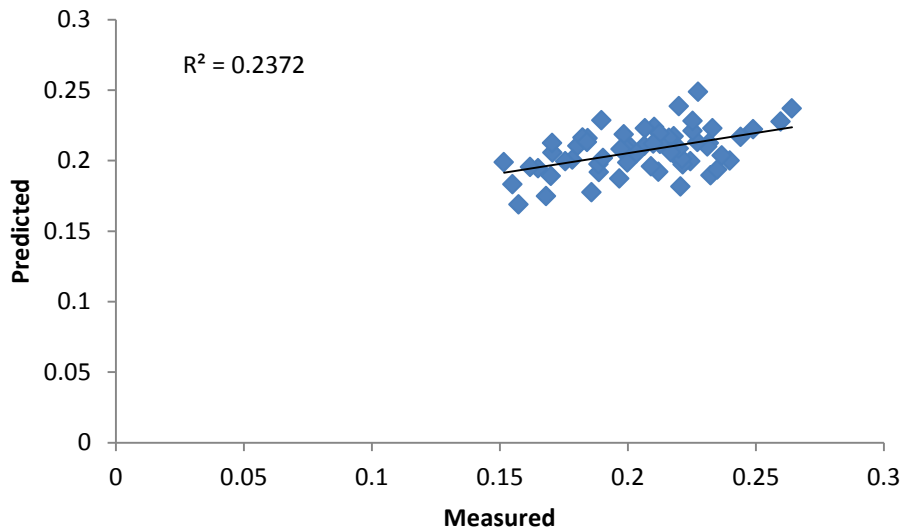


Figure 3. Measured (x-axis) vs. predicted (y-axis) sugar yields. Predictions are from PLS modeling of NIR spectra. R^2 describe the proportion of variation in the laboratory predictions that the PLS model was able to account for using two latent factors.

APPENDIX

Appendix A. Summary of 300 samples used for sugar yield characterization with enzymatic hydrolysis. Both sample sets were harvested from the same Plum Creek Timber Company/CellFor Inc. clonal test series.

Harvested	Harvest Location	# samples	# Clones
Winter 2012	Oliver, GA (1 test)	150	30
Fall 2008	Oliver, GA (2 tests); Holly Hill, SC (1 test)	150	24

CHAPTER 4

MODELING LOBLOLLY PINE IN A BIOREFINERY PROCESS MODEL

Abstract

Cellulosic ethanol as an alternative fuel has been garnering interest in recent years. The use of lignocellulosic biomass for energy has been an area of research interest in the U.S. since the 1970's oil crisis, and recent increases in fuel prices and uncertainty about future supply have renewed these concerns. Many potential sources for this biomass are being considered, from food crop residues (e.g., corn stover, which is not eaten by humans) to grasses and trees. Two of the primary advantages in using lignocellulosic biomass are that there are large amounts available, and the use of this biomass for fuel does not compete with food markets. However, the biotechnology of biomass conversion to fuel is still developing. To encourage investment in this area, it is important to understand the impact of feedstock characteristics on the biorefining process. Economic analysis of sugar yields from loblolly pine (*Pinus taeda* L.) using a biorefinery process model indicates that ethanol biomass is not cost-competitive with gasoline. However, if oil prices go up or bioconversion technology improves, making ethanol from biomass may well become a profitable venture.

Introduction

Investment in bioenergy facilities will be hampered by uncertainty about feedstock availability. For a species to be used as a biofuel feedstock, plantations must be able to produce sufficient quantities of biomass to supply a biorefinery. The landowner/grower needs to make a profit, and the delivered cost to the biorefinery must not be so high that the price of the end product (biofuel) is priced out of the market. Discounted cash flow analysis

using net present value (NPV) and/or soil expectation value (SEV) can be used to evaluate the profitability of a forest plantation (Clutter et al. 1983, Gregory 1987).

For evaluating the profitability of investment in biorefineries, process models can be developed (Gonzalez 2011a). These models account for the various inputs and outputs of energy and materials for a particular biorefinery and can produce an internal rate of return (IRR) that indicates the profitability of the biorefinery. It is expensive to develop pilot plants, so using models of the biorefining processes to test various assumptions is essential. While computer models and spreadsheets cannot be a perfect simulacrum of a real biorefinery, it can be a useful starting off point for investigators and investors in bioenergy.

Another obstacle to the development of a biofuels industry is that bioconversion techniques are still developing, and, at the same time, the type and quantity of locally produced feedstock available will guide the selection of conversion technology. This seems to be especially true for biochemical conversion technologies. Biochemical techniques focus on hydrolytic processes that retrieve the sugars from the cellulose and hemicellulose. These techniques can use a variety of chemical and/or enzyme treatments to help loosen the grip of lignin on the carbohydrates and to break down those carbohydrates. Biochemical conversion is typically assisted by chemical and/or enzymatic agents, and fermentation of the sugars can be improved through specially tailored microbes (Mousdale 2008). For cellulosic ethanol, (which can be made from woody biomass and energy grasses – the most likely sources of significant production of biomass in the next few years), the main difficulty is extracting the

valuable sugars from biomass with as little input and as little sugar loss as possible. The structural properties of these lignocellulosic feedstocks are variable, which means that biochemical conversion techniques will need to be calibrated for specific feedstocks. This study does not go into the details of feedstocks other than loblolly pine, but the work of Gonzalez et al. (2011b) established that loblolly pine is competitive with other feedstocks (including *Eucalyptus*, grasses, and agricultural biomass) on a delivered cost basis for carbohydrate and BTUs. The challenge, then, is to improve conversion of normally-recalcitrant softwood biomass.

The difficulty of successful commercial biorefinery development is highlighted by recent failures. In 2011, North Carolina's first ethanol plant, run by Clean Burn Fuels LLC, entered Chapter 11 bankruptcy, citing an increase in corn prices that made the biorefinery unprofitable (Calhoun 2011). Another high profile-example failure in the southeast is Range Fuels LLC, whose thermochemical-based bioenergy plant in Soperton, Georgia closed in 2011. The site was auctioned off in 2012 and is being developed by LanzaTech, which is intending to move the project toward commercial-scale production of ethanol (Lane 2013).

There are scientific issues to be addressed both in the areas of biomass supply and biofuel production in terms of increasing the types and yields of available feedstocks and improving efficiency in fuel production. Information on biomass yields and conversion efficiency is essential to fostering industrial-scale biomass development. This study uses sugar yield data from a loblolly pine conversion experiment as input into a developed biorefinery process

model to test the effects of variation in several factors (including sugar yield and biomass pricing) on ethanol price.

Methods

Data

17 clonal varieties of loblolly pine, chosen for a diverse range of chemical and physical wood properties, were tested to characterize variation in yield of fermentable sugars from enzymatic hydrolysis (Chapter 2). A cluster analysis based on near-infrared (NIR) spectra of ground wood samples from multiple individual trees, or ramets, of each of the clonal genotypes was used to divide the clones into groups. NIR spectra reflect chemical and physical wood properties, so the clustering should have produced groups of clones that are similar for some combination of these traits. One cluster was identified that is superior for sugar yield, and this could provide a quick way of identifying additional superior genotypes for sugar production.

In this experiment, three different pretreatments were tested: dilute acid, alkaline, and alkaline with mechanical beating (10,000 rotations in a PFI-type beater). The purpose of pretreatment is to release the cellulose from the lignin, thereby making the cellulose available for enzymatic hydrolysis. One level of enzyme (20 FPU, or filter paper units) was used for the dilute acid pretreatment, and three levels of enzyme (10, 20, and 40 FPU) were used for the alkaline and alkaline with mechanical beating pretreatments, for a total of seven treatments. Each treatment was applied to a mixture of wood chips representing each clone.

This mixture was composed of three ramets of a given clone. Mean sugar yields were calculated for each treatment (Table 1). The alkaline pretreatments with mechanical beating produced more sugars than the dilute acid or alkaline pretreatments. Additionally, increasing enzyme levels also produced greater yields of sugars.

The Process Model

The sugar data from the experiment was analyzed using Greenfield Mill Dilute Acid and Greenfield Mill Green Liquor process models for a loblolly biorefinery processing 500,000 bone dry tons of biomass per year. The financial analysis is based on biorefinery process models developed by colleagues in the Department of Forest Biomaterials at NCSU under a grant from the Biofuels Center of NC. The models take into account the various inputs and outputs of capital, expenses, energy, materials, labor, and products that would be found in an operational plant (Gonzalez et al. 2011). The variables that this analysis tested are described below.

The goal of the analysis is to estimate the effect of conversion methods, cluster membership, clone, enzyme costs, and biomass costs on the economic feasibility of an industrial-scale biorefinery using loblolly pine as a feedstock. All scenarios modeled use a 12% internal rate of return (IRR), and the output is the selling point ethanol that is required to break even with a 12% IRR. The bulk sugar yields (Table 1.), are simple tabulations of sugar data provided by high pressure liquid chromatography (HPLC) analysis of the samples after pretreatment (for dilute acid) and enzymatic hydrolysis (dilute acid and green liquor pretreatments). These

sugar data are broken down into glucan (cellulose fraction) and as well as xylan and mannan (5- and 6-carbon hemicelluloses, respectively). These different sugars were input separately into the biorefinery simulator, as they respond differently during bioconversion.

Mean sugar yields for each treatment method were modeled. Four different delivered biomass prices were modeled using these mean sugar yields: \$27.5, \$33, \$38.5, and \$44 per green Mg (this corresponds to \$25, \$30, \$35, and \$40 on a green ton basis). \$27.5/Mg is close to delivered pine biomass prices in North Carolina from 2006 to 2010 (Timber Mart-South 2010) and was used as the biomass price for the sensitivity analyses of other factors (described below). Higher prices were modeled because 1) \$33-\$38.5/Mg delivered biomass prices are not atypical in other parts of the Southeast, and 2) if biorefineries become a reality, demand for softwood biomass could potentially increase local delivered biomass prices.

Enzyme costs were also subjected to a sensitivity analysis using mean sugar yields for each treatment. Mean sugar yields were analyzed using enzyme costs of \$0.026, \$0.013, and \$0.003/FPU/liter (corresponds to \$0.10, \$0.05, and \$0.01/FPU/gallon) in order to test the effect of enzyme cost on ethanol price. For the other sensitivity analyses, enzyme cost was kept at \$0.026/FPU/liter. The base cost of \$0.026/FPU/liter was used as the baseline for other sensitivity analyses. With the enzyme concentrations that were used to get sugar yields for this experiment, this would indicate enzyme costs ranging from approximately \$0.25 to \$1.00 per liter using the base enzyme cost. In comparison, Klein-Marcuschamer et al. (2012) found that enzyme costs were \$0.39 per liter under normal sugar yields from corn stover.

The PFI mechanical beating method utilized for three of the treatments was accounted for in the modeling. Based on the energy usage of a PFI beater (Kerekes 2005), the amount of energy required for 10,000 revolutions was estimated at 517 kilowatt hours per Mg of biomass. However, when using a mechanical beating technology at an industrial scale to produce an equivalent product, there may well be an order-of-magnitude reduction in energy cost (Koo et al. 2011). Accordingly, a second analysis of these treatments was done using 52 kilowatt hours per Mg as the energy cost.

To account for potential genetic impacts on conversion efficiency and thus on ethanol price, the sugar yields of the best and worst clone for each treatment were analyzed. Additionally, membership in NIR cluster three was contrasted with the average sugar yields of all other clusters for each treatment, given that previous analysis had shown that cluster three was significantly higher-yielding than other clusters for some treatments.

Results and Discussion

There is considerable variation between treatments in the cost/liter of ethanol required to achieve a 12% IRR (Table 2). The dilute acid treatment was clearly the worst, and this was due to lower sugar yields and higher chemical costs than the alkaline treatments. The alkaline treatments have a lower ethanol cost/liter, but there is not a clear positive relationship between ethanol price and mean sugar yields (Table 1), indicating the presence of other important factors that affect ethanol cost. The two best prices have low to moderate enzyme doses and use mechanical beating.

Reduction of enzyme costs (Figure 1) produces a dramatic reduction in ethanol cost. The relative steepness of the decline is indicative of the amount of enzyme used for a given treatment. It is clear that enzyme costs of \$0.026/FPU/liter make the ethanol price too high for the most intensive enzymes doses to be economically feasible. However, when enzyme costs are reduced by 50% or 90%, it becomes affordable to use more. If enzyme prices were to decrease to \$0.003/FPU/liter, several of the treatments could produce ethanol at a cost less than \$1.00/liter, and the ALK40PFI treatment comes in at less than \$0.78/liter. Enzyme manufacturers have been working to improve enzyme costs and/or efficiency (Lane 2012), but how much of a reduction is possible remains to be seen.

Reduction of energy costs for the three treatments with mechanical beating (Figure 2) produces modest declines in ethanol cost. If industrial-scale mechanical beating can reduce energy demands by the amount modeled, it will result in reducing costs by \$0.05-\$0.08/liter, which may be significant to the consumer.

Similarly, biomass cost increases have a modest effect on final ethanol costs (Figure 3). For every \$5.5/Mg increase in delivered biomass cost, there is an average increase of about \$0.05/liter in ethanol cost. Also, there is an apparent relationship between enzyme levels and ethanol price increases due to biomass cost. For those treatments that used less enzyme, the increase in ethanol price due to biomass cost increase is somewhat greater than for those treatments that used higher amounts of enzyme. This is probably due to the fact that

treatments using less enzyme release less sugar per Mg of biomass. Though, if enzyme prices remain high, enzyme cost will have a much bigger impact on ethanol prices than will biomass costs. However, in planning for a biorefinery, it is nonetheless important to gauge the effect that demand from the plant may have on local biomass costs.

Choice of feedstock genetics may also be an important factor in final ethanol cost. If clones that produce superior sugar yields can be identified, it may be possible to increase efficiency with only a small investment in the right genetics. The price difference between the best and worst clones for each treatment ranged from \$0.20 to \$1.18 per liter (Figure 4). It must be cautioned, however, that additional data are needed to test the magnitude of these genetic differences and to make a robust estimation of how much genetic gain is feasible. A genetic analysis of sugar yields from powdered wood core samples that were pretreated with dilute acid and then subjected to enzymatic hydrolysis indicates that the genetic variance component is significant (Chapter 3). However, more sugar yield data from alkaline conversion methods would be useful to see how clonal variation and heritability might change when a greater proportion of sugars are being extracted.

Clones from cluster three were shown to have higher sugar yields than those in other clusters (Chapter 2). Cluster three was compared to all other clusters for the seven treatments to see if there was any general effect on ethanol prices (Figure 5). It is apparent that cluster three membership is generally associated with lower ethanol prices, though the magnitude of the difference varies from treatment to treatment. Interestingly, for the treatments that showed

the clearest statistical differences in sugar yield for cluster three (ALK20PFI and ALK40PFI), the magnitude of the difference in price is not as great relative to other treatments. This is likely due to the intensive nature of these treatments (high enzyme dose with mechanical beating), which produced a very high sugar yield for all clusters.

Conclusion

The alkaline pretreatment obtained higher sugar yields than the dilute acid pretreatment that used equivalent enzyme levels, so the alkaline pretreatment resulted in an overall lower ethanol cost compared to dilute acid. Increasing enzyme levels will augment sugar yields, but at a significant cost. At high prices/FPU, enzyme costs outweigh the sugar yield gains at the high level of dosage, making lower levels of enzyme dosing more economical. If enzyme prices can be significantly reduced, however, it may become feasible to use higher doses of enzyme and thereby produce more sugars from a given amount of feedstock. Additionally, mechanical beating is an economically effective tool to promote sugar release, as energy costs are more than offset by increased sugar yield. Production of biofuel from biomass will require some combination of advances in feedstocks and/or biorefining; feedstock advances are needed both to increase yields as well as select trees with superior digestibility.

Literature Cited

- Burkheisser, E.V. ed. 2010. Biological barriers to cellulosic ethanol. Nova Science Publishers Inc., New York. 251 p.
- Calhoun, J. 2011, April 4th. Clean Burn Fuels files for Chapter 11 bankruptcy. *The Fayetteville Observer*. Available online: <http://www.fayobserver.com/articles/2011/04/04/1083859>. Last accessed 8-9-11.
- Clutter, J.L., Fortson, J.C., Pienaar, L.V., Brister, G.H., and Bailey, R.L. 1983. Timber Management: A Quantitative Approach. John Wiley & Sons, New York. 329p.
- Gonzalez, R., Treasure, T., Phillips, R., Jameel, H., and Saloni, D. 2011a. Economics of cellulosic ethanol production: green liquor pretreatment for softwood and hardwood, greenfield and repurpose scenarios. *Bioresources* 6(3):2551-2567.
- Gonzalez, R., Phillips, R., Saloni, D., Jameel, H., Abt., R., Pirraglia, A., and Wright, J. 2011b. Biomass to energy in the southern United States: supply chain and delivered cost. *Bioresources* 6(3):2954-2976.
- Gregory, G.R. 1987. Resource economics for foresters. John Wiley & Sons, New York. 477 p.
- Kerekes, R. 2005. Characterizing refining action in PFI mills. *Tappi J* 4(3):9-13.
- Klein-Marcuschamer, D., Oleskowicz-Popiel, P., Simmons, B.A., and Blanch, H.W. 2012. The challenge of enzyme cost in the production of lignocellulosic biofuels. *Biotechnol Bioeng* 109(4):1083-1087.
- Koo, B., T.H. Treasure, H. Jameel, R.B. Phillips, H. Chang, and S. Park. 2011. Reduction of enzyme dosage by oxygen delignification and mechanical refining for enzymatic hydrolysis of green liquor-pretreated hardwood. *Appl Biochem Biotechnol*: published online 6-7-11.
- Lane. J. 2012. The Enzyme Wars. *Biofuels Digest* February 24, 2012. Available online: <http://www.biofuelsdigest.com/bdigest/2012/02/24/the-enzyme-wars/>. Last accessed 9-23-13.
- Lane. J. 2013. LanzaTech: Biofuels Digest's 5-Minute Guide (Q1 2013 update). *Biofuels Digest* January 8, 2013. Available online: <http://www.biofuelsdigest.com/bdigest/2013/01/08/lanzatech-biofuels-digests-5-minute-guide-q1-2013-update/>. Last accessed 1-9-13.

Mousdale, D.M. 2008 Biofuels: Biotechnology, Chemistry, and Sustainable Development. Boca Raton : CRC Press. 424 p.

Timbermart-South. Market News Quarterly - 4th Quarter, 2010. Available online at http://www.timbermart-south.com/pdf/4Q2010news_abridged.pdf; last accessed April 28, 2011.

Tables and Figures

Table 1. Mean sugar yields by treatment. Standard deviation is in parentheses.

Treatment ¹	Mean Sugar Yield (mg sugar/g wood)
DA20	210 (19.9)
ALK10	228 (52.8)
ALK20	332 (62.6)
ALK40	414 (59.9)
ALK10PFI	341 (60.6)
ALK20PFI	444 (65.5)
ALK40PFI	521 (34.9)

¹DA20 is dilute acid pretreatment with 20 FPU enzyme. ALK10, ALK20, ALK30 are alkaline pretreatments with 10, 20, 40 FPU of enzyme, respectively. ALK10PFI, ALK20PFI, ALK40PFI are alkaline pretreatments with mechanical beating and 10, 20, 40 FPU of enzyme, respectively.

Table 2. Cost/liter of ethanol required to achieve a 12% IRR, by treatment. DA20 is dilute acid pretreatment with 20 FPU enzyme, ALK10 is an alkaline pretreatment with 10 FPU of enzyme, and so forth. ALK10PFI is an alkaline pretreatment with mechanical beating and 10 FPU.

Treatment ¹	Cost/liter
DA20	\$2.33
ALK10	\$1.61
ALK20	\$1.43
ALK40	\$1.72
ALK10PFI	\$1.28
ALK20PFI	\$1.27
ALK40PFI	\$1.63

¹DA20 is dilute acid pretreatment with 20 FPU enzyme. ALK10, ALK20, ALK30 are alkaline pretreatments with 10, 20, 40 FPU of enzyme, respectively. ALK10PFI, ALK20PFI, ALK40PFI are alkaline pretreatments with mechanical beating and 10, 20, 40 FPU of enzyme, respectively.

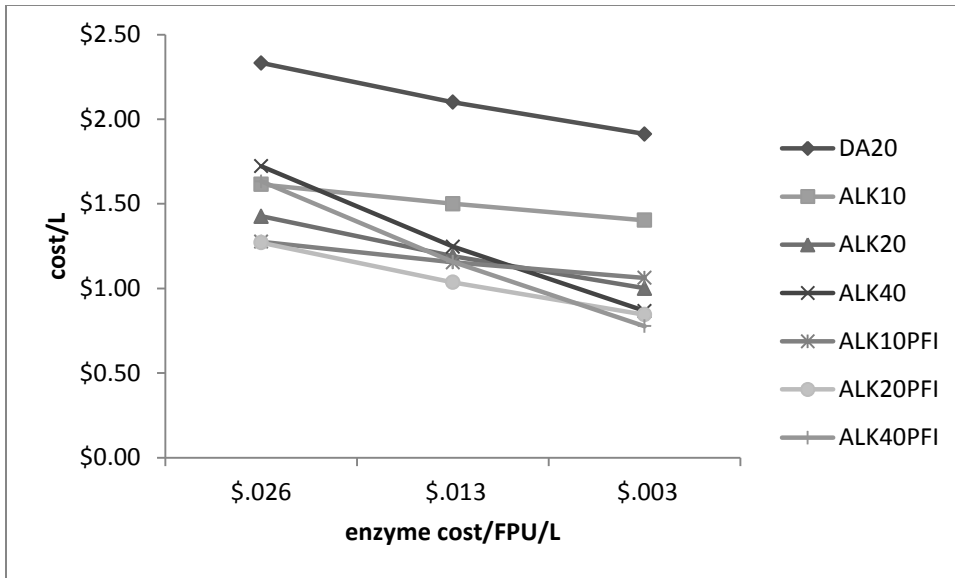


Figure 1. Cost/liter of ethanol with enzyme costs varying for each treatment¹.

¹DA20 is dilute acid pretreatment with 20 FPU enzyme. ALK10, ALK20, ALK30 are alkaline pretreatments with 10, 20, 40 FPU of enzyme, respectively. ALK10PFI, ALK20PFI, ALK40PFI are alkaline pretreatments with mechanical beating and 10, 20, 40 FPU of enzyme, respectively.

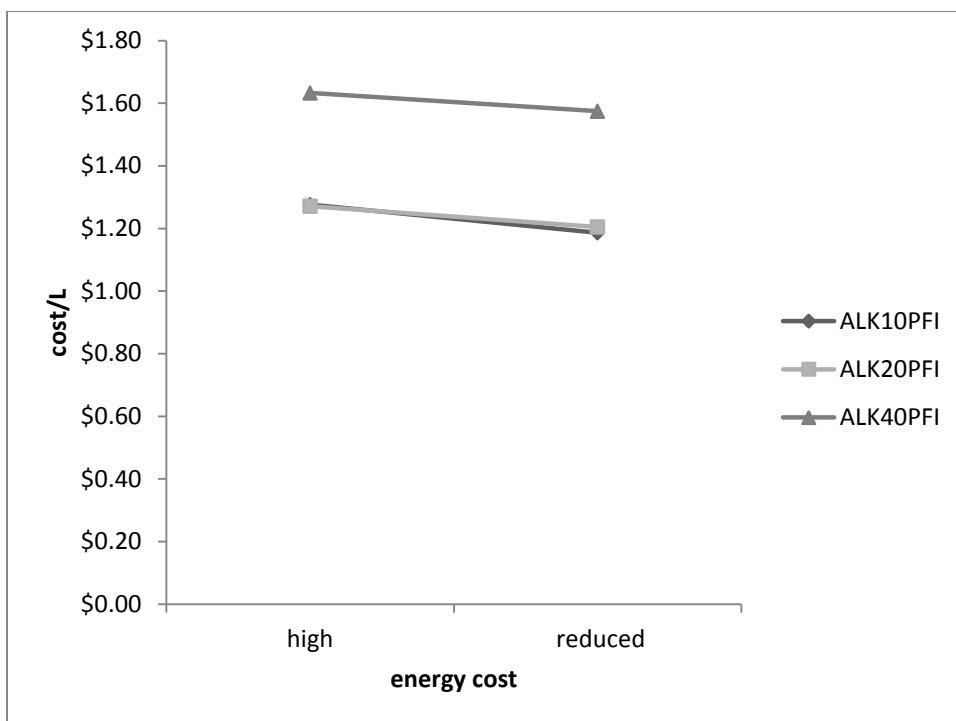


Figure 2. Cost/liter of ethanol with energy cost varying for each treatment¹.

¹DA20 is dilute acid pretreatment with 20 FPU enzyme. ALK10, ALK20, ALK30 are alkaline pretreatments with 10, 20, 40 FPU of enzyme, respectively. ALK10PFI, ALK20PFI, ALK40PFI are alkaline pretreatments with mechanical beating and 10, 20, 40 FPU of enzyme, respectively.

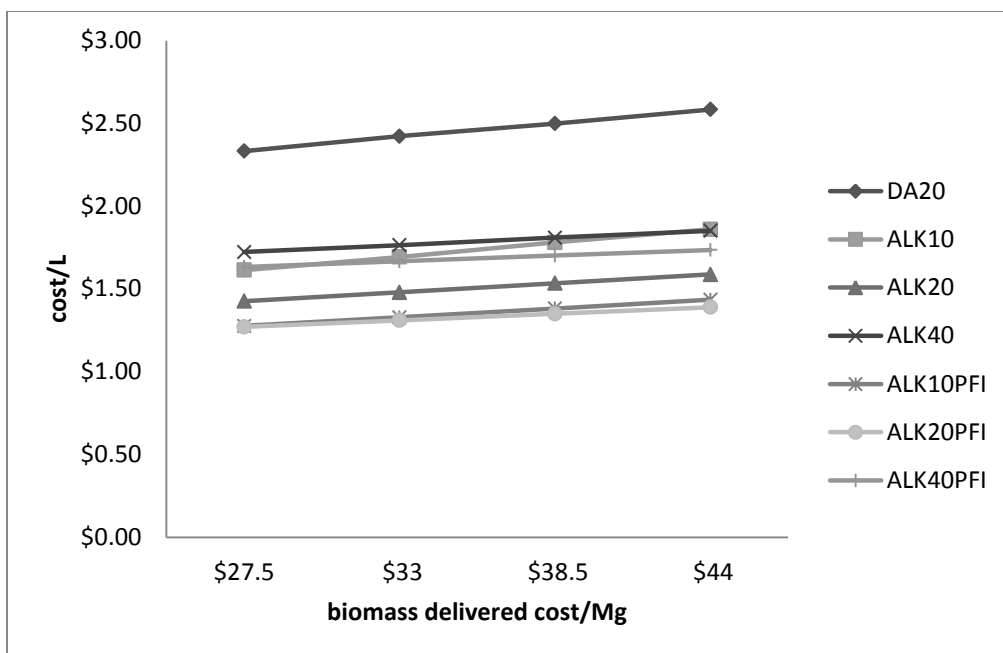


Figure 3. Cost/liter of ethanol with delivered biomass price varying for each treatment¹.

¹DA20 is dilute acid pretreatment with 20 FPU enzyme. ALK10, ALK20, ALK30 are alkaline pretreatments with 10, 20, 40 FPU of enzyme, respectively. ALK10PFI, ALK20PFI, ALK40PFI are alkaline pretreatments with mechanical beating and 10, 20, 40 FPU of enzyme, respectively.

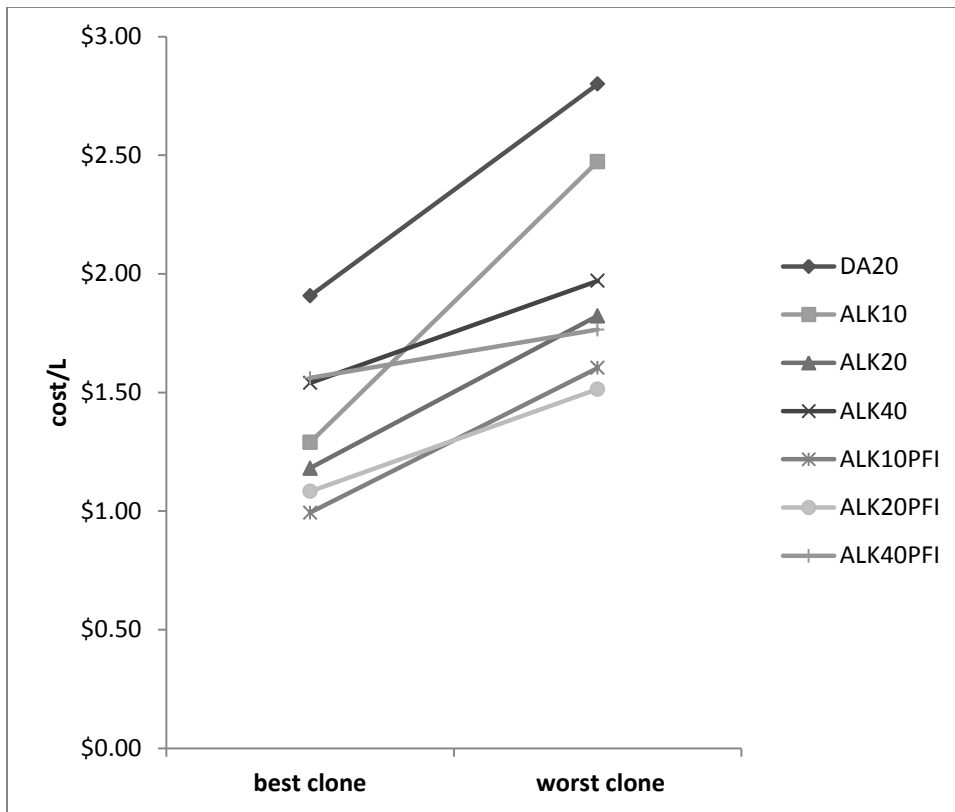


Figure 4. Cost/liter of ethanol comparing best- and worst-yielding clones for each treatment¹.

¹DA20 is dilute acid pretreatment with 20 FPU enzyme. ALK10, ALK20, ALK30 are alkaline pretreatments with 10, 20, 40 FPU of enzyme, respectively. ALK10PFI, ALK20PFI, ALK40PFI are alkaline pretreatments with mechanical beating and 10, 20, 40 FPU of enzyme, respectively.

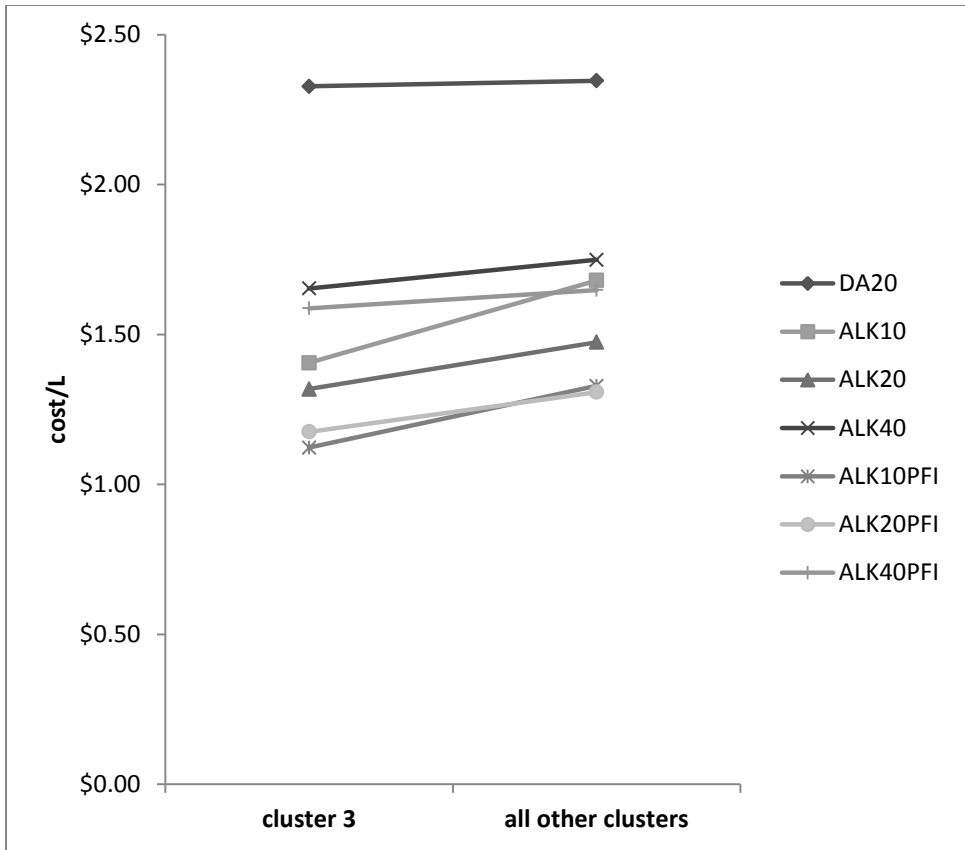


Figure 5. Cost/liter of ethanol comparing cluster three against other clusters for each treatment¹.

¹DA20 is dilute acid pretreatment with 20 FPU enzyme. ALK10, ALK20, ALK30 are alkaline pretreatments with 10, 20, 40 FPU of enzyme, respectively. ALK10PFI, ALK20PFI, ALK40PFI are alkaline pretreatments with mechanical beating and 10, 20, 40 FPU of enzyme, respectively.