ABSTRACT

CUMBIE, WILLIAM PATRICK. Association Genetics for Growth, Carbon Isotope Discrimination, and Stem Quality in Loblolly Pine. (Under the direction of Dr. Barry Goldfarb.)

Association genetics offers the potential to identify molecular markers which account for variation in phenotypic traits of interest in loblolly pine. The detection of associations could be used to explain the underlying genetics architecture of complex quantitative traits, and potentially the identification of markers which could be used for selection in breeding programs to capture additional genetic variation and make greater genetic gains through selection and breeding cycles. Three experiments were conducted to explore the genetic variation in growth, water use efficiency, and stem form in loblolly pine.

The first experiment involved an association genetics approach using an unimproved population of 425 clonally replicated unrelated trees to test 3,938 single nucleotide polymorphisms (SNPs) for association with phenotypic variation in carbon isotope discrimination, total tree height, and foliar nitrogen concentration after two growing seasons. Best linear unbiased prediction was used with a spatial adjustment to remove additional environmental variation from phenotypic data. After correction for multiple testing a total of 14 SNPs were associated with carbon isotope discrimination, height, and foliar nitrogen concentration.

The second experiment was a quantitative analysis of genetic variation in growth and stem form traits related to sawtimber quality in an elite population of loblolly pine. Progeny from an elite population of loblolly pine were bred in a diallel mating design and

planted at four sites across the lower coastal plain of the southeastern United States. Growth, disease incidence, stem quality and a sawtimber potential score were measured after six growing seasons. There were significant differences among families for all traits measured. Individual-tree narrow-sense heritability estimates ranged from 0.06 to 0.22. Height and volume were highly correlated with the sawtimber potential score of individual trees. From multiple regression, 79% of the variation in sawtimber potential breeding values could be attributed to variation in volume, rust incidence, stem sweep, and forking breeding values. The potential economic value of loblolly pine was increased as much as 162% over local checks when both volume and sawtimber potential were used to select the 10 best parents from the population. Implementation of a selection index on currently measured traits is a promising opportunity to make gains in the proportion of sawtimber produced from improved germplasm of loblolly pine in the southeastern United States.

The third experiment was an attempt to identify single SNPs associated with variation in growth and stem form traits in loblolly pine (*Pinus taeda* L.). Associations were tested between 4,200 SNPs and breeding values for a population of 200 largely unrelated selections of loblolly pine. We identified 13 SNP-phenotype associations for sawtimber index, volume, and stem straightness after multiple testing correction. Individual SNPs explained from 0% to 27% of the variance in breeding values used as phenotypes. The most significant SNPs were used to estimate genetic values for an independent population of 153 clonally replicated trees. The correlation between marker based estimated genetic

values and the BLUP predictions for volume was highest when 10 to 25 SNP loci were used (r=0.27). Gain estimates from marker based selection scenarios were compared to seedling and clonal progeny testing scenarios to explore the needed reliability of marker-based estimates to assess the incorporation of marker based selection in loblolly pine breeding programs. Comparisons revealed that even low repeatability values for marker-based selection may be a potential consideration in tree breeding programs.

Association Genetics for Growth, Carbon Isotope Discrimination, and Stem Quality in Loblolly Pine

by
William Patrick Cumbie

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

Raleigh, NC

2010

APPROVED BY:

_____	_____
Dr. Barry Goldfarb Chair of Advisory Committee	Dr. John King
_____	_____
Dr. Bailian Li	Dr. Dahlia Nielsen

_____
Dr. Ross Whetten

BIOGRAPHY

William Patrick Cumbie was born and raised in Charlotte, North Carolina. Patrick graduated from Myers Park High School in Charlotte in 1996 and attended North Carolina State University for his undergraduate education. After completing a B.S. in forest management in 2000, Patrick enrolled in a master's program at NC State under the direction of Barry Goldfarb and Bailian Li. In 2002 Patrick married Sarah Efird, completed his master's degree and accepted a position with Weyerhaesuer Company in Macon, Georgia as a research forester and tree breeder.

After a positive experience in forest research with Weyerhaeuser Company, Patrick decided to return to NC State to pursue a PhD in forest genetics. In the fall of 2005, Patrick enrolled in his PhD program in the Department of Forestry and Environmental Resources. In the spring of 2006 Patrick had the opportunity to join the staff of the NCSU-Cooperative Tree Improvement Program as a research assistant.

During Patrick's PhD program and employment with the Tree Improvement Program, Patrick and Sarah had two children, Will and McLean. Being a student, employee, husband, and father has been demanding, challenging, and very rewarding for Patrick.

ACKNOWLEDGEMENTS

This research would not have been possible without support from the Neale lab at the University of California at Davis and the staff and members of the NCSU-Cooperative Tree Improvement Program. Their assistance in funding, in-kind work contributions and guidance along the way are gratefully acknowledged.

Barry Goldfarb served as my committee chair, and John King, Bailian Li, Dahlia Nielsen, and Ross Whetten served as committee members. I am thankful for the help, counsel, challenges, and support they have offered during my program.

I would like to thank Dr. Fikret Isik and Dr. Steve McKeand for their support and counsel during my graduate education and employment with the Tree Improvement Program.

I am very thankful for my wife, Sarah Efird Cumbie, for her love, patience, and support in my seemingly never-ending education. Her sacrifices to allow me the time to pursue my research interests will forever be remembered. I am excited to find out what the future holds for us as we journey together.

Many graduate students, undergraduates, and part-time staff have been a tremendous help in completing my work. I specifically thank Maria Wirth Wilkes for her help with rooted cutting production, Jesus Espinoza for field data collection in the Lower Gulf Elite trials, and the Tree Improvement staff for their help in countless ways.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

GENERAL INTRODUCTION

The southeastern United States is one of the most world's most productive regions for industrial timber and currently produces 60% of the nation's wood needs (Prestemon and Abt 2002). Each year, hundreds of millions of loblolly pine (*Pinus taeda* L.) seedlings are planted to establish pine plantations in the southeastern United States (McKeand et al. 2003). Intense management of these plantations can achieve productivity levels that exceed 10 tons of wood per acre per year (Stanturf et al. 2003). Tremendous gains in growth, form, and disease resistance have already been made during two cycles of selective genetic improvement of southeastern U.S. loblolly pine (Li et al. 2000). As forest plantations are more intensively managed for greater levels of timber productivity, the loblolly pine genotypes deployed must have the potential to produce desired growth rates and stem form quality under variable site conditions.

Productivity gains through genetic improvement of loblolly pine exceeded 12% in rotation age volume in rogued first-generation seed orchards and 30% in rogued second-generation seed orchards when compared to unimproved planting stock (Li et al. 2000). These gains have added substantial value for forest landowners who purchase improved seedlings and implement currently-recommended silvicultural practices (McKeand et al. 2006a). To realize maximum growth potential on plantation sites, deployed genotypes must have attributes that will enable them to survive under variable environmental conditions and generate the wood quality needed for production of higher value forest products.

Important limitations that affect timber productivity on sites in the southeastern United States include water and nutrient deficiencies. Albaugh et al. (2004) reported that nitrogen, phosphorus and boron deficiencies of loamy and sandy soils were more critical than water limitations for stand development, while Dougherty and Gresham (1988) emphasized the critical effects of water availability. To produce greater amounts of timber on fewer acres, improvements in growth efficiencies in the deployed genotypes could reduce operational costs of plantation management. McKeand et al (1997) observed that good growing families of loblolly pine always outperformed poor growing families under a variety of site conditions and resource availabilities, suggesting that genetic gains are more closely related to growth efficiency than leaf area.

The value of gains in genetic quality for plantation forestry has recently been assessed in loblolly pine (McKeand et al. 2006a). Volume gains alone produced increases of $50 to $300 in net present value per acre in simulations of pulpwood and sawtimber silvicultural regimes. It has been suggested that if stem quality improvements were included, the added genetic quality gain value might double (McKeand et al. 2006b), but to date the value of genetic gain in stem quality has not been quantified. The value of sawtimber is substantially greater than that of pulpwood. Sawtimber values can be four to seven times greater than values of pulpwood produced on similar sites (Timber Mart-South 2007). This difference in value has become an important focus in loblolly pine breeding programs as the value of genetic improvements is quantified. Assessments of stem form genetics in different pine species have generally produced low heritability estimates. Composite form

trait assessments in radiata pine (*Pinus radiata* D.Don) exhibited low heritability across several sites (Jayawickrama 2001), and Busby (1983) reported low repeatability of subjective crown grading in loblolly pine as well. Crown grading in jack pine (*Pinus banksiana*) also revealed lower heritability estimates as compared to those for growth and branch traits, and crown form was negatively correlated with tree height at age 7 years (Adams and Morgenstern 1991).

In loblolly pine breeding programs, the application of quantitative genetics has been effective in producing significant gains in growth, stem form and disease resistance but breeders should explore other tools to capture genetic quality gains and understand the genetic variation of important economic and ecological traits. Many important quantitative heritable traits in plants and animals are complex, due to the fact that multiple genes can contribute to the variation in a single trait. Genetic association studies, also called linkage disequilibrium or association mapping, are methods used to identify phenotypic trait and genetic marker relationships based on linkage disequilibrium at the population level (Flint-Garcia et al. 2003). Recent increases in the amount of genetic data and improved high-throughput technology have made association studies more feasible in forest trees such as loblolly pine. Identification of specific genes that contribute to an important phenotypic trait could increase quality gains in forest tree breeding programs as well as contribute to a greater understanding of gene functions. Several characteristics of loblolly pine make it a favorable species for application of genetic association methods. These include large out-crossing populations, nucleotide diversity, the ability to clone genotypes for precise

production of desired phenotypes and rapid decay of linkage disequilibrium (Neale and Savolainen 2004).

While forest trees are complex organisms and difficult to work with because of their large genomes, they present a unique opportunity to use an almost natural population for association mapping of complex traits. This has a direct application to forest tree breeding programs. Using traditional quantitative methods, breeders select for complex desired traits such as growth rate, disease resistance or wood properties without knowing which particular genes are actually important. QTL studies have been successful in identifying markers related to traits such as wood properties (Devey et al. 2004), but offer limited application because QTL markers must be confirmed in any pedigree where selection is to take place. Association studies offer an alternative to traditional QTL studies as a method for marker aided breeding and selection research. Association studies are geared toward population-level genetic characteristics rather than individual pedigrees, which make them ideal for study of a large out-crossing population (Neale and Savolainen 2004). If an association study is successful, specific genes or loci that determine economically important traits can be identified. Selections can even be made from the association test and deployed operationally, since the populations will likely be clonally maintained. Alternatively, the markers or genes identified in association studies can then be used for screening and selection in breeding populations.

Association mapping could be useful for forest tree breeding programs by 1) finding SNPs which explain the underlying genetic architectures of complex quantitative traits and 2)

identifying SNPs that explain significant variation in important traits that could be incorporated into breeding programs.

Results of three studies conducted to explore the potential use of association mapping in loblolly pine are reported here. The first experiment involved the development of a clonal association population of 425 unrelated genotypes that were genotyped for 3839 SNPs. These were tested for associations with two-year tree height, carbon isotope discrimination, and foliar nitrogen content. Association testing was conducted in a two-step process where phenotypes where analyzed in a mixed model with a spatial residual to remove effects of environmental variation and then tested for associations with SNPs.

The second experiment sought to quantify growth and stem quality trait variation in an elite population of loblolly pine. Genetic parameters were estimated for individual traits and a composite visual assessment of each tree's potential sawtimber quality was made. Multiple regression was used to estimate the impacts of individual traits on sawtimber potential assessment. The coefficients from the multiple regression were then used as weights in a selection index, and trait gain estimates were compared to gains achieved using selection for growth only and for the composite sawtimber potential trait.

The third experiment applied the results of the second experiment by using the sawtimber potential selection index as a trait for association mapping in addition to the individual traits of commercial importance that have been followed in the NCSU Cooperative Tree Improvement Program (NCSU-CTIP). Two hundred unrelated parents from the NCSU-CTIP were used to test 4200 SNPs for associations with tree height, wood volume, stem

straightness, rust incidence, stem forking and sawtimber index score. The effects of marker genotypes on wood volume were estimated for the association population. A second population consisting of 153 clonally-replicated genotypes was available to validate marker effects estimated for volume in the association population.

Literature Cited

Adams, G.W., and E.K. Morgenstern. 1991. Multiple-trait selection in jack pine. *Can. J. For. Res.* 21(4):439-445.

Albaugh, T.J., K.H. Johnsen, P.M. Dougherty, and H.L. Allen. 2004. Long term growth responses of loblolly pine to optimal nutrient and water resource availability. *Forest ecology and management* 192(1):3-19.

Busby, C.L. 1983. Crown-quality assessment and the relative economic importance of growth and crown characters in mature loblolly pine. *Proc. 17th Sou. For. Tree Imp. Conf.*(39):121-130.

Devey, M.E., S.D. Carson, M.F. Nolan, A.C. Matheson, C.T. Riini, and J. Hohepa. 2004. QTL associations for density and diameter in *Pinus radiata* and the potential for marker aided selection. *Theoretical and Applied Genetics* 108:516-524.

Dougherty, P.M., and C.A. Gresham. 1988. Conceptual analysis of southern pine plantation establishment and early growth. *Southern journal of applied forestry* 12(3):160-166.

Flint-Garcia, S.A., J.M. Thornsberry, and E.S.B. IV. 2003. Structure of Linkage Disequilibrium in Plants. *Annual Reviews in Plant Biology* 54:357-374.

Jayawickrama, K.J.S. 2001. Genetic parameter estimates for radiata pine in New Zealand and New South Wales: a synthesis of results. *Silvae genet.* 50(2):45-53.

Li, B., S. McKeand, and R. Weir. 2000. Impact of forest genetics on sustainable forestry--results from two cycles of loblolly pine breeding in the U.S. *J. Sust. For.* 10(1/2):79-85.

McKeand, S., and J. Svensson. 1997. Sustainable management of genetic resources. *Journal of forestry* 95(3):4-9.

McKeand, S., T. White, T. Byram, and T. Mullin. 2003. Deployment of genetically improved loblolly and slash pine 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. 2006a. What are the best loblolly pine genotypes worth to landowners? *J For.* 104:352-358.

McKeand, S.E., E.J. Jokela, D.A. Huber, T.D. Byram, H.L. Allen, B. Li, and T.J. Mullin. 2006b. Performance of improved genotypes of loblolly pine across different soils, climates, and silvicultural inputs [electronic resource]. *For. Ecol. Manag.* 227:178-184.

Neale, D.B., and O. Savolainen. 2004. Association genetics of complex traits in conifers. *Trends in Plant Sciences* 9(7):325-330.

Prestemon, J., and R.C. Abt. 2002. Timber products supply and demand. in The southern forest assessment final technical report. USDA Asheville, NC.

Stanturf, J.A., S.B. Jones, F.S. Broerman, and R.C. Kellison. 2003. Productivity of the southern pine plantations: where are we and how did we get here? *Journal of forestry* 101(3):26-31.

Timber Mart-South 2007. Available online at www.TimberMart-South.com/tmart/; last accessed June 1, 2007.

CHAPTER 1

Association genetics of carbon isotope discrimination, height, and foliar nitrogen in a natural population of *Pinus taeda* L.

(In the format appropriate for submission to Heredity)

Association genetics of carbon isotope discrimination, height, and foliar nitrogen in a natural population of *Pinus taeda* L.

Patrick Cumbie[1*], Andrew Eckert[2,3], Jill Wegrzyn[4], Ross Whetten[1],

David Neale[4,5], Barry Goldfarb[1]

[1] Department of Forestry and Environmental Resources
North Carolina State University
Campus Box 8002
Raleigh, NC 27695-8002

[2] Section of Evolution and Ecology
University of California at Davis
Davis, CA 95616, USA

[3] Center for Population Biology
University of California at Davis
Davis, CA 95616, USA

[4] Department of Plant Sciences
University of California at Davis
Davis, CA 95616, USA

[5]Institute of Forest Genetics
USDA Forest Service
Davis, CA 95616, USA

*Corresponding Author (email: patrick_cumbie@ncsu.edu)

Abstract: Water use efficiency is important for understanding the adaptability of forest trees to the variable environments in which they grow. Loblolly pine, *Pinus taeda* L., is one of the most widely planted, commercially and ecologically important tree species in North America. We took an association genetics approach using an unimproved population of 425 clonally replicated unrelated trees to test 3,938 single nucleotide polymorphisms (SNPs) for association with phenotypic variation in carbon isotope discrimination, total tree height, and foliar nitrogen concentration after two growing seasons. Best linear unbiased prediction was used with a spatial adjustment to remove additional environmental variation from phenotypic data derived from a common garden experiment. After correction for multiple testing a total of 14 SNPs were associated with carbon isotope discrimination (n = 7), height (n = 1), and foliar nitrogen concentration (n = 6). Tails of the population were compared for allele frequency differences revealing 10 SNPs with at least one tail significantly different from the overall population. Associated SNPs were in sequences similar to known genes such as an AP2 transcription factor related to carbon isotope discrimination and glutamate decarboxylase associated with foliar nitrogen concentration and others were from unknown genes without homologs in *Arabidopsis*. Results from this experiment demonstrate the utility of association genetics to explore the underlying genetic variation in quantitative traits in forest trees.

Keywords: water use efficiency, forest tree, population genetics, loblolly pine

Introduction

The limitation of available water is a critical factor affecting plant growth and survival. The regulation and control of water use by plants is a complex system, and is difficult to measure in a high-throughput assay suitable for a population-scale study. Carbon isotope discrimination (Δ) has been used in a wide range of plant species to assess water use efficiency. The ratio of C^{12} to C^{13} (δ) has been established as an indirect, but integrated, measure of stomatal conductance and photosynthetic activity (Farquhar *et al* 1989). The allocation of carbon in plants is dependent on several factors including water availability, nutrition, and light intensity, quality and duration (Hunt and Lloyd 1987). Plants generally favor above-ground growth when CO_2 and light are limiting, while below-ground growth is favored when water and nutrients are limiting. The response by plants to limitations in water has implications for both an understanding of adaptive variation in natural populations as well as economic importance for increasing production or yield in economically important plant species. Carbon isotope discrimination has also been utilized in crop species in an attempt to improve water use efficiency (Anyia *et al* 2007; Condon *et al* 2004; Rebetzke *et al* 2006).

Genetic variation in carbon Δ or δ has been reported for several species of forest trees. Foliar carbon isotope discrimination was under moderate genetic control in *Picea mariana* and was strongly correlated with growth, making it a candidate for indirect selection to improve growth (Johnsen *et al* 1999). Studies in *Araucaria cunninghamii* (Prasolova et al 2000), *Pinus elliotii x Pinus caribea hybrids* (Prasolova *et al* 2005) , and *Pinus pinaster*

(Brendel *et al* 2002) report variable levels of inheritance for carbon isotope discrimination in foliage and wood samples where individual tree h^2 *l*evels ranged from 0.07 to 0.72. Carbon isotope discrimination in loblolly pine appears to be under a lower level of genetic control when compared to *Picea*, *Araucaria*, and *Pinus pinaster*. Recent work by Baltunis (2008) reported a low individual tree heritability (h^2=0.09) from a clonally replicated, controlled mating design, field-grown experiment in loblolly pine. However, the population-wide genetic variation in carbon isotope discrimination has not been well documented in loblolly pine. Carbon isotope discrimination has been the target trait in genetic marker-based studies in other forest trees. Brendel (2002) reported four QTL explaining ~25% of the phenotypic variance for carbon isotope discrimination in *Pinus pinaster*. Previous association testing in loblolly pine found SNPs from four candidate genes (*dhn-1*, *sod-chl*, *wrky-like*, and *lp5-like*) potentially associated with carbon isotope discrimination (González-Martínez *et al* 2008), in which candidate sequences were selected based on putative functions in drought response. However, to date a large number of potential loci have not been tested in a large population for water use efficiency in loblolly pine.

Understanding the relationship between genotype and phenotype is essential for the improvement of complex traits in economically important plant species. Improvements in genomic technology and knowledge gained from research in model organisms are creating opportunities for large scale genomic research in commercially important species. Association genetics studies in humans have demonstrated the potential to discover DNA

sequence variants that are correlated with disease phenotypes (Smith and Newton-Cheh 2009). Both candidate gene and genome-wide approaches are potentially successful methods for the discovery of variants that are either causal, or are linked to, the causal variant for disease and quantitative traits (Hirschhorn and Daly 2005). Association studies for herbaceous plants have been successful in identifying polymorphisms related to phenotypic variation in adaptive traits in *Arabidopsis* (Chan *et al* 2009) as well as economically important traits in maize (Buckler *et al* 2009), sugar beet (Stich *et al* 2008b), and wheat (Jing *et al* 2007).

Forest trees have recently been used in several association studies (Eckert *et al* 2009c; González-Martínez *et al* 2008; González-Martínez *et al* 2007; Ingvarsson *et al* 2008). Conifers are well-suited for use in association genetics studies due to their large random mating populations, nucleotide diversity, rapid decay of linkage disequilibrium, and haploid tissue obtainable from seeds (Neale and Savolainen 2004). Economic and adaptive traits have been explored in conifers including Douglas-fir (*Pseudotsuga menziesii*) and loblolly pine and in angiosperms such as *Populus* (Neale and Ingvarsson 2008). Results of association studies for wood properties and carbon isotope discrimination in loblolly pine have been published revealing potential associations using a candidate gene approach (González-Martínez *et al* 2008; González-Martínez *et al* 2007). The objectives of this study were to estimate the variation in carbon isotope discrimination, height, and foliar nitrogen concentration in a population of unrelated *P. taeda* trees and to use an association genetics approach to identify SNPs related to the previously mentioned traits.

Materials & Methods

Plant material

Four hundred and twenty-five (n = 425) unrelated genotypes of loblolly pine were selected from the North Carolina State University Cooperative Tree Improvement Program and the Western Gulf Forest Tree Improvement Program at the Texas Forest Service (henceforth referred to as the NCSU population). Trees were grown from seed from natural stand selections from the first generation of improvement to represent the natural range of loblolly pine (Figure 1). A small number of plantation selection seedlots which had known seed sources were used to cover geographic areas from which natural stand selections were not available. All seeds were open-pollinated, except for a few control-pollinated seedlots. Control-pollinated seedlots came from 2^{nd}- or 3^{rd}-cycle mating, but care was taken to ensure no relatedness to other entries in the population. Trees were grown from seed for one year and then hedged for stem cutting production using established methods for loblolly pine (Lebude *et al* 2004). In the spring of 2006, rooted cuttings of each genotype were planted in a raised nursery bed comprised of a loamy sand from the coastal plain of North Carolina with a soil texture composed of 85% sand, 12.2% silt, and 2.8% clay (Gocke 2006). The bed was 1.5m x 40m with a soil depth of 30cm and a randomized complete block design was used, with two replicates planted per clone.

The trees were grown for two growing seasons and foliage was collected for carbon isotope discrimination after the end of the second growing season in December 2007. Isotope analysis was performed at the COIL Cornell stable isotope facility (www.cobsil.com). The

relative abundance of ^{13}C to ^{12}C was determined using 3mg samples of pine needle tissue. The carbon isotope ratio $\delta^{13}C$ was reported against a standard of Pee Dee Belemnite (Craig 1954) and then converted to Δ, the carbon isotope discrimination value (CID) using

[1] $$\Delta = \frac{\delta_a - \delta_p}{1 + \delta_p}$$

where δ_a is the atmospheric isotope composition (assumed to be -8‰) and δ_p is the leaf tissue isotope composition. Foliar nitrogen was estimated from mass spectroscopy at COIL. Total tree height (cm) was measured after the second growing season.

Analysis of phenotypic data

A two stage approach was used for association testing. Phenotypic data were analyzed using a mixed model with the standard form:

[2] $$\mathbf{y = Xb + Za + e}$$

where **y** is the vector of response variables (observations); **X** is the design matrix relating individual observations to the fixed effects in the model, **b** is the vector of fixed effect factors that includes the overall mean, site and replication within site effects; **Z** is the incidence matrix relating observations to random effects and **a** is the vector of random effects that includes clone effects and interactions with blocks; and **e** is the vector of random residual terms. The expectation of fixed terms is E(**y**) = **Xb**. The random terms are assumed to have zero means and variances var(**a**)=**G**, var(**e**)=**R.** The variance-covariance

matrix of observations (vector y) is Var(**y**) = $\mathbf{ZGZ}^T$ + **R** where **G** and **R** are covariance matrices corresponding to **a** and **e**, respectively. The **G** matrix accounts for the genetic effects, while **e** accounts for random residual effects. The **G** is a block diagonal matrix defined as $\mathbf{I}\sigma^2_C$ for the variance among clones and $\mathbf{I}\sigma^2_{CB}$ is the variance of the clone by block interaction where I is an identity matrix of dimension n_i x n_i (n_i = number of levels of the ith term). The $\mathbf{R}=\mathbf{I}\sigma^2_e$ is a diagonal matrix with the residual error variances (σ^2_e) in the diagonal and zero covariances (null submatrix) in the off diagonals when errors are independent; **I** is the identity matrix of dimension equal to the number of observations. In this experiment a spatial residual structure was implemented which divides **e** into spatially dependent (ξ) and spatially independent (η) residuals (Dutkowski *et al* 2002). For spatially dependent residuals a covariance structure was specified using a first-order autoregressive process in rows and columns:

[3] $$\mathbf{R} = \sigma^2_{\xi}[AR1(\rho_{col}) \otimes AR1(\rho_{row})] + \sigma^2_{\eta}\mathbf{I},$$

where σ^2_{ξ} is the spatial residual variance, σ^2_{η} is the independent residual variance, **I** is an identity matrix equal to the number of observations, and the error term is included in the individual residual variance. AR1(ρ) is a first order autoregressive correlation matrix with the form:

[4] $$\mathrm{AR1}(\rho) = \begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^n \\ \rho & 1 & \rho & \cdots & \\ \rho^2 & \rho & 1 & \dots & \\ \vdots & \vdots & \vdots & \ddots & \\ \rho^n & \dots & \dots & \dots & 1 \end{bmatrix}$$

Clone mean heritability was estimated using the formula:

[5] $$H_c^2 = \frac{\sigma_C^2}{\sigma_C^2 + \frac{\sigma_{CB}^2}{b} + \frac{\sigma_E^2}{bt}}$$

where σ^2_C is the variance among clones, $\sigma^2_{CB}$is the variance due to the clone by block interaction, σ^2_E is the residual variance after spatial residuals are removed, *b* is the number of blocks (2), and *t* is the trees per plot (1). Standard errors of the heritability estimates were calculated using the Taylor series expansion (Gilmour *et al* 2006). The best linear unbiased prediction (BLUP) values for clonal genotypes were used as phenotypes in the association analysis. Bivariate analyses were used to estimate genetic covariances among traits, where **G** and **R** include variances and covariances of the two traits. Genetic correlations among clones were estimated using the formula:

[6] $$r_{g(XY)} = \frac{\sigma_{XY}}{\sqrt{\sigma^2_{g(X)}\sigma^2_{g(Y)}}}$$

where $r_{g(xy)}$ is the genetic correlation between traits x and y; σ_{xy} is the genetic (clonal) covariance between traits; $\sigma^2_{g(x)}$ is the clonal variance for trait x; and $\sigma^2_{g(y)}$ is the clonal

variance for trait y. All statistical analysis was performed using the ASReml 2.0 statistical software package (Gilmour *et al* 2006).

Genotypic Data

Genotypes for single nucleotide polymorphisms (SNPs) were obtained using the Illumina Infinium™ assay (Illumina, San Diego, CA). Smaller Illumina platforms have been shown to work well within the large and complex genome of conifers (Eckert *et al* 2009a). The development of these SNPs is available online at http://dendrome.ucdavis.edu/adept2/. Briefly, SNPs were detected and genotyped for 7508 resequenced amplicons generated from all available unique EST contigs representing all pine ESTs known to date using an Infinium™ genotyping chip. From the resequenced amplicons, roughly 22,000 SNPs were discovered and 7216 were selected for genotyping. Based on quality, reliability of reads and polymorphism, 3938 SNPs were selected using the BeadStudio ver. 3.1.3.0 software (Illumina), Genotypic data were available for 380 of the 425 clones measured in this experiment.

Association Analyses

Association testing was performed in TASSEL using both a general linear model (GLM) and a mixed model (MLM) approach (Bradbury *et al* 2007). Population structure covariates were estimated from a set of 23 nuclear microsatellite markers using STRUCTURE with a cluster number of five (Eckert et al., in review). Marker based kinship was estimated using a function in the EMMA (Efficient Mixed Model Analysis)

package (Kang *et al* 2008) in the R programming environment (RDevelopmentCoreTeam 2005). We used the positive false discovery rate approach to adjust p-values for multiple testing (Storey 2003), using the qvalue package in R with a false discovery rate of 0.05.

We chose to compare the use of population and marker-based kinship to determine if these factors enhanced analysis of the NCSU association population. Yu et al (2006) demonstrated the value of accounting for population structure and relatedness through the incorporation of genomic control and marker-based kinship in mixed model association testing. We compared observed p-values from association testing against a uniform distribution of expected p-values using the mean of the squared differences (MSD) of potential models for association testing (Stich *et al* 2008a). We compared four models to examine the distribution of p-values from association tests: a general linear model with no structure or kinship effects (GLM), a general linear model with covariates to account for population structure (Q), a mixed model with a marker-based kinship component (K), and a mixed model that incorporated both population structure and marker-based kinship estimates (QK).

SNP effects and Genotypic Frequencies

For each associated SNP, we estimated the additive (a) and dominance (d) effects using ASReml 2.0. SNPs were treated as fixed effects to test for significance and to generate best linear unbiased estimates (BLUEs) for genotype classes of each SNP. Additionally, variance estimates of significant SNPs with three genotypic classes were generated by adding a SNP effect into the mixed model analyzing phenotypic data, where the SNP effect

was treated as a random effect with 2 degrees of freedom. To further evaluate the effect of potential associations with phenotypes, we compared the genotypic frequencies of the population tails to the frequencies observed in the entire population. The population tails were truncated at > 1.5 standard deviations above and below the population mean for each trait. Genotypic frequencies and tests for Hardy-Weinberg Equilibrium were performed in the Allele procedure in SAS (SAS 1989).

To obtain annotations for SNPs, flanking sequences of the corresponding EST contig were obtained from the Dendrome database (http://dendrome.ucdavis.edu/treegenes) for all SNPs associated with a trait after multiple testing correction. We performed a BLASTx query against the NCBI non-redundant protein database (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Gene and site annotations for the strongest hits (lowest e-value) for each sequence are reported.

Results

Phenotypic variation

Carbon isotope discrimination, height, and %N displayed significant variation among clones. Use of the spatial model allowed additional environmental variation to be removed, which enhanced heritability estimates and removed environmental bias from the clonal BLUP values. Figure 2 displays a variogram of the residuals from the spatial model indicating spatial trends in the nursery bed that would not have been accounted for without the spatial model in the two-replicate design. Broad-sense clone mean heritability estimates

(H^2_c) revealed that 50% of the variation in $\Delta^{13}C$ (Table 1) and 43% of the variation in second-year height and %N were due to variation among clones. Bivariate analyses (Table 2) revealed that carbon isotope discrimination was moderately correlated with height ($r_{g(xy)}$ = 0.38) and %N ($r_{g(xy)}$ = 0.42), while height and %N were weakly correlated ($r_{g(xy)}$ = 0.22). Clonal BLUPs for carbon isotope discrimination ranged from 20.7 to 23.4 with a population mean of 22.11. Height ranged from 116 cm to 240 cm with a mean of 194 cm and %N ranged from 1.15 to 1.52 with a mean of 1.40. Best linear unbiased predictions were used as the phenotype in subsequent association testing.

SNP-Trait Associations

Significant associations were found for all traits. Prior to multiple testing correction the number of associations significant at the $p < 0.05$ level ranged from 144 (K) to 193 (GLM) for $\delta^{13}C$, 133 (K and QK) to 172 (GLM) for height, and 135 (QK) to 176 (GLM) for %N, across the four different models tested. After correction for multiple testing, we found four to six SNPs associated with $\delta^{13}C$, one SNP associated with height, and five to six SNPs associated with %N, depending on the model used for testing. For each trait, the model with the highest number of significant associations following multiple testing correction was reported: QK for $\Delta^{13}C$, GLM for height, and GLM for %N. The four models used in this analysis identified similar numbers of SNPs as associated with all traits at the four levels of p-value thresholds (see supplemental information).

The minor allele frequencies for SNPs that remained significant after multiple testing correction ranged from 2% to 35% (Table 4), suggesting that associations are from generally rare alleles. The associations are largely caused by SNPs segregating rare alleles (MAF 2-35%) with small effects (r^2: 4-9%) on $\delta^{13}C$, height, and %N. All significantly associated SNPs failed to depart from Hardy Weinberg Equilibrium except for SNP 2_1501_01_109. Additive effects were significantly different from zero for six loci associated with $\delta^{13}C$ and five loci associated with %N. Dominance effects were significant for the locus associated with height, four of the loci associated with $\delta^{13}C$, and five of the loci associated with %N (Table 4).

SNP Effects

When individual SNPs were used in a mixed model as an independent random effect with clones, no SNP variance estimates were significantly different from zero using a chi-square test of -2logliklihood model values (not shown). Marker r^2 values for significantly associated SNPs ranged from 0.04 to 0.08 indicating that individual SNP loci account for small amounts of the variation among clones (Table 4). Using a model that included the effect of clones and individual SNPs as random effects, individual SNP effects accounted for small amounts of the phenotypic variance ranging from less than 0.001% up to 3.2% in $\delta^{13}C$ and as high as 7% for %N.

Additive and dominance estimates were estimated for all SNPs observed to be in Hardy Weinberg Equilibrium with all three genotype classes present. Additive effects were

significant for six SNPs associated with $\Delta^{13}C$ but the BLUE estimates of SNP effects were not precise as standard errors were greater than half the value of the estimate (Table 4). The ratio of dominance to additive effects (d/a) showed dominance effects similar in magnitude to additive effects ranging from -0.79 to 1.30 for $\Delta^{13}C$ (Table 5). Dominance effects for %N were similar in range (-0.75 to -1.04) with the exception of locus *CL1074Contig1_03_101* where dominance was only 1% of the additive effect. The dominance effect for tree height at locus 0_14415_01_190 was the largest dominance effect observed at more than three times the additive effect (Table 4).

Genotypic Frequencies

We tested for differences in genotypic frequency in the tails of the population for all loci significantly associated with $\delta^{13}C$, height, and %N. For each trait, we compared the extremes of the population, which we considered to be clonal BLUP values greater than 1.5 standard deviations above and below the mean of clonal values. Using the population mean as the expected genotypic frequencies we observed nine loci with significant frequency changes ($p<0.05$) from the mean for one of the phenotypic tails (Table 5). For each trait, we observed SNPs with departures from the overall population genotype frequencies (one SNP for height (0_14415_01_190), four SNPs for $\delta^{13}C$ (0_17030_01_94, 0_10921_01_353, CL599Contig1_07_109, and 0_8304_02_414), and five SNPs for %N (0_17195_01_417, 2_1087_01_86, 2_4191_01_104, 2_7865_01_156, and CL1074Contig1_03_101). The single SNP that survived multiple testing correction for height displayed an increased

frequency in the heterozygous class for the tail with lower height ($p<0.01$), while SNPs for $\delta^{13}C$ and %N reflected changes in genotypes with the minor alleles in one tail.

Annotations for associated SNPs

Association analyses revealed SNPs located in regions of sequences of previously described function and also in unknown sequences (Table 4). BLASTx queries using the flanking sequences around the associated SNPs found similar proteins in the Genbank database for six out of 14 SNPs. Putative orthologs were a mitochondrial protein (0_8304_02_414), a heme activated DNA-binding protein (CL599Contig1_07_109), and an AP2 domain transcription factor (0_3648_01_357) for $\delta^{13}C$-associated SNPs. For %N-associated SNPs, these analyses revealed a receptor protein kinase-like protein (2_7865_01_156), and glutamate decarboxylase (0_17195_01_417) (Table 4).

Discussion

The ability for plants to respond to different levels of available water is variable and complex. Members of the genus *Pinus* employ a drought avoidance strategy where under well-watered conditions water use is maximized, but decreases quickly when water is limiting to avoid low water potential (Martinez-Vilalta *et al* 2004). The population of 425 unrelated loblolly pine clones in this study displayed a substantial amount of variation for carbon isotope discrimination. Moderate correlations indicate an increase in height growth and nitrogen content in foliage as $\delta^{13}C$ increases, supporting a mechanism of drought avoidance in loblolly pine. Previous analyses of height and carbon isotope discrimination

revealed similar results suggesting that carbon isotope analysis may be used to improve water use efficiency in loblolly pine populations (Baltunis *et al* 2008).

Significantly associated SNPs in this experiment were found in both known and unknown gene sequences and in some cases SNPs were in functional genes related to similar traits in other plant species. The sequence flanking SNP 0_3648_01_357 was similar to an AP2 domain transcription factor. The AP2 domain family of transcription factors has been associated with ABA-sensitive abiotic stress response in *Arabidopsis* (Finkelstein *et al* 1998) in seed and leaf tissue. In *Pinus strobus* the over-expression of the ERF/AP2 transcription factor *CaPF1* conferred increased drought and freeze tolerance in young plants (Tang *et al* 2007). SNP 0_3648_01_357 results in a nonsynonomous codon change, and accounted for 3.8% of the phenotypic variation in this experiment suggesting that a response to stress is linked to carbon isotope discrimination levels in this experiment. Our results are consistent with those of González-Martínez et al. (2008) with new SNPs in previously unidentified sequences as well as those found in known sequences. Sequence similarity search results from this study and candidate genes from the work of González-Martínez et al. (2008) in carbon isotope association testing in loblolly pine reveal functional annotations of genes that are involved in abiotic stress response rather than growth functions, which suggests that Δ is more closely related to stomatal conductance than photosynthetic capacity. In both studies, the most significant SNPs accounted for small portions of the phenotypic variance supporting a polygenic response to regulate water in loblolly pine.

Nitrogen is critical to plant growth and photosynthetic activities, and it is often a limiting nutrient for growth in forest plantations (Fox *et al* 2007). To date, genes in loblolly pine affecting the use or uptake of nitrogen have not yet been identified. The sequence flanking SNP 0_17195_01_417 is similar to glutamate decarboxylase (*GAD*) which is involved in the production of GABA, nitrogen metabolism and C:N ratio balance in *Arabidopsis* (Bouché and Fromm 2004). SNP 2_7865_01_156 was identified as a receptor-like protein kinase-like (RLK) protein, a class of proteins which are important for many plant functions including nitrogen fixation in legumes such as alfalfa, soybean, and peas (Morris and Walker 2003).

A spatial model was implemented to remove environmental variability that would be not removed by the experimental design. The individual SNP models regressed the BLUP of each clone for each trait on individual SNPs. BLUP predictions were based on 2 replicates of each genotype planted in a relatively small, homogenous nursery bed. R-square values and the strength of any relationship between a SNP polymorphism and clonal BLUP values may not be repeatable in a larger, more heterogeneous field trial, which is typical of forest tree progeny trials. Future experiments should incorporate efficient designs to remove environmental variation and improve the precision of phenotypes used in association studies.

Association analyses have been performed in populations of loblolly pine for wood quality traits (González-Martínez *et al* 2007), carbon isotope discrimination (González-Martínez *et al* 2008), and cold-tolerance in Douglas-fir (Eckert *et al* 2009c) using a candidate gene

approach. To date, associated SNPs for traits in loblolly pine explained the highest proportion of phenotypic variance in wood quality traits (20% of phenotypic variance) while only 7% of the phenotypic variance was accounted for in this experiment for the $\delta^{13}C$-associated SNPs. Recent association analyses in conifers have explained small proportions of the phenotypic variation with individual SNPs (Eckert *et al* 2009c; González-Martínez *et al* 2008; González-Martínez *et al* 2007), supporting the treatment of these traits as polygenic and quantitative in nature (Falconer 1989). The results of this association analysis support the treatment of $\delta^{13}C$, height, and %N as polygenic traits in conifers, and provide evidence to support further analysis of the underlying genetic complexity of these quantitative traits in forest trees.

Improved methods and larger populations are needed to have greater precision in estimates of marker effects. F-tests show significant additive and dominance effects at several SNP loci associated with measured phenotypes, but best linear unbiased estimates (BLUE) of the genotypic effects revealed large standard errors. The estimated magnitudes of allelic effects in this population are likely to be biased upward, because effects are estimated from a truncated distribution (Xu 2003). In addition, the majority of minor allele frequencies observed for significant SNPs in this population were low (0.05 to 0.10) with one exception at 0.35, thus the small number of observations with these alleles were more heavily weighted in the estimation of allelic effects and variances. A larger population is likely to more accurately estimate the true effect of any associated SNP polymorphism.

In this population the magnitude of the dominance effects was similar to that of additive effects. Excluding two SNPs, 0_14415_01_190, which was associated with height, and CL1074Contig1_03_101, which was associated with %N, the ratio of dominance to additive ranged from 0.75 to 1.3 (Table 5). Eckert (2009c) reported similar ratios of dominance to additive effects for cold-tolerance-related traits in Douglas-fir, where most dominance effects were 0.8x to 1.3x the additive effect. Additive effects have been the focus of tree breeding programs for loblolly pine but the use of non-additive effects may be an opportunity for capturing desirable trait attributes in deployment populations.

The largest dominance effect observed in this study was the effect of SNP 0_14415_01_190 on height. This effect is supported by the increased number of heterozygotes in the low-end tail for height in the population (Table 5). This difference in genotypic frequency was significant ($p<0.01$), and the proportion of heterozygotes in the low-end tail increases from 0.41 in the entire population to 0.68 in the low-end tail. However, variance in height due to this SNP was very small (<0.001%) and was not significant, highlighting the challenge to discover variants and estimate the magnitude of their effects in field trials. Eckert et al. (2009c) reported that 86% of the SNPs tested for associations with cold tolerance in Douglas-fir showed non-additive effects. If such SNPs are validated and account for significant variation, non-additive effects could be utilized by deploying specific clones to improve growth.

We observed significant changes in genotypic frequency between at least one tail of the population and the overall population for nine SNPs, lending support that the SNP variants

are having some impact on the phenotypes observed in this study. Finding significant changes in allele frequency between the tails of the population suggests that a pooled sampling approach may be of value in forest trees. Pooling individuals based on phenotype for SNP discovery and allele frequency estimation has shown promise in effectively discovering polymorphic SNPs and estimating allele frequencies within the pooled phenotypic classes in human blood and disease studies (Craig *et al* 2009; Druley *et al* 2009). Such an approach could significantly reduce genotyping cost as compared to association studies in large populations; and identify candidate SNPs for future studies based upon allele frequency differences.

Recent characterization of population structure in loblolly pine, which included the NCSU population, identified 24 SNPs as F_{st} outliers and 5 loci associated with geographic variation for potential evapotranspiration (Eckert *et al* 2009b). Two of the F_{st} outliers were associated with traits in this analysis: 0_14415_01_190 for height and 2_1087_01_86 for %N. Selection may be influencing these two SNPs in this population by either having an effect on the alleles of the genes in which these SNPs are located or other alleles that are in linkage disequilibrium with these SNPs. Since 90% of the genes in loblolly pine have not been sequenced the amount of bias that may be introduced is unknown and requires further scrutiny.

An understanding of the genes involved in carbon isotope discrimination and foliar nitrogen content are valuable for the analysis of both the environmental services provided by natural forests managed with low intensity and wood productivity in more intensively managed

plantation forests. Understanding the genetic variation of water related traits in forest trees is also important in light of potential climate change. Stomatal conductance and leaf hydraulic conductance were reduced when trees were exposed to increased CO_2 (Domec *et al* 2009). The underlying genetic variation in water use efficiency may be valuable for modeling forest stand dynamics as well as for producing more wood products in a sustainable and efficient manner. In the future we will likely be able to associate alleles of genes with more specific traits to dissect complex traits of economic and ecological value (Nelson and Johnsen 2008). Carbon isotopes have shown promise for use in improving water use efficiency in pine species, but adequate sampling across environments should be taken into account. Tools such as spatial models may be useful to remove additional environmental variation to improve phenotypic data for future genomic studies.

Conclusion

We identified 14 new marker-trait associations for $\Delta^{13}C$, height, and %N in loblolly pine. Results suggest that $\Delta^{13}C$, height, and %N are under polygenic control in loblolly pine, with the effect of each associated SNP having a small impact on the observed phenotype. To date, data largely point to additive variation to explain complex traits (Hill *et al* 2008), but results from this study suggest that there are both additive and non-additive effects on adaptive and potentially economic traits. The application of analytical tools such as spatial models will help remove the environmental variation from phenotypic data and may help strengthen future association testing in field grown trees. Traditional breeding methods based on quantitative genetic approaches have been successful in the improvement of

loblolly pine populations, but the development of forest tree genomic resources will aid the improvement and understanding of ecological and economically valuable traits in forest trees.

Acknowledgements

The authors wish to thank the members and staff of the North Carolina State University Cooperative Tree Improvement Program and the Western Gulf Forest Tree Improvement Program, Texas Forest Service for their contribution of germplasm to this project.; Dr. Fikret Isik for consultation on statistical analysis; and Dr. Anthony Lebude for assistance in vegetative propagation. This work was supported by the National Science Foundation (grant DBI-0501763).

Conflict of interest.

The authors declare no conflict of interest.

Literature Cited

Anyia AO, Slaski JJ, Nyachiro JM, Archambault DJ, Juskiw P (2007). Relationship of Carbon Isotope Discrimination to Water Use Efficiency and Productivity of Barley Under Field and Greenhouse Conditions *J Agron Crop Sci* 193: 313-323.

Baltunis BS, Davis JM, Huber DA, Martin TA (2008). Inheritance of foliar stable carbon isotope discrimination and third-year height in Pinus taeda clones on contrasting sites in Florida and Georgia [electronic resource]. *Tree gen & genom* 4: 797-807.

Bouché N, Fromm H (2004). GABA in plants: just a metabolite? *Trends Plant Sci* 9(3): 110-115.

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007). TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* 23(19): 2633-2635.

Brendel O, Pot D, Plomion C, Rozenberg P, Guehl JM (2002). Genetic parameters and QTL analysis of delta13C and ring width in maritime pine. *Plant, Cell Environ* 25(8): 945-953.

Buckler ES, Holland JB, Bradbury PJ, Acharya CB, Brown PJ, Browne C *et al* (2009). The Genetic Architecture of Maize Flowering Time. *Science* 325(5941): 714-718.

Chan EKF, Rowe HC, Kliebenstein DJ (2009). Understanding the Evolution of Defense Metabolites in Arabidopsis thaliana Using Genome-Wide Association Mapping. genetics.109.108522.

Condon AG, Farquhar GD, Rebetzke GJ, Richards RA (2004). Breeding for high water-use efficiency. *J Exp Bot* 55: 2447-2460.

Craig H (1954). Carbon-13 in plants and the relationship between carbon-13 and carbon-14 variations in nature. *J Geology* 62: 115-149.

Craig JE, Hewitt AW, McMellon AE, Henders AK, Ma L, Wallace L *et al* (2009). Rapid inexpensive genome-wide association using pooled whole blood. *Genome Res* 19(11): 2075-2080.

Domec J-C, Palmroth S, Ward E, Maier CA, Theresien M, Oren R (2009). Acclimation of leaf hydraulic conductance and stomatal conductance of Pinus taeda (loblolly pine) to long-term growth in elevated CO_2 (free-air CO_2 enrichment) and N-fertilization. *Plant, Cell Environ* 32(11): 1500-1512.

Druley TE, Vallania FLM, Wegner DJ, Varley KE, Knowles OL, Bonds JA *et al* (2009). Quantification of rare allelic variants from pooled genomic DNA. *Nat Meth* 6(4): 263-265.

Dutkowski GW, Costa E Silva J, Gilmour AR, Lopez GA (2002). Spatial analysis methods for forest genetic trials. *Can J For Res* 32(12): 2201-2214.

Eckert A, Bower AD, Wegrzyn JL, Pande B, Jermstad KD, Krutovsky KV *et al* (2009a). Association Genetics of Coastal Douglas-fir (Pseudotsuga menziesii var. menziesii, Pinaceae) I. Cold-hardiness Related Traits.

Eckert A, Heerwaarden Jv, Wegrzyn J, Nelson CD, Ross-Ibarra J, González-Martínez SC *et al* (2009b). Patterns of population structure and environmental associations with aridity across the range of loblolly pine (Pinus taeda L., Pinaceae). *Genetics* in press.

Eckert AJ, Bower AD, Wegrzyn JL, Pande B, Jermstad KD, Krutovsky KV *et al* (2009c). Association Genetics of Coastal Douglas Fir (Pseudotsuga menziesii var. menziesii, Pinaceae). I. Cold-Hardiness Related Traits. *Genetics* 182(4): 1289-1302.

Falconer DS (1989). Introduction to quantitative genetics Burnt Mill, Harlow, Essex, England : Longman, Scientific & Technical ; New York : Wiley, 1989. Vol. 3rd ed.

Farquhar GD, Ehleringer JR, Hubick KT (1989). Carbon isotope discrimination and photosynthesis. *Annu Rev Plant Physiol Plant Mol Biol*: 503-537.

Finkelstein RR, Wang ML, Lynch TJ, Rao S, Goodman HM (1998). The Arabidopsis Abscisic Acid Response Locus ABI4 Encodes an APETALA 2 Domain Protein. *Plant Cell* 10(6): 1043-1054.

Fox TR, Allen HL, Albaugh TJ, Rubilar R, Carlson CA (2007). Tree Nutrition and Forest Fertilization of Pine Plantations in the Southern United States. *South J Appl For* 31: 5-11.

Gilmour AR, Gogel BJ, Cullis BR, Thompson R. (2006). VSN International Ltd: Hemel Hempstead, UK, pp 320pp.

Gocke MH (2006). Production system influences the survival and morphology of rooted stem cuttings of loblolly pine (Pinus taeda L.) and sweetgum (Liquidambar styraciflua L.). Master of Science thesis, North Carolina State University, Raleigh.

González-Martínez SC, Huber D, Ersoz E, Davis JM, Neale DB (2008). Association genetics in Pinus taeda L. II. Carbon isotope discrimination. *Heredity* 101(1): 19-26.

González-Martínez SC, Wheeler NC, Ersoz E, Nelson CD, Neale DB (2007). Association Genetics in Pinus taeda L. I. Wood Property Traits. *Genetics* 175(1): 399-409.

Hill WG, Goddard ME, Visscher PM (2008). Data and Theory Point to Mainly Additive Genetic Variance for Complex Traits. *PLoS Genet* 4(2): e1000008.

Hirschhorn JN, Daly MJ (2005). Genome-wide association studies for common diseases and complex traits. *Nat Rev Genet* 6(2): 95-108.

Hunt R, Lloyd PS (1987). Growth and partitioning. *New Phytol* 106(1,suppl.): 235-249.

Ingvarsson PK, Garcia MV, Luquez V, Hall D, Jansson S (2008). Nucleotide polymoirphism and phenotypic associations within and around the phytochrome B2 locus in European aspen (Populus tremula, Salicaceae). *Genetics* 178(4): 2217-2226.

Jing H-C, Kornyukhin D, Kanyuka K, Orford S, Zlatska A, Mitrofanova OP *et al* (2007). Identification of variation in adaptively important traits and genome-wide analysis of trait marker associations in Triticum monococcum. *J Exp Bot* 58(13): 3749-3764.

Johnsen KH, Flanagan LB, Huber DA, Major JE (1999). Genetic variation in growth, carbon isotope discrimination, and foliar N concentration in Picea mariana: analyses from a half-diallel mating design using field-grown trees. *Can J For Res* 29(11): 1727-1735.

Kang HM, Zaitlen NA, Wade CM, Kirby A, Heckerman D, Daly MJ *et al* (2008). Efficient Control of Population Structure in Model Organism Association Mapping. *Genetics* 178(3): 1709-1723.

Lebude AV, Goldfarb B, Blazich FA, Wise FC, Frampton J (2004). Mist, substrate water potential and cutting water potential influence rooting of stem cutting of loblolly pine. *Tree Physiol* 24: 823-831.

Martinez-Vilalta J, Sala A, Pinol J (2004). The hydraulic architecture of Pinaceae - a review. *Plant ecol* 171: 3-13.

Morris ER, Walker JC (2003). Receptor-like protein kinases: the keys to response. *Curr Opin Plant Biol* 6(4): 339-342.

Neale DB, Ingvarsson PrK (2008). Population, quantitative and comparative genomics of adaptation in forest trees. *Curr Opin Plant Biol* 11: 149-155.

Neale DB, Savolainen O (2004). Association genetics of complex traits in conifers. *Trends Plant Sci* 9(7): 325-330.

Nelson CD, Johnsen KH (2008). Genomic and physiological approaches to advancing forest tree improvement. *Tree physiol* 28(7): 1135-1143.

Prasolova MV, Xu ZH, Farquhar GD, Saffigna PG, Dieters MJ (2000). Variation in branchlet delta13C in relation to branchlet nitrogen concentration and growth in 8-year-old hoop pine families (Araucaria cunninghamii) in subtropical Australia. *Tree physiol*(15): 1049-1055.

Prasolova NV, Prasolova NV, Lundkvist K, Xu ZH (2005). Genetic variation in foliar nutrient concentration in relation to foliar carbon isotope composition and tree growth with clones of the F1 hybrid between slash pine and Caribbean pine. *For Ecol Manage* 210: 173-191.

RDevelopmentCoreTeam. (2005). Computing RFfS (ed.): Vienna, Austria.

Rebetzke GJ, Richards RA, Condon AG, Farquhar GD (2006). Inheritance of Carbon Isotope Discrimination in Bread Wheat (Triticum Aestivum L.). 150: 97-106.

SAS (1989). *SAS/STAT User's Guide, Version 6*, 4th edn. SAS Institute Inc.: Cary, NC.
Smith JG, Newton-Cheh C (2009). Genome-wide association study in humans. *Methods Mol Biol (Totowa, NJ, U S)* 573: 231-258.

Stich B, Mohring J, Piepho H-P, Heckenberger M, Buckler ES, Melchinger AE (2008a). Comparison of Mixed-Model Approaches for Association Mapping. *Genetics* 178(3): 1745-1754.

Stich B, Piepho H-P, Schulz B, Melchinger A (2008b). Multi-trait association mapping in sugar beet (Beta vulgaris L.). *Theor Appl Genet* 117(6): 947-954.

Storey JD (2003). The positive false discovery rate: A Bayesian interpretation and the q-value. *Ann Stat* 31(6): 2013-2035.

Tang W, Newton R, Li C, Charles T (2007). Enhanced stress tolerance in transgenic pine expressing the pepper CaPF1 gene is associated with the polyamine biosynthesis. *Plant Cell Rep* 26(1): 115-124.

Xu S (2003). Theoretical basis of the Beavis effect. *Genetics* 165(4): 2259-2268.

Yu J, Pressoir G, Briggs WH, Vroh Bi I, Yamasaki M, Doebley JF *et al* (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat Genet* 38(2): 203-208.

Table 1. Phenotypic mean (Mean $_{Pop}$) and standard error (Std. Error $_{Pop}$), clonal variance (σ^2_c), error variance (σ^2_ε), and clone mean heritability (H^2_c) and standard errors (H^2_c Std. Error) for $\delta^{13}C$, Height, and %Nitrogen.

Trait	Mean $_{Pop}$	Std. Error $_{Pop}$	σ^2_c	σ^2_ε	H^2_c	H^2_c Std. Error
$\Delta^{13}C$	21.46	0.05	0.25	0.37	0.50	0.05
Height (cm)	194	14	1440	2787	0.43	0.05
Nitrogen (%)	1.41	0.08	0.0064	0.0128	0.42	0.06

Table 2. Phenotypic (below-diagonal) and Clonal (above-diagonal) correlations (and standard errors) among $\delta^{13}C$, Height, and %N.

	$\Delta^{13}C$	Height	%N
$\Delta^{13}C$	-	-0.37 (0.22)	-0.42 (0.21)
Height	0.02 (0.04)	-	0.32 (0.16)
%N	-0.55 (0.03)	0.04 (0.04)	-

Table 3. SNP loci annotations and significance values[1] for second-year height (Ht),carbon isotope ratio ($\delta^{13}C$), and foliar nitrogen level (%N); nc noncoding; nc nonsynonymous; syn synonymous.

SNP Locus	Trait	Annotation	P value	Q value	E-value
0_14415_01_190^{NC}	Ht	Unknown	2.38E-05	8.87E-02	-----
0_17030_01_94^{NC}	$\Delta^{13}C$	Unknown	4.95E-07	1.91E-05	-----
0_8304_02_414^{NS}	$\Delta^{13}C$	mitochondrial protein	2.84E-05	4.04E-02	5E-66
0_17543_01_196^{SYN}	$\Delta^{13}C$	Unknown	3.98E-05	2.97E-03	-----
0_10921_01_353^{NC}	$\Delta^{13}C$	Unknown	8.67E-05	7.88E-03	-----
2_1501_01_109^{NC}	$\Delta^{13}C$	Unknown	1.43E-05	1.97E-02	-----
CL599Contig1_07_109^{SYN}	$\Delta^{13}C$	putative heme activated protein	3.79E-04	2.24E-02	1E-53
0_3648_01_357^{NS}	$\Delta^{13}C$	AP2 domain transcription factor	6.26E-04	3.63E-02	1E-20
UMN_6338_01_99^{NS}	%N	Predicted protein	2.05E-07	2.11E-04	7E-58
2_7865_01_156^{NC}	%N	receptor protein kinase-like protein	3.53E-07	2.11E-04	9E-29
2_1087_01_86^{NC}	%N	Hypothetical protein	1.98E-06	1.08E-03	5E-37
CL1074Contig1_03_101^{NC}	%N	predicted protein	6.17E-06	2.77E-03	5E-54
0_17195_01_417^{NC}	%N	glutamate decarboxylase	1.89E-05	4.58E-03	3E-47
2_4191_01_104^{NC}	%N	Unknown	1.55E-04	2.71E-02	-----

[1] P-values and Q-values are reported for the model which yielded the most significant associations ($Q<0.05$) for individual traits.

Table 4. Description of significant SNP loci, minor allele frequencies (MAF), r^2, %variance estimates, additive and dominance effects.

SNP Locus	Trait	SNP[A]	MAF	r^{2}_{B}	% Phenotypic Variance explained by SNP	Additive Effect[C] a	Dominance Effect[D] d
0_14415_01_190	Ht	[T/A]	0.35	0.06	<0.01%	-1.28	-3.99*
0_17030_01_94	$\Delta^{13}C$	[A/G]	0.05	0.08	0.4%	0.83*	-0.79*
0_8304_02_414	$\Delta^{13}C$	[A/G]	0.11	0.06	<0.01%	0.34*	-0.27*
0_17543_01_196	$\Delta^{13}C$	[A/G]	0.08	0.06	<0.01%	0.51*	-0.47
0_10921_01_353	$\Delta^{13}C$	[A/G]	0.06	0.05	0.2%	0.47*	-0.48
2_1501_01_109	$\Delta^{13}C$	[A/G]	0.02	0.05	NE[F]	NE	NE
CL599Contig1_07_109	$\Delta^{13}C$	[G/C]	0.11	0.05	2.7%	0.28*	-0.28*
0_3648_01_357	$\Delta^{13}C$	[G/A]	0.09	0.04	3.8%	0.24*	0.31*
UMN_6338_01_99	%N	[A/T]	0.06	0.09	<0.01%	-0.10*	0.11*
2_7865_01_156	%N	[A/G]	0.06	0.08	NA	-0.10*	0.09*
2_1087_01_86	%N	[G/A]	0.05	0.07	<0.01%	-0.06	0.05*
CL1074Contig1_03_101	%N	[A/G]	0.03	0.07	<0.01%	0.04*	<-0.01
0_17195_01_417	%N	[A/C]	0.04	0.06	7.0%	-0.06*	0.04*
2_4191_01_104	%N	[G/A]	0.08	0.05	<0.01%	-0.05*	0.05*

[A] SNPs listed in order of common allele/minor allele for each locus

[B] r-square of the markers from the single marker models regressed on Clonal BLUP values.

[C] *Additive effect is significantly different from zero from ANOVA ($p<0.05$)

[D] *Dominance effect is significantly different from zero from ANOVA ($p<0.05$)

[E] NE (Not Estimated) indicates the SNP locus did not meet criteria of 3 genotypic classes and/or departed from Hardy Weinberg Equilibrium

Table 5. Genotypic frequencies by SNP and phenotypic class (+/- 1.5 std dev) with χ^2 tests for significant differences from the population; XX is the major allele homozygote, XY the heterozygote, and YY is the minor allele homozygote.

Trait	Locus	Class	XX	XY	YY	#obs	$p>\chi^2$
$\Delta^{13}C$	0_17030_01_94	Population	311	33	1	345	
		High	20	0	1	21	0.03
		Low	24	1	0	25	0.61
	0_10921_01_353	Population	308	36	2	346	
		High	18	2	1	21	0.04
		Low	20	5	0	25	0.27
	CL599Contig1_07_109	Population	274	64	6	344	
		High	14	4	3	21	<0.001
		Low	20	5	0	25	0.79
	0_17543_01_196	Population	287	57	8	352	
		High	17	3	1	21	0.73
		Low	21	4	0	25	0.74
	0_3648_01_357	Population	285	54	5	344	
		High	15	5	1	21	0.24
		Low	18	7	0	25	0.21
	0_8304_02_414	Population	276	66	4	346	
		High	15	4	2	21	0.0016
		Low	22	3	0	25	0.56
Height	0_14415_01_190	Population	152	140	46	338	
		High	12	5	1	18	0.17
		Low	8	21	2	31	0.01
%N	UMN_6338_01_99	Population	313	32	1	346	
		High	15	3	0	18	0.54
		Low	17	1	1	19	0.84
	0_17195_01_417	Population	318	25	3	346	
		High	18	0	0	18	0.45
		Low	15	3	1	19	0.03
	2_1087_01_86	Population	311	32	2	345	
		High	18	0	0	18	0.37
		Low	16	2	1	19	0.026
	2_4191_01_104	Population	297	46	3	346	
		High	15	3	0	18	0.85
		Low	16	1	2	19	<0.001
	2_7865_01_156	Population	309	36	1	346	
		High	18	0	0	18	0.34
		Low	17	1	1	19	0.0002
	CL1074Contig1_03_101	Population	328	17	1	346	
		High	12	5	1	18	<0.001
		Low	19	0	0	19	0.59

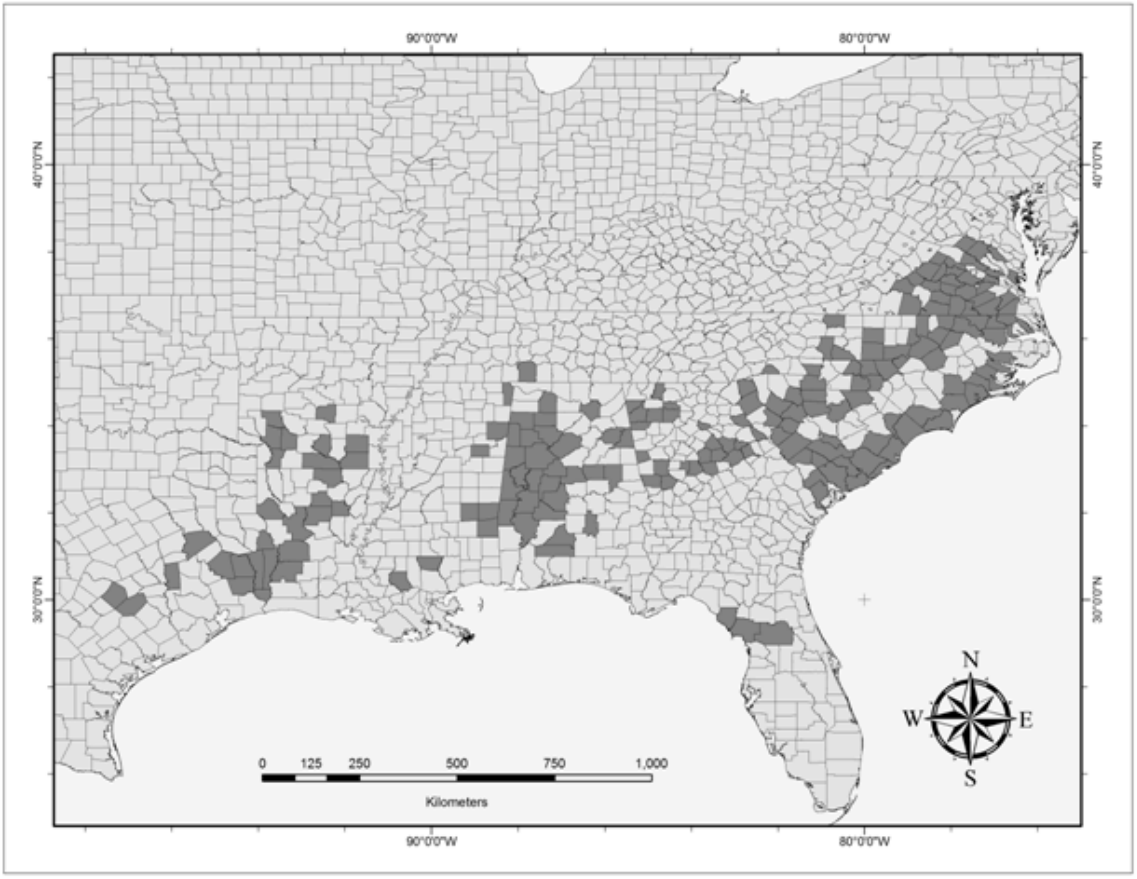

Figure 1. Counties of origin for genotypes selected for the NCSU association population. Shaded counties represent a least one genotype.

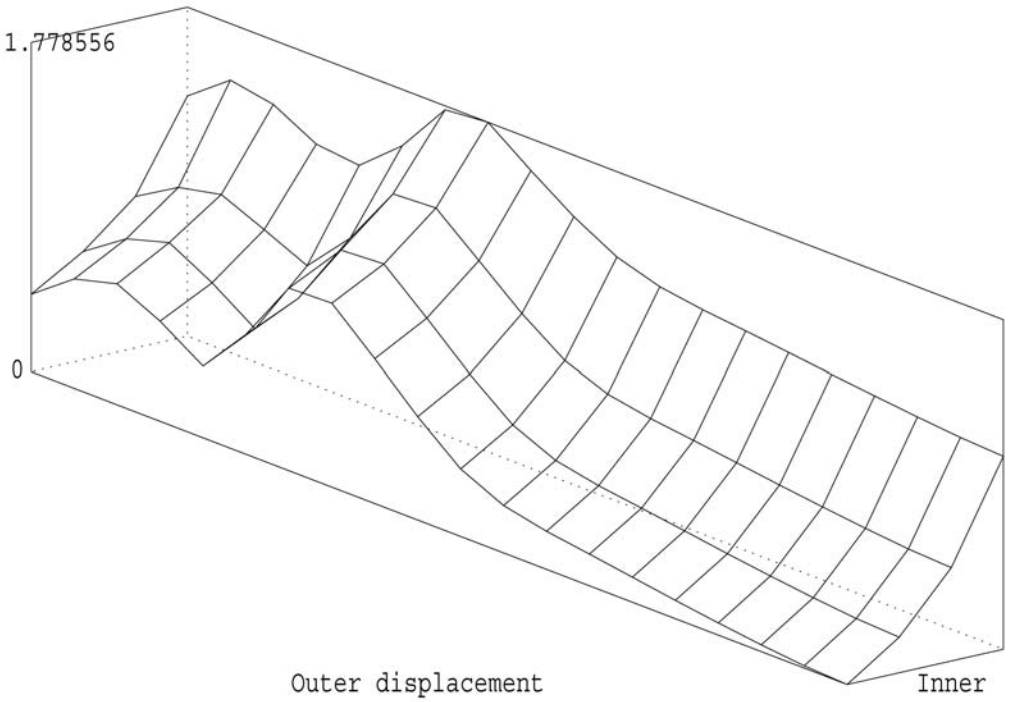

Figure 2. Variogram plot of residuals for $\Delta^{13}C$. The three dimensional plot reflects the similarity of trees due to environmental trends in a row and column spatial configuration.

Supplemental Information

Table S1. Mean of the Squared Difference (MSD) and the number of associated SNP loci by model, p-value and q-value. GLM =model with no covariates; Q=model with population structure, K= model with marker-based kinship; QK =model with both population structure and marker-based kinship.

$\Delta^{13}C$

Model	MSD	p<0.2	p<0.1	p<0.05	p<0.01	q<0.05
GLM	0.00054	414	244	193	65	4
Q	0.00009	439	202	173	55	4
K	0.16977	374	178	144	47	4
QK	0.00063	364	186	145	44	7

Height

Model	MSD	p<0.2	p<0.1	p<0.05	p<0.01	q<0.05
GLM	0.00054	443	221	172	46	1
Q	0.00010	415	221	152	32	0
K	0.00029	391	201	133	31	0
QK	0.00003	387	199	133	31	0

%N

Model	MSD	p<0.2	p<0.1	p<0.05	p<0.01	q<0.05
GLM	0.00013	395	200	176	60	6
Q	0.00013	376	206	155	56	5
K	0.00099	357	217	149	53	6
QK	0.00112	345	200	135	46	5

Table S2. Significant associations for $\Delta^{13}C$, height (Ht), and %N using a threshold of $p<0.01$ uncorrected for multiple testing.

Model	Trait	SNP	p-value	q-value
GLM	%N	UMN-6338-01-99	6.6E-08	0.0002
GLM	%N	2-7865-01-156	1.1E-07	0.0002
GLM	%N	2-1087-01-86	8.2E-07	0.0011
GLM	%N	CL1074Contig1-03-101	2.8E-06	0.0028
GLM	%N	0-17195-01-417	5.8E-06	0.0046
GLM	%N	2-4191-01-104	4.1E-05	0.0271
GLM	%N	0-18356-02-396	1.7E-04	0.0973
GLM	%N	0-9347-01-328	2.8E-04	0.1354
GLM	%N	0-886-02-324	4.4E-04	0.1912
GLM	%N	0-16924-03-96	5.2E-04	0.2040
GLM	%N	UMN-3841-01-114	0.001	0.3475
GLM	%N	2-1447-02-71	0.001	0.3475
GLM	%N	CL4733Contig1-01-61	0.001	0.3475
GLM	%N	0-16610-01-506	0.001	0.3475
GLM	%N	2-8909-02-75	0.001	0.3475
GLM	%N	2-9190-01-446	0.002	0.3475
GLM	%N	0-8271-02-56	0.002	0.3475
GLM	%N	0-7639-01-77	0.002	0.3523
GLM	%N	2-85-02-400	0.002	0.3523
GLM	%N	2-3211-01-505	0.002	0.4135
GLM	%N	0-14873-01-155	0.002	0.4313
GLM	%N	0-4750-01-582	0.003	0.4475
GLM	%N	CL1085Contig1-01-629	0.003	0.4573
GLM	%N	0-13816-01-36	0.003	0.4573
GLM	%N	0-1802-01-424	0.003	0.4573
GLM	%N	0-18755-01-72	0.003	0.4573
GLM	%N	0-1549-01-523	0.003	0.4573
GLM	%N	0-3215-01-104	0.004	0.4573
GLM	%N	2-2702-01-69	0.004	0.4573
GLM	%N	UMN-2378-01-410	0.004	0.4573
GLM	%N	UMN-2675-02-45	0.004	0.4573
GLM	%N	0-9729-01-92	0.004	0.4594
GLM	%N	0-2223-01-106	0.004	0.4594
GLM	%N	0-8408-01-328	0.004	0.4594
GLM	%N	CL1905Contig1-03-377	0.004	0.4594
GLM	%N	CL1905Contig1-06-351	0.004	0.4594
GLM	%N	2-3128-01-306	0.005	0.4776
GLM	%N	0-13391-02-213	0.005	0.4776
GLM	%N	0-3604-02-88	0.005	0.4776
GLM	%N	CL1530Contig1-04-89	0.005	0.4776
GLM	%N	0-15548-01-24	0.005	0.4776
GLM	%N	0-4762-01-133	0.005	0.4776
GLM	%N	0-7682-01-200	0.005	0.4776

Table S2 Continued

GLM	%N	0-3261-01-545	0.005	0.4776
GLM	%N	0-13066-01-114	0.006	0.4776
GLM	%N	CL4736Contig1-01-135	0.006	0.4776
GLM	%N	UMN-4815-01-95	0.006	0.4776
GLM	%N	2-6434-01-87	0.006	0.4840
GLM	%N	0-15010-01-259	0.006	0.5144
GLM	%N	0-16872-01-283	0.007	0.5671
GLM	%N	0-5973-02-148	0.008	0.5868
GLM	%N	CL1591Contig1-04-165	0.009	0.6541
GLM	%N	0-16864-01-251	0.009	0.6541
GLM	%N	2-2196-02-297	0.009	0.6541
GLM	%N	UMN-2789-01-471	0.009	0.6541
GLM	%N	UMN-1835-02-116	0.010	0.6541
GLM	%N	2-5483-02-355	0.010	0.6541
GLM	%N	0-14099-02-60	0.010	0.6541
GLM	%N	0-9488-01-346	0.010	0.6541
GLM	Ht	0-14415-01-190	1.2E-05	0.0425
GLM	Ht	0-10207-01-280	2.4E-04	0.4431
GLM	Ht	0-13026-02-330	4.0E-04	0.4818
GLM	Ht	CL1767Contig1-02-318	6.2E-04	0.5611
GLM	Ht	0-16096-01-71	9.2E-04	0.6690
GLM	Ht	UMN-7021-02-156	0.001	0.6810
GLM	Ht	CL3739Contig1-06-209	0.001	0.6810
GLM	Ht	CL558Contig1-07-47	0.002	0.6810
GLM	Ht	0-13348-01-522	0.002	0.7668
GLM	Ht	0-1688-02-561	0.002	0.7673
GLM	Ht	0-5024-01-185	0.004	0.7673
GLM	Ht	2-3177-03-168	0.004	0.7673
GLM	Ht	CL2416Contig1-06-360	0.004	0.7673
GLM	Ht	0-16068-01-350	0.004	0.7673
GLM	Ht	0-17701-01-115	0.005	0.7673
GLM	Ht	CL1791Contig1-03-50	0.005	0.7673
GLM	Ht	UMN-4764-02-149	0.005	0.7673
GLM	Ht	0-3900-01-431	0.005	0.7673
GLM	Ht	0-9340-01-203	0.005	0.7673
GLM	Ht	0-1666-01-844	0.005	0.7673
GLM	Ht	0-2214-02-59	0.006	0.7673
GLM	Ht	UMN-4856-01-462	0.007	0.7673
GLM	Ht	0-4910-01-103	0.007	0.7673
GLM	Ht	2-3150-02-90	0.007	0.7673
GLM	Ht	0-7001-01-143	0.007	0.7673
GLM	Ht	CL2342Contig1-01-407	0.007	0.7673
GLM	Ht	0-16213-01-75	0.007	0.7673
GLM	Ht	0-2398-01-131	0.008	0.7673
GLM	Ht	2-4484-02-238	0.008	0.7673
GLM	Ht	2-2866-01-512	0.008	0.7673
GLM	Ht	2-2199-01-139	0.008	0.7673

Table S2 Continued

GLM	Ht	UMN-3386-01-277	0.009	0.7673
GLM	Ht	0-13455-01-122	0.009	0.7673
GLM	Ht	0-14694-01-399	0.009	0.7673
GLM	Ht	CL2055Contig1-03-78	0.009	0.7673
GLM	Ht	0-1433-02-78	0.009	0.7673
GLM	Ht	0-2214-02-367	0.009	0.7673
GLM	Ht	0-14885-02-61	0.009	0.7673
GLM	Ht	CL3094Contig1-03-125	0.009	0.7673
GLM	Ht	0-15958-01-145	0.010	0.7673
GLM	Ht	2-4102-01-738	0.010	0.7673
GLM	Ht	2-776-03-84	0.010	0.7673
GLM	Ht	2-6804-01-289	0.010	0.7673
GLM	Ht	2-5989-02-136	0.010	0.7673
GLM	Ht	UMN-7253-01-407	0.010	0.7673
QK	$\Delta^{13}C$	0-17030-01-94	4.9E-09	0.00002
QK	$\Delta^{13}C$	0-17543-01-196	1.5E-06	0.0030
QK	$\Delta^{13}C$	0-10921-01-353	6.0E-06	0.0079
QK	$\Delta^{13}C$	2-1501-01-109	2.0E-05	0.0197
QK	$\Delta^{13}C$	CL599Contig1-07-109	2.8E-05	0.0224
QK	$\Delta^{13}C$	0-3648-01-357	5.5E-05	0.0363
QK	$\Delta^{13}C$	0-8304-02-414	7.2E-05	0.0404
QK	$\Delta^{13}C$	0-1191-01-125	8.0E-04	0.3635
QK	$\Delta^{13}C$	0-13127-01-168	9.9E-04	0.3635
QK	$\Delta^{13}C$	CL1879Contig1-02-192	0.001	0.3635
QK	$\Delta^{13}C$	0-15732-01-588	0.001	0.3635
QK	$\Delta^{13}C$	2-1321-01-285	0.001	0.3635
QK	$\Delta^{13}C$	UMN-991-01-295	0.001	0.3635
QK	$\Delta^{13}C$	2-83-01-503	0.002	0.5063
QK	$\Delta^{13}C$	0-8671-01-61	0.002	0.5251
QK	$\Delta^{13}C$	2-6538-01-49	0.002	0.5907
QK	$\Delta^{13}C$	2-987-02-146	0.003	0.5907
QK	$\Delta^{13}C$	2-6538-01-334	0.003	0.5907
QK	$\Delta^{13}C$	CL910Contig1-03-196	0.003	0.5907
QK	$\Delta^{13}C$	UMN-2867-01-361	0.003	0.5907
QK	$\Delta^{13}C$	CL4422Contig1-03-167	0.003	0.6376
QK	$\Delta^{13}C$	2-2325-02-138	0.004	0.6444
QK	$\Delta^{13}C$	CL1389Contig1-01-203	0.005	0.7757
QK	$\Delta^{13}C$	0-976-01-67	0.005	0.7757
QK	$\Delta^{13}C$	2-4095-01-278	0.005	0.7757
QK	$\Delta^{13}C$	CL4628Contig1-01-38	0.005	0.7757
QK	$\Delta^{13}C$	2-8658-01-44	0.006	0.7757
QK	$\Delta^{13}C$	0-18604-01-94	0.006	0.7757
QK	$\Delta^{13}C$	0-8975-02-113	0.006	0.7757
QK	$\Delta^{13}C$	UMN-5129-03-244	0.006	0.7757
QK	$\Delta^{13}C$	CL4028Contig1-03-105	0.006	0.7757

Table S2 Continued

QK	$\Delta^{13}C$	UMN-CL34Contig1-03-89	0.006	0.7757
QK	$\Delta^{13}C$	CL1074Contig1-03-101	0.007	0.7757
QK	$\Delta^{13}C$	0-3215-01-104	0.007	0.7760
QK	$\Delta^{13}C$	CL4662Contig1-01-215	0.007	0.7989
QK	$\Delta^{13}C$	CL3330Contig1-01-106	0.007	0.8083
QK	$\Delta^{13}C$	0-6409-02-252	0.008	0.8083
QK	$\Delta^{13}C$	2-9790-01-129	0.008	0.8083
QK	$\Delta^{13}C$	0-886-02-324	0.009	0.8592
QK	$\Delta^{13}C$	CL4663Contig1-02-55	0.009	0.8592
QK	$\Delta^{13}C$	0-7948-01-115	0.009	0.8592
QK	$\Delta^{13}C$	2-3041-01-77	0.009	0.8592
QK	$\Delta^{13}C$	2-5746-02-630	0.010	0.8592
QK	$\Delta^{13}C$	0-10116-01-165	0.010	0.8592

CHAPTER 2

Genetic Improvement of Sawtimber Potential in Loblolly Pine

(In the format appropriate for submission to Forest Science)

Genetic Improvement of Sawtimber Potential in Loblolly Pine

Patrick Cumbie[1], Fikret Isik, Steve McKeand

Cooperative Tree Improvement Program,

North Carolina State University

Campus Box 8002

Raleigh, NC 27695, USA

[1]Corresponding author (email: Patrick_Cumbie@ncsu.edu).

Abstract: Progeny from 48 elite parents of loblolly pine (*Pinus taeda* L.) were bred in a disconnected diallel mating design and were planted at four sites across the lower coastal plain of the southeastern United States. Height, DBH, volume, fusiform rust incidence, stem forking, stem sweep, branch angle, branch diameter, branch frequency, and a sawtimber potential score were measured after six growing seasons. There were significant differences among families for all traits measured. Individual-tree narrow-sense heritability estimates ranged from 0.06 to 0.22, and half-sib family-mean heritability estimates ranged from 0.73 to 0.98. Height and volume were highly correlated with the sawtimber potential score of individual trees. From multiple regression, 79% of the variation in sawtimber potential breeding values can be attributed to variation in volume, rust incidence, stem sweep, and forking breeding values. The potential dollar value of loblolly pine was increased as much as 162% over local checks when both volume and sawtimber potential were used to select the 10 best parents from the population. Implementation of a selection index on currently measured traits is a promising opportunity to make gains in the proportion of sawtimber produced from improved germplasm of loblolly pine in the southeastern United States.

Introduction

Tree improvement programs have focused on increasing the value of forest plantations in the southeastern United States for over fifty years. After two cycles of breeding, the North Carolina State University Cooperative Tree Improvement Program (hereafter called 'the NCSU Cooperative') has made significant contributions to increases in growth rate, development of resistance to fusiform rust (caused by the fungus *Cronartium quercuum* (Berk.) Miyabe ex Shirai f. sp. *fusiforme*), and improvements in stem straightness in loblolly pine (*Pinus taeda* L.) (Li et al. 2000). Silvicultural improvements, including tree improvement, could increase productivity of loblolly pine plantations to exceed 10 tons per acre per year (Allen et al. 2005) . As intensive timber production is focused on fewer acres of highly managed land, increasing the value, not only the volume, of timber harvested must be a high priority for silvicultural improvements. Increases in growth rate through better silviculture have clearly had an impact on plantation value, but improvements in timber quality can also be made. Previous research suggested that crown characteristics have an impact on lumber quality such as veneer grade bolts, but the major economic factor of interest to date has been timber volume (Busby 1983).

Stem quality in loblolly pine plantations has rarely been quantified. Amateis and Burkhart (2005) analyzed a long-term thinning study for the distribution of peeler, sawtimber, and pulpwood stems. Their study showed that thinning significantly influenced timber product distribution. Choi et al. (2008) recently modeled the effect of stem defects on loblolly pine stem quality in non-thinned plantations. Through age 15 years, there was a relatively low

incidence of forked stems (4%), broken tops (5%) and diseased trees (12%) while stem sweep was a more common defect (41%). Over the 15-year life of the stand, 37% of the forked stems recovered to single stems and 83% of the trees with broken tops recovered to a normal top (Choi et al. 2008).

In genetically improved loblolly pine seedlings, improvement of growth rate alone will result in substantial financial gains (McKeand et al. 2006). Improvements in stem form and rust have been recognized (Lambeth 2000; Li et al. 2000) and positive impacts on sawtimber quality have been implied. However, heritable variation of sawtimber traits that could potentially be incorporated into southeastern United States loblolly pine breeding programs has not been quantified. An increased importance of sawtimber production could justify additional emphasis on selecting and deploying superior genotypes from breeding populations for improved stem and branching characteristics.

There are few reports describing genetic analyses of composite crown and stem form traits in the literature. Busby (1983) found that branching traits significantly affected the quality of veneer bolts in loblolly pine but did not affect dollar value in comparison to overall tree volume. The composite crown scores of the first generation of improvement in the NCSU Cooperative were poorly inherited most likely due to multiple subjective traits scored by many graders (Busby 1983).

Attempts have been made to assess crown and branch traits of other species for incorporation into breeding programs with varying results. Adams and Morganstern (1991)

found negative correlations between height and crown traits in Jack Pine (*Pinus banksiana* Lamb.), making improvements in both growth and crown form difficult. In radiata pine (*Pinus radiata* D.Don), growth was most important for increasing value in plantations, while stiffness (Modulus of Elasticity or MoE) was more critical for added value in sawmills (Ivkovic et al. 2006). Composite stem form assessments exhibited low heritability in radiata pine but were stable across sites (Jayawickrama 2001). Jayawickrama (2001) reported that the within-site narrow-sense heritability for percent-acceptable stems ranged from 0.02 to 0.15 and was less heritable than stem straightness. The percent-acceptable trait was highly correlated with stem straightness (0.97) and moderately correlated (0.68) with a malformation score (forking). Jayawickrama and Low (1999) reported that regional sources of radiata pine differed significantly in stem acceptability, malformation, and internode length using a categorical scoring system.

Understanding the impact of genetic variation in economically important form traits is necessary for timber value improvement in loblolly pine. The objectives of this experiment were to (*i*) quantify genetic variation of growth and form traits affecting sawtimber yield in loblolly pine; (*ii*) estimate genetic correlations among growth, form, and a composite sawtimber trait; (*iii*) assess the impact of growth, stem form, and branch traits on sawtimber quality; and (*iv*) determine if index selection is a viable option to make future improvements in sawtimber quality traits in loblolly pine.

Materials and Methods

Plant material

Forty-eight first- and second-generation elite parents from the NCSU Cooperative, the University of Florida Cooperative Forest Genetic Research Program, and the Western Gulf Forest Tree Improvement Program at the Texas Forest Service were selected based on previous progeny test results. Parents were selected from three provenances: Atlantic Coastal Plain (ACP), Western Gulf (WG), and Florida (FL). Elite parents were mated using 8-tree disconnected half-diallels. Each diallel contained 4 ACP selections, 3 FL selections, and 1 WG selection. Breeding took place in the late 1990s, and seedlings were planted in the field in the winter of 2001. Four test locations in Georgia (Brunswick and Bainbridge), Alabama (Phenix City), and Florida (Yulee) were measured for this experiment (Table 1). The experimental design was a randomized complete block with single tree plots randomized in 20 blocks at each of the four sites.

At the end of the sixth growing season, total tree height, stem diameter at breast height (dbh), fusiform rust incidence, forking incidence, stem sweep, branch angle, branch diameter, branch frequency, and sawtimber potential were measured. Volume was estimated according to Goebel and Warner (1966). Rust incidence and forking incidence were scored as binary traits (0, 1). Stem sweep, branching traits and sawtimber potential were measured using a categorical scale for each trait. The following procedures were used to assess the form, branching and sawtimber potential traits:

(1) Stem sweep (hereafter termed "sweep") was measured in 1.27 cm increments at the point of maximum deflection in the first 3.5m of the bole with a 2.5m straight edge.

(2) Branch angle was measured at the base of the live crown and was scored on a 1 to 6 category of 15 degree increments from flat (0 degrees) to vertical (90 degrees).

(3) Branch diameter measured in the first 1.5 logs (7m) was scored using 1.27 cm increments that were rounded up to the nearest 1.27 cm.

(4) Branch frequency (inter-node length) was scored in 15 cm increments. This was a general characterization of the tree. The categorical score indicates the average spacing between whorls on the main stem of the tree by visual assessment.

(5) Sawtimber potential was scored using a 1 (best) to 4 (worst) scale where:

(1) = Sawtimber: no quality defects in the first 1.5 logs (7m) of the tree.

(2) = Some Sawtimber: minor defects in the first 1.5 logs but still likely to be a sawtimber tree. Minor defects included minor sweep (<7.5cm), high forks (above first log or 4.9m), and small ramicorn branches.

(3) = Pulpwood: defects present such as low forks, major ramicorn branches, large branches in the crown > 6cm, branch angles over 45 degrees, below average volume, sweep >7.5cm.

(4) = non-merchantable: Multiple major stem defects such as stem rust, forking, extremely poor growth, or poor branching characteristics.

Statistical analyses

The growth (height and volume) and form traits (branching characteristics and sawtimber) were analyzed using linear mixed models with the general form:

[1] $$\mathbf{y = Xb + Zu + e}$$

where **y** is the vector of response variable (observations); the **X** is the design matrix relating individual observations to the fixed effects in the model, **b** is the vector of fixed effect factors that includes the overall mean, site and replication within site effects; **Z** is the incidence matrix relating observations to random effects and **u** is the vector of random effects that includes general combining ability (GCA) effects, specific combining ability (SCA) effects and their interactions with site; **e** is the vector of random residual terms. The expectation of fixed terms is E(**y**) = **Xb**. The random terms are assumed to have zero means and variances var(**u**)=**G**, var(**e**)=**R**. The variance covariance matrix of observations (vector y) is Var(**y**) = $\mathbf{ZGZ}^T$ + **R,** where **G** and **R** are covariance matrices corresponding to **u** and **e**, respectively. The **G** is a block diagonal matrix defined as $\mathbf{A}\sigma^2_G$ for GCA effects, $\mathbf{I_n}\sigma^2_S$ for SCA effects, $\mathbf{I_n}\sigma^2_{GT}$ for GCA by site interactions and $\mathbf{I_n}\sigma^2_{ST}$ for SCA by site interactions, where A is the numerator relationship matrix and I is an identity matrix of dimension n_i x n_i (n_i = number of levels of the ith term). The $\mathbf{R}=\mathbf{I}\sigma^2_e$ is a diagonal matrix with the residual error variances (σ^2_e) in the diagonal and 0 covariances (null submatrix) in the off diagonals; **I** is the identity matrix of dimension equal to the number of observations. Bivariate analyses

were used to estimate genetic covariances among traits, where **G** and **R** include variances and covariances of the two traits.

For binary traits (fork and rust disease) the following generalized linear mixed model was fit to data to obtain variance components.

[2] $$g^{-1}(\boldsymbol{\eta}) = g^{-1}(\mathbf{Xb} + \mathbf{Zu})$$

where $\boldsymbol{\eta}$ is the linear predictor ($\boldsymbol{\eta}$ = logit($\boldsymbol{\mu}$)), g is the link function, $\boldsymbol{\mu}$ is the conditional mean. The remaining terms are as previously explained. The covariance matrix of $\boldsymbol{\eta}$ is $\text{Cov}(\boldsymbol{\eta}) = \mathbf{ZGZ}^T + \mathbf{R}$ where, $\mathbf{G}=\sigma^2_i\mathbf{I}$ is the diagonal matrix of random effects **u,** and $\mathbf{R}=\sigma^2_e\mathbf{I}$ is the diagonal random errors variance (Schall 1991). The variance of a binary trait is mean dependent as $\mu/(1-\mu)$. Statistical analysis of data with a mean incidence outside certain boundaries (e.g., 0.30 to 0.70) is not advised (Gilmour et al. 1985, 1987), but studies have shown that low heritability traits (<0.3) are not severely biased when trait incidences are below 0.25 (Lopes et al. 2000; Mantysaari et al. 1991) . An analysis of 123 field trials revealed that forking is a low heritability trait in loblolly pine (Xiong et al. 2009), and in our data the overall incidence was within such boundaries for rust incidence (0.31) but was slightly lower for forking (0.18) suggesting that estimates from the data may not be strongly biased.

Significance of random terms in the models (variance components) were tested using the log likelihood ratio test. Heritability estimates were obtained for all traits across sites using the variance components generated. The individual-tree narrow-sense (h^2_i), half-sib family-

mean narrow-sense (h^2_{hs}), and full-sib family-mean broad-sense (H^2_{fs}) heritabilities for traits across sites were estimated by the following equations:

[3] $$h_i^2 = \frac{4\sigma_G^2}{2\sigma_G^2 + \sigma_S^2 + 2\sigma_{GT}^2 + \sigma_{ST}^2 + \sigma_E^2}$$

[4] $$h_{hs}^2 = \frac{\sigma_G^2}{\left[p\sigma_G^2 + \sigma_S^2 + \frac{p\sigma_{GT}^2}{t} + \frac{\sigma_{ST}^2}{t} + \frac{\sigma_E^2}{tbn} \right] \frac{1}{p-1}}$$

[5] $$H^2_{fs} = \frac{2\sigma_G^2 + \sigma_S^2}{2\sigma_G^2 + \sigma_S^2 + \frac{2\sigma_{GT}^2}{t} + \frac{\sigma_{ST}^2}{t} + \frac{\sigma_E^2}{tbn}}$$

where σ^2_G is the GCA variance; σ^2_S is the SCA variance; σ^2_{GT} and σ^2_{ST} are the GCA by site and SCA by site interactions, respectively; σ^2_E is the residual variance; t is the number of sites (t=4); b is the number of replications within sites (b=20); p is the number of parents (p = 48); and n is the number of trees per plot (n = 1). When the heritability of binary traits (forking and rust disease) were calculated, the residual variance was set to $\pi^2/3=3.29$ because a binary trait has a residual variance of 3.29 on the underlying scale (Gilmour et al. 1985, 1987). Standard errors of heritabilities were obtained using the Taylor series expansion. All the genetic data analyses were carried out using ASReml software (Gilmour et al. 2006). Best linear unbiased predictions (BLUP) of parental breeding values were obtained as two times the general combining ability (BV=2GCA). For binary traits, the

predictions from the mixed model solutions (logit scale) were back transformed to the probability scale as suggested by Isik et al.(2008).

Genetic correlations (r_g) were calculated between traits as follows;

[7] $$r_{g(xy)}(xy) = \frac{\sigma_{xy}}{\sqrt{\sigma^2_{g(x)}\sigma^2_{g(y)}}}$$

where $r_{g(xy)}$ is the genetic correlation between traits x and y; σ_{xy} is the GCA covariance between traits; $\sigma^2_{g(x)}$ is the GCA variance for trait x; and $\sigma^2_{g(y)}$ is the GCA variance for trait y. In cases of correlations among two binary traits, the correlation of parental breeding values was used. We used the Type B genetic correlation to explore the extent of genotype by environment (GxE) interaction for GCA effects across test sites (Burdon 1977; Yamada 1962) using

[8] $$r_{g_{B\_GCA}} = \frac{\sigma^2_G}{\sigma^2_G + \sigma^2_{GT}}$$

where $r_{g\ B\_GCA}$ is the Type B genetic correlation and variance components are as defined previously.

Using approximate tonnage values for the southeastern United States at the time of data collection (Timber Mart-South 2007), sawtimber potential scores and volume were used to

estimate the individual tree value at 6 years of age. The following factors were used to estimate economic value: Sawtimber potential score of 1 = large sawtimber value of $30/ton; 2 = small sawtimber value of $15/ton; 3 = pulpwood value of $5/ton; and 4 = non-merchantable value of $0/ton. Volume was converted to weight with a simple conversion of 1 US ton (0.907 tonnes) = 100 cubic feet (2.83 cubic meters). For each tree, the entire volume was assumed to have a value based on the sawtimber potential score assigned at the time of grading at age six. We did not merchandize individual stems into different product classes.

Multiple linear regression analysis was used to determine the importance of traits on the composite sawtimber potential score. Parental breeding values for traits were treated as explanatory variables in the expression

[8] $$\hat{Y} = \beta_0 + \beta_1 X_1 + ... + \beta_i X_i + \varepsilon$$

where $\hat{Y}$ is the predicted value of sawtimber potential; β_0 is the intercept parameter; β_1 is the slope parameter for trait 1; X_1 is the first independent variable; β_i is the slope parameter for the i-th trait; X_i is the i-th independent variable; ε is the random error with independently and identically distributed random error variance. The ratio of regression coefficients was then used as economic weights in the selection index. A step-wise model selection method in the GLM procedure of SAS software was used to obtain the final model (SAS 1989).

The variance component estimates were used to generate a selection index to estimate potential population improvement of both growth and sawtimber quality. Index coefficients (**b**) were estimated for each trait using the following formula:

[9] $$b = P^{-1}Ca$$

where **b** is the optimal weights vector based upon the economic value vector **a**, the phenotypic variance-covariance matrix **P**, and the additive genetic variance-covariance matrix **C** (Lin 1978). Genetic gain (**ΔG**) from the index weights was then estimated as follows:

[10] $$\Delta G = i\sqrt{b' Pb}$$

where *i* is the selection intensity, **b** is the optimal weights from the index, and **P** is the phenotypic variance-covariance matrix. Selection intensity *i* for this experiment was 1.372 based upon a proportion of selection of 20% of parents (10 out of 48 parents). The vector of economic weights, **a**, is typically the economic value of improving a trait by one unit. However, the economic value of unit increases is not well known and has not been defined for use in loblolly pine selection indices. Three indices were compared: 1) weight volume as the only economically important trait; 2) weight traits according to the coefficients from multiple regression analysis; and 3) weight sawtimber potential only.

Results

Site differences

The effect of test location was significantly different for all growth traits at the $p<0.001$ level (Table 1.) Trees at the Bainbridge and Brunswick test sites were similar for volume, height, and dbh, while trees at the Yulee and Phenix City test sites were slower growing. Rust incidence was lower at Yulee and Brunswick (23% and 31%, respectively), while Bainbridge and Phenix City both had higher rust infection (40% and 56%, respectively). Sweep was similar across all tests ranging from 1.6 in Phenix City to 2.2 in Brunswick. Branch angle site differences were small but the effect was significant ($p<0.03$). Site effects for branch diameter and branch frequency were significant ($p<0.001$) but differences were small with the exception of branch diameter at the Yulee site which had a lower value than any other test.

Sawtimber potential was significantly different among sites ($p<0.001$), and differences ranged from 2.5 at the Bainbridge and Brunswick sites to 2.8 at the Phenix City site. The mean score of sawtimber potential at the Bainbridge and Brunswick tests indicated that 50% of the trees were scored as sawtimber while the Phenix City test had significantly more trees with scores of 3 (pulpwood) or 4 (cull).

Variance components and heritability estimates

GCA variance was significant for all traits, ranging from 6% of the total variance in rust incidence to 1% of the total variance for sawtimber potential (Table 2). SCA variance was significant for all traits except branch angle and forking, and was highest for height and dbh. The GCA by site interactions were generally significant for the traits, but that was not the case for SCA by site interactions. As expected, the residual variance explained a high proportion of total phenotypic variance for traits, ranging from 90% (dbh) to 96% (branch diameter). Type B genetic correlations were generally high, ranging from 0.72 for sawtimber potential to 1.0 for branch frequency which had no GCA by test variance (Table 2). Type B correlations for height and rust incidence were also very high (0.92 and 0.96, respectively).

Individual-tree heritability ranged from 0.06 for sawtimber potential up to 0.22 for rust incidence and height (Table 2). Growth traits were in the range of expected values, and branching traits were slightly lower than growth traits. Sweep was more similar in individual-tree heritability to growth traits, and forking and rust incidences were higher than growth traits. Half-sib family-mean heritability ranged from a low of 0.73 for branch angle to a high of 0.98 for branch frequency. Full-sib family-mean heritability ranged from a low of 0.72 for branch angle to a high of 0.91 for height.

Genetic correlations

Sawtimber potential was highly genetically correlated with tree height (-0.77) (Table 3). Since a lower number was superior in the sawtimber scoring, the negative correlation with height is favorable. Sawtimber potential was also favorably correlated with dbh, Volume, Sweep and rust incidence. Although forking and branch traits were components of the subjective sawtimber potential trait, relationships with sawtimber were not significant. Sweep was uncorrelated with height, dbh and volume, suggesting that stem straightness and growth can be improved simultaneously. Branch diameter was moderately correlated with height, dbh, and volume (0.45, 0.63, and 0.56 respectively) suggesting larger trees tend to have had larger branches. Branch angle was weakly negatively correlated with dbh, volume, and sweep (-0.27, -0.28, -0.27 respectively). Branch frequency was moderately positively correlated with height (0.56) indicating a relationship of longer spaces between whorls on taller trees. Forking incidence was positively correlated (unfavorable) with growth traits and rust incidence was independent of growth traits.

Assessment of value for stem quality and growth

The favorable correlation between growth and sawtimber potential is visible when the proportion of sawtimber (generated from sawtimber potential scoring) and volume genetic values are plotted (Figure 1). Values of the proportion of sawtimber ranged from 0.25 to 0.76 for the 48 half-sib family means. Using a simple conversion from wood prices in the southeastern United States, an estimate of individual tree value was generated using the

sawtimber potential and volume traits (Table 4). Individual tree value was low, which was to be expected at 6 years of age, as rotation ages generally exceed 20 years.

The top ten parents from the population for value (incorporating product class and volume) versus the top ten parents for volume alone were selected and compared to the local unimproved checklot in the tests. Selecting either set of parents adds substantial value and volume over the checklot, but selection of volume alone does not produce the most value (Table 4). Selecting the top ten parents for value resulted in 12% less volume gain but an increase of 24% in value over the local check as compared to the top ten crosses for volume alone. Average tree value for the highest value parents was $0.55 per tree, while the highest volume parents were $0.50 per tree. Both were more than twice the value of the unimproved check. A ranking of the 15 best parents based on volume gain (% over local check) shows that if sawtimber quality is desired, some of the faster growing parents may be less desirable, while some parents with a slightly lower volume gain have larger proportion of sawtimber grade trees at age 6 (Figure 2).

Multiple regression of growth and form traits on sawtimber potential

A multiple regression model using the growth and form trait breeding values of parents explained 79% of the variation in sawtimber potential breeding values (Table 5). The final regression model included volume, rust, sweep, and forking, which had significant effects on sawtimber value. Based upon the regression coefficients, volume had the greatest positive effects on sawtimber potential while rust, sweep, and forking had negative

coefficients, reflecting the penalty a tree receives for these defects. The ratio of volume to rust, sweep, and forking traits was rounded to 3:1. This ratio was used in the selection index as one possible set of economic weights to generate index values.

Trait improvement through a selection index

If volume alone is the only selection criterion in this population (Index 1), sawtimber potential would improve by 5% but rust incidence, forking, and sweep would all shift in the unfavorable direction (Table 6). By adding weights for rust incidence, forking, and sweep, gains in volume and sawtimber potential would not be penalized, and rust incidence would be reduced by 0.01. Sweep and forking would increase, but with slightly less magnitude compared to Index 1 which weighted volume alone. Index 3 demonstrates that if the composite sawtimber potential trait were the only selection criterion, there would be greater improvements in sweep (-11%), rust incidence (-0.27) and sawtimber potential (-7%). Forking would increase but substantially less than in Indices 1 and 2. Volume gain would be less compared to the other indices in this index, but still a significant improvement of 12% over the population mean.

Discussion

Genetic variation

GCA variance was significant for all traits, but SCA was higher than expected compared to previous estimates in loblolly pine (Isik et al. 2005). This may be an effect of the small

diallel size or the crosses among provenances in this population. It is promising that the GCA variance, while small, was significant for sawtimber potential suggesting that improvements could be made through capturing additive genetic variation in breeding programs.

While parents in this study were tested on four different sites covering a large geographic range, GCA by site interactions accounted for a small amount of the total variance. Results from this analysis suggest that ranking parents for sawtimber potential should be stable across sites. Interactions could be due to differences in forking incidence, rust infection, wind damage, or other non-quantified differences among the sites. Sierra-Lucero et al. (2002) reported moderate GxE for Florida source loblolly pine in a provenance trial of the southeastern United States coastal plain, in which Florida families accounted for 60% of the GxE observed in the study. The mixture of provenances in this study, which includes the Florida provenance, may be a contributing factor to the GCA by test interactions in several traits. McKeand et al (1997) reported potentially large GxE effects were possible for composite traits in conifer breeding populations, especially if traits were negatively correlated. The Type B correlation for sawtimber potential was not as high as individual growth and stem traits, but did not exhibit sufficient interactivity to be of concern. Provenance differences could influence results from this study, but were not accounted for in the analysis because of inadequate representation of the three provenances in the original mating design.

Heritability estimates for growth traits fall within the range of previous work (Baltunis et al. 2007; Isik et al. 2005; Li et al. 1996; McKeand et al. 2008). The low individual-tree heritability for sawtimber potential could be attributed to a relatively subjective score based on several characteristics. Busby (1983) reported that individual branch trait measurements were more repeatable than a subjective crown score. While branching traits were generally more heritable at the individual tree level in this study, sawtimber potential appears to be a useful screening trait at the family selection level.

Rust incidence and sweep were moderately correlated with sawtimber potential. This is expected given the relationship between the penalty for having stem rust infection or excessive sweep in the sawtimber potential score. Incidence of forking was unfavorably correlated (r_g=0.18) with growth in analyses of second-generation progeny tests of loblolly pine (Xiong et al. 2009). In this study, forking incidence was uncorrelated with sawtimber potential, but moderately correlated with volume and height (an unfavorable effect). The positive correlation between forking and growth traits is potentially a problem for multiple trait improvement. However, the relatively low incidence of forking across the four test sites makes the genetic parameter estimates less reliable. Heritability estimates for branching traits demonstrate that improvements could be made through breeding and selection. The individual tree heritability for branch angle was lower in this study than that reported previously by Adams and Morganstern in *Pinus banksiana* (1991) and Isik and Isik (1999) in *Pinus brutia*, but these were studies of trees of different ages or in different species, which makes it difficult to draw conclusions. We found that branch diameter was

positively correlated with stem diameter, which may limit the reduction of branch and knot size in trees if growth continues to be improved. While branching traits are ultimately important for end-product quality, they were not as influential in assessing sawtimber potential as growth, rust incidence, and stem form.

Potential to capture greater value in forest plantations

Improvements in sawtimber potential may be possible through indirect selection of traits currently under selection in the NCSU Cooperative since it is heritable at the family level and is favorably correlated with growth and form traits. Multiple regression results suggest that selection for increased volume and reduced incidence of rust, sweep, and forking will increase sawtimber potential (Table 5). Historically, truncation selection has been practiced within the NCSU Cooperative to remove genotypes with undesirable rust and sweep breeding values, but a weighting based on economic value or impact may be more efficient for future selection and breeding. Figure 1 demonstrates that emphasis on both sawtimber potential and growth can be used to select elite parents for deployment. Potential gains in sawtimber proportion or volume could be lost if the emphasis is not placed on both traits. Relatively small increases in the proportion of sawtimber that is harvested in forest plantations can dramatically increase the value at harvest. By deploying seedlings from families based only on volume, the potential sawtimber quality of the plantation may be reduced if quality traits are not emphasized appropriately.

The assignment of value based on volume and the sawtimber potential score revealed a substantial increase in value if emphasis is placed on sawtimber quality rather than growth alone (Table 4). While the individual tree value is low, the assignment of value to each tree based on the sawtimber potential score demonstrates that volume alone may not account for economic value. Traits affecting the stem quality must be incorporated into breeding programs to improve the future sawtimber quality in loblolly pine plantations.

These results are based on visual grading in a non-thinned progeny test at age six and does not take into account stand improvements that can be made during mid-rotation activities such as thinning and pruning. A minor defect that graded trees into lower grades may not be apparent at rotation age. Large rust galls, low forks that are clearly visible at early ages will persist, but branching characteristics and small amounts of sweep could be difficult to see at harvest and may not be an economic factor when grading standing timber. Choi et al. (2008) reported that 37% of forked trees recovered to a single stem by age 15 years, so we recognize that not all defects observed at an early age will have an impact on visual stem quality at later ages.

Implications for multiple trait improvement

Interpreting categorical scores can be difficult to translate to meaningful values. However, these indices show that by selecting on sawtimber potential, there would be no change in branch angle, a reduction in branch diameter, and an increase in branch frequency (internode length) (Table 6). Branch angle is essentially unchanged in all indices, but

branch diameter increases by 5% in Indices 1 and 2. Branch frequency increases the greatest when volume is weighted, but it is improved in all three indices.

Sawtimber potential is improved in all three indices, but the largest improvement is made when sawtimber potential alone is the economically important trait. A 5 to 7 % reduction in the mean score translated to having a higher proportion of trees scored in sawtimber categories versus being scored as pulpwood or as non-merchantable. If sawtimber potential were used as the only selection criterion (Index 3), there would be an improvement in percent sawtimber from 38% to 58% in the population. Selection on volume, rust incidence, sweep, and forking (Index 2) would also improve percent sawtimber to 50% or higher.

Index 2 represents the improvement in traits that would be possible in growth and form traits using the trait data available in the NCSU Cooperative. While the greatest improvement in sawtimber potential would be made by weighting it over other traits, using an index with the available BLUP breeding values for volume, rust incidence, stem form (sweep or straightness) and forking would improve sawtimber potential. Since these are the traits that the NCSU Cooperative has used for selection, improvements made in NCSU Cooperative populations have increased the likelihood of producing sawtimber quality trees in loblolly pine plantations.

The potential to increase tree form and volume appear possible in this study, but wood quality traits were not included in this analysis. Atwood et al. (2002) reported a weak

negative genetic correlation between specific gravity and tree volume in Florida provenance loblolly pine, but demonstrated through varying weights in a selection index that gains in both traits were possible. Ivkovic et al. (2006) found that improving growth, MoE, and stem form traits by ten percent in radiata pine increased the financial returns in forest plantations, but the most valuable trait improvement varied according to the business model. For plantation growers, volume increases yielded the highest net present value (NPV) increases, but for integrated companies an increase in MoE was most valuable. The true economic value of improving growth versus form for sawtimber quality in loblolly pine may vary according to markets. Further efforts to quantify the economic importance of individual traits is needed to appropriately weight traits for tree improvement programs.

There are several limitations of the current study. Sawtimber potential was assessed at age 6, thus it is not a true measure of rotation age sawtimber quality, but it is an approximation of actual sawtimber quality for genetic trials in loblolly pine. One limitation to simultaneous gains in volume and sawtimber potential in this population is the moderate positive (unfavorable) correlation between volume and forking (0.32). Reducing the incidence of forking while improving growth is not possible in this population and many other loblolly pine populations (Xiong et al. 2009), but it may be possible to use a restriction index to minimize the increase of forking incidence if a reduction in volume gain is taken. However, only a low fork in the first log (4.8m) was considered a defect for sawtimber. A tree with a high fork could still be graded as sawtimber if there was a merchantable log in the tree below the fork, but it is not known if there are genetic

differences in the height of forking. The correlation between growth and forking incidence may not be consistently limiting for improvement of both growth and sawtimber potential.

Utilization of a selection index to improve both growth rate and sawtimber quality is a potentially valuable tool for loblolly pine breeding programs to consider. As programs progress in the development of elite breeding populations with specific trait combinations and end products, the proper emphasis on important quality traits will be a critical component of future genetic improvements in loblolly pine.

A major challenge for incorporating additional traits into breeding programs is determining economic weights for index selection. Wood property improvement has been debated since tree improvement programs began in the southeastern United States (Lowe et al. 1999). Wood properties can be expensive to measure, and the desired objectives can vary among companies and mills. Unlike the issues of wood property improvement, stem form improvement could readily increase value to the timber grower and improve wood going into the mills with the existing fee structure. Value for improved sawtimber quality would be assessed on the stump, therefore plantations with higher proportions of sawtimber would be graded as such and timber buyers would be more willing to pay premium prices for stumpage based on their ability to merchandize the wood at harvest. Payment by mills for improved wood properties or reduced occurrence of timber defects is already defined by the log grading schemes that are commonly used in the industry.

Conclusions

A composite trait to score the potential sawtimber quality of loblolly pine shows great promise to be a useful tool for breeding programs in the southeastern United States. In this study of 48 parents, family variation was significant for growth and form traits that influence sawtimber potential. Traits that are currently measured in the North Carolina State University Cooperative Tree Improvement Program explain 79% of the variation in sawtimber potential, suggesting that efforts made in the previous two cycles of improvement have increased the potential sawtimber quality of loblolly pine breeding populations. The value of trees at 6 years of age was substantially increased when sawtimber quality was also used as a selection criterion demonstrating the importance of form and quality traits for economic return. Improving growth, disease resistance, and sawtimber quality through a selection index appears to be a promising opportunity to increase quality and add value to future forest plantations.

Literature Cited

Adams, G.W., and E.K. Morgenstern. 1991. Multiple-trait selection in jack pine. *Can. J. For. Res.* 21(4):439-445.

Allen, H.L., T.R. Fox, and R.G. Campbell. 2005. What is ahead for intensive pine plantation silviculture in the South? *S. J. Appl. For.* 29:62-69.

Amateis, R.L., and H.E. Burkhart. 2005. The influence of thinning on the proportion of peeler, sawtimber, and pulpwood trees in loblolly pine plantations. *S J. Appl. For.* 29:158-162.

Atwood, R.A., T.L. White, and D.A. Huber. 2002. Genetic parameters and gains for growth and wood properties in Florida source loblolly pine in the southeastern United States. *Can. J. For. Res.* 32(6):1025-1038.

Baltunis, B.S., B.S. Baltunis, H.X. Wu, and M.B. Powell. 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(11):2164-2174.

Burdon, R.D. 1977. Genetic correlation as a concept for studying genotype environment interaction in forest tree breeding. P. 168-175 in Silvae genet.

Busby, C.L. 1983. Crown-quality assessment and the relative economic importance of growth and crown characters in mature loblolly pine. *Proc. 17th Sou. For. Tree Imp. Conf.*(39):121-130.

Choi, J., R.L. Amateis, and H.E. Burkhart. 2008. Modeling trends in stem quality characteristics of loblolly pine trees in unthinned plantations. *Can. J. For. Res.* 38:1446-1457.

Gilmour, A.R., R.D. Anderson, and A.L. Rae. 1985. The analysis of binomial data by a generalized linear mixed model. *Biometrika* 72(3):593-599.

Gilmour, A.R., R.D. Anderson, and A.L. Rae. 1987. Variance components on an underlying scale for ordered multiple threshold categorical data using a generalized linear mixed model. *J. Anim. Breed. Genet.* 104(3):149-155.

Gilmour, A.R., B.J. Gogel, B.R. Cullis, and R. Thompson. 2006. ASReml User Guide Release 2.0. P. 320pp. VSN International Ltd, Hemel Hempstead, UK.

Goebal, N.B., and J.R. Warner. 1966. Volume Tables for Small Diameter Loblolly, Shortlead, and Virginia Pine in the Upper South Carolina Piedmont. Clemson University. For. Res. No. 7.

Isik, F., D.D. Boos, and B. Li. 2005. The distribution of genetic parameter estimates and confidence intervals from small disconnected diallels. *Theor. Appl. Gen.* 110:1236-1243.

Isik, F., B. Goldfarb, X. Sun, M. Gumpertz, and B. Li. 2008. Analysis of cellulose microfibril angle using a linear mixed model in Pinus taeda clones. *Can. J. For. Res.* 38(6):1676-1689.

Isik, K., and F. Isik. 1999. Genetic variation in Pinus brutia Ten. in Turkey. II. Branching and crown traits. *Silvae genet.* 48(6):293-302.

Ivkovic, M., H.X. Wu, T.A. McRae, and M.B. Powell. 2006. Developing breeding objectives for radiata pine structural wood production. I. Bioeconomic model and economic weights. *Can. J. For. Res.* 36:2920-2931.

Jayawickrama, K.J.S. 2001. Genetic parameter estimates for radiata pine in New Zealand and New South Wales: a synthesis of results. *Silvae genet.* 50(2):45-53.

Jayawickrama, K.J.S., and C.B. Low. 1999. Pinus radiata selections from different regions in New Zealand differ in branch habit, form, and growth rate. *N. Z. J. For. Sci.* 29(1):3-24.

Lambeth, C. 2000. Realized genetic gains for first generation improved loblolly pine in 45 tests in coastal North Carolina. *S J. Appl. For.* 24(3):140-144.

Li, B., S. McKeand, and R. Weir. 2000. Impact of forest genetics on sustainable forestry--results from two cycles of loblolly pine breeding in the U.S. *J. Sust. For.* 10(1/2):79-85.

Li, B., S.E. McKeand, and R.J. Weir. 1996. Genetic parameter estimates and selection efficiency for the loblolly pine breeding in the southeastern U.S. P. ix + 264 *in Tree improvement for sustainable tropical forestry. Proceedings of the QFRI-IUFRO conference, Caloundra, Queensland, Australia, 27 October-1 November 1996. Volume 2.*, Dieters, M.J., A.C. Matheson, D.G. Nikles, C.E. Harwood, and S.M. Walker (eds.). Queensland Forestry Research Institute, Gympie; Australia.

Lin, C.Y. 1978. Index selection for genetic improvement of quantitative characters. *Theor. Appl. Gen.* 52(2):49-56.

Lopes, U.V., D.A. Huber, and T.L. White. 2000. Comparison of methods for prediction of genetic gain from mass selection on binary threshold traits. *Silvae genet.*(1):50-56.

Lowe, W.J., T.D. Byram, and F.E. Bridgwater. 1999. Selecting loblolly pine parents for seed orchards to minimize the cost of producing pulp. *For. Sci.* 45(2):213-216.

Mantysaari, E.A., R.L. Quaas, and Y.T. Grohn. 1991. Simulation study on covariance component estimation for two binary traits in an underlying continuous scale. *J. Dairy Sci.* 74(2):580-591.

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? *J For.* 104:352-358.

McKeand, S.E., B. Li, J.E. Grissom, F. Isik, and K.J.S. Jayawickrama. 2008. Genetic parameter estimates for growth traits from diallel tests of loblolly pine throughout the southeastern United States. *Silvae genet.* 57(3):101-110.

SAS. 1989. SAS/STAT User's Guide, Version 6. SAS Institute Inc., Cary, NC. 846 p.

Schall, R. 1991. Estimation in generalized linear models with random effects. *Biometrika* 78(4):719-727.

Sierra-Lucero, V., S.E. McKeand, D.A. Huber, D.L. Rockwood, and T.L. White. 2002. Performance differences and genetic parameters for four coastal provenances of loblolly pine in the Southeastern United States. *For. Sci.* 48(4):732-742.

Timber Mart-South 2007. Available online at www.TimberMart-South.com/tmart/; last accessed June 1, 2007.

Xiong, J.S., F. Isik, S. McKeand, and R. Whetten. 2009. Genetic Variation of Stem Forking in Loblolly Pine. *Submitted to Forest Science*.

Yamada, Y. 1962. Genotype by environment interaction and genetic correlation of same trait under different environments. *Jpn. J Gen.* 37(6):498-509.

Table 1. Means (and standard errors) of traits by test location. Traits were measured at age six in the field trials.

	Test Location			
Traits	**Yulee, FL N30.65° W81.75°**	**Brunswick, GA N31.39° W81.65°**	**Bainbridge, GA N30.87° W84.63°**	**Phenix City, AL N32.26° W85.19°**
Volume (dm^3)	31.4 (0.34)	40.6 (0.36)	38.1 (0.32)	34.0 (0.33)
Height (m)	8.0 (0.02)	8.4 (0.02)	8.2 (0.02)	8.0 (0.02)
Dbh (cm)	11.0 (0.05)	12.3 (0.06)	12.4 (0.04)	11.8 (0.05)
Fusiform Rust Freq.	0.23 (0.01)	0.31 (0.01)	0.40 (0.01)	0.56 (0.01)
Forking Freq	0.09 (0.01)	0.07 (0.01)	0.32 (0.01)	0.21 (0.01)
Sweep	2.0 (0.05)	2.2 (0.05)	2.1 (0.05)	1.6 (0.05)
Branch Angle	2.7 (0.1)	2.6 (0.01)	2.7 (0.01)	2.7 (0.01)
Branch Diameter (cm)	1.3 (0.01)	1.6 (0.01)	1.9 (0.01)	1.7 (0.01)
Branch Frequency	1.6 (0.02)	1.6 (0.02)	1.8 (0.02)	1.7 (0.02)
Sawtimber Potential	2.7 (0.01)	2.5 (0.01)	2.5 (0.01)	2.8 (0.01)

Table 2. Variance components, percent variance explained by terms, and heritability estimates (with standard errors) for 6-year old traits; GCA= General Combining Ability, SCA = Specific Combining Ability, GCAxT = GCA by Test interaction, SCAxT = SCA by Test interaction, h^2_i = individual tree narrow-sense heritability, h^2_{HS} = half-sib narrow-sense family mean heritability, H^2_{FS} = Full-sib broad-sense heritability, and $^rg_{B\_GCA}$= Type B GCA correlation.

Trait/ Term	GCA	SCA	GCAxT	SCAxT	Error	h^2_i	h^2_{HS}	H^2_{FS}	$^rg_{B\_GCA}$
Volume	9.080 * 4%	6.562* 3%	2.068* 1%	1.165 0%	191.8 91%	0.16 (0.04)	0.91 (0.03)	0.84 (0.03)	0.81 (0.11)
Height	0.06* 5%	0.08* 7%	0.01* 1%	0.003 0%	0.867 86%	0.22 (0.06)	0.94 (0.02)	0.91 (0.02)	0.92 (0.04)
Dbh	0.17* 4%	0.23* 5%	0.06* 1%	0.03 1%	4.4 90%	0.13 (0.04)	0.89 (0.04)	0.83 (0.03)	0.79 (0.08)
Rust	0.20* 6%	0.06* 2%	0.01 0%	0.02 1%	3.29 91%	0.22 (0.05)	0.96 (0.02)	0.91 (0.02)	0.96 (0.04)
Forking	0.14* 4%	0.01 0.5%	0.04 1%	0.01 0.5%	3.29 94%	0.15 (0.04)	0.92 (0.04)	0.82 (0.06)	0.84 (0.11)
Sweep	0.19* 4%	0.04* 1%	0.03* 1%	0.0 0%	4.17 94%	0.16 (0.04)	0.95 (0.02)	0.83 (0.03)	0.89 (0.05)
Branch Angle	0.01* 4%	0.001 0%	0.002* 1%	0.0 0%	0.257 95%	0.16 (0.04)	0.73 (0.33)	0.72 (0.16)	0.84 (0.06)
Branch Diameter	0.01* 3%	0.007* 1%	0.001* 0%	0.002 0%	0.386 96%	0.11 (0.03)	0.95 (0.02)	0.77 (0.05)	0.87 (0.07)
Branch Frequency	0.016 * 3%	0.004* 1%	0.000 0%	0.004* 1%	0.462 95%	0.13 (0.03)	0.98 (0.01)	0.80 (0.04)	1.00 (0.00)
Sawtimber Potential	0.00 5* 1%	0.010* 3%	0.002* 1%	0.000 0%	0.317 95%	0.06 (0.02)	0.85 (0.07)	0.76 (0.04)	0.72 (0.11)

* indicates the variance component estimate is significantly different from zero using a Likelihood Ratio Test with 1 DF.

Table 3. Additive genetic correlations (and standard errors in parentheses) among 6-year growth and stem quality traits[1]

	Dbh	Volume	Sweep	Rust	Fork	BA	BD	BF	Saw
Height	0.87 (0.04)	0.88 (0.04)	0.05 (0.16)	-0.03 (0.20)	0.42 (0.15)	-0.05 (0.16)	0.45 (0.14)	0.56 (0.12)	-0.77 (0.07)
Dbh	*	0.97 (0.01)	0.23 (0.15)	0.05 (0.16)	0.29 (0.16)	-0.27 (0.15)	0.63 (0.10)	0.23 (0.16)	-0.71 (0.09)
Volume		*	0.18 (0.16)	0.00 (0.16)	0.32 (0.15)	-0.28 (0.15)	0.56 (0.11)	0.33 (0.15)	-0.67 (0.09)
Sweep			*	0.26 (0.15)	0.16 (0.16)	-0.27 (0.15)	0.36 (0.15)	0.14 (0.16)	0.39 (0.15)
Rust				*	-0.08[†]	0.13 (0.16)	0.45 (0.13)	0.09 (0.16)	0.41 (0.14)
Forking					*	0.29 (0.15)	-0.07 (0.17)	0.27 (0.16)	0.05 (0.18)
BA						*	-0.46 (0.13)	0.01 (0.16)	0.03 (0.17)
BD							*	0.37 (.15)	0.03 (0.17)
BF								*	-0.01 (0.17)

[1] BA = Branch Angle, BD = Branch Diameter, BF = Branch Frequency, Saw = Sawtimber Potential.

[†]The correlation among 2 binary traits was estimated using parental breeding values.

Table 4. Summary of value and volume improvement over local checklot for the 10 highest value parents and the 10 highest volume parents based on breeding values

Grouping	Top 10 Value	Top 10 Volume	Local Checklot
Average Tree Value (USD)	$ 0.55	$ 0.50	$ 0.21
Average Tree Volume (dm^3)	39.6	43.0	28.9
Volume Gain Over Local Checklot	37%	49%	-
Value Gain Over Local Checklot	162%	138%	-

Table 5. Multiple regression coefficients for growth and form trait parental breeding values on Sawtimber Potential ($R^2 = 0.79$). Stepwise regression was used to select the best 4 traits.

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	1	0.34	0.04	8.48	<.0001	0
Volume	1	0.44	0.03	15.09	<.0001	1.11
Rust	1	-0.16	0.03	-5.77	<.0001	1.05
Forking	1	-0.16	0.03	-5.06	<.0001	1.12
Sweep	1	-0.17	0.01	-11.73	<.0001	1.07

Table 6. Trait weights and genetic gains over the population mean from three selection indices to improve growth and sawtimber quality.

			Index1 (volume only)			Index2 (multiple traits)			Index3 (sawtimber only)	
Trait	**Mean**	**Weight**	**Index Mean**	**Gain %**	**Weigh t**	**Index Mean**	**Gain %**	**Weigh t**	**Index Mean**	**Gain %**
Height (m)	8.22	0	8.51	3	0	8.52	4	0	8.04	-2
Dbh (cm)	11.96	0	12.17	2	0	12.17	2	0	11.84	-1
Volume (dm^3)	36.52	1	42.58	17	3	42.56	17	0	40.80	12
Sweep (cm)	1.92	0	2.05	7	-1	2.02	5	0	1.72	-11
Rust Incidence[1]	0.37	0	0.41	0.04	-1	0.36	-0.01	0	0.10	-0.27
Forking[1]	0.17	0	0.56	0.39	-1	0.54	0.37	0	0.42	0.25
Branch Angle	2.68	0	2.66	-1	0	2.66	-1	0	2.70	1
Branch Diameter	1.62	0	1.72	6	0	1.71	6	0	1.65	2
Branch Frequency	1.70	0	1.85	9	0	1.84	8	0	1.79	5
Sawtimber Potential	2.62	0	2.49	-5	0	2.48	-5	-1	2.44	-7

[1]For binary traits Rust Incidence and Forking, the Gain% is calculated as the change in probability of incidence .

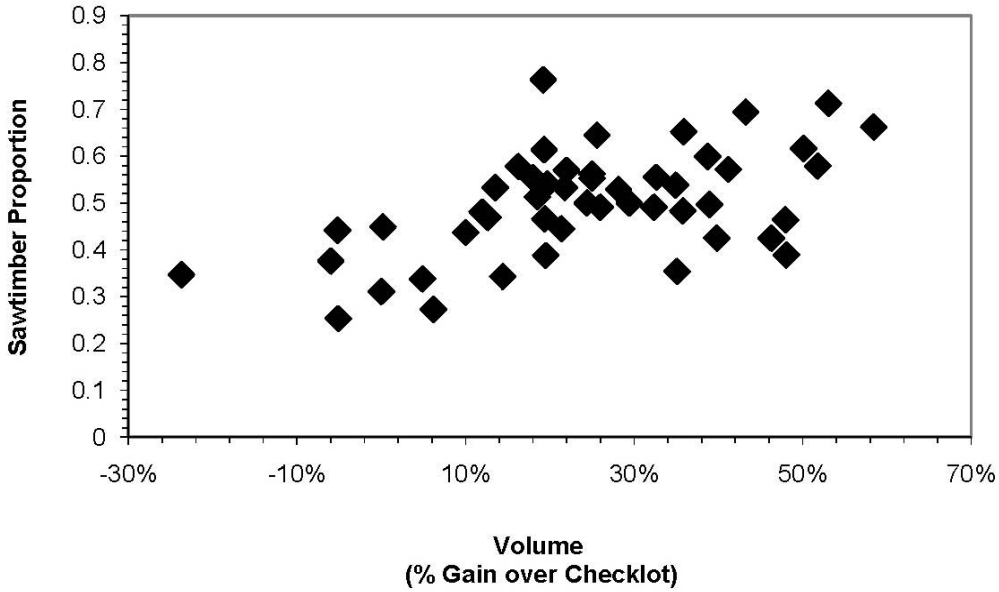

Figure 1. Proportion of sawtimber trees versus volume gain (%) by parental (half-sib) breeding value. The 48 parents used in this experiment demonstrate the variation in volume growth and the proportion of potential sawtimber trees based on age 6-year progeny test data.

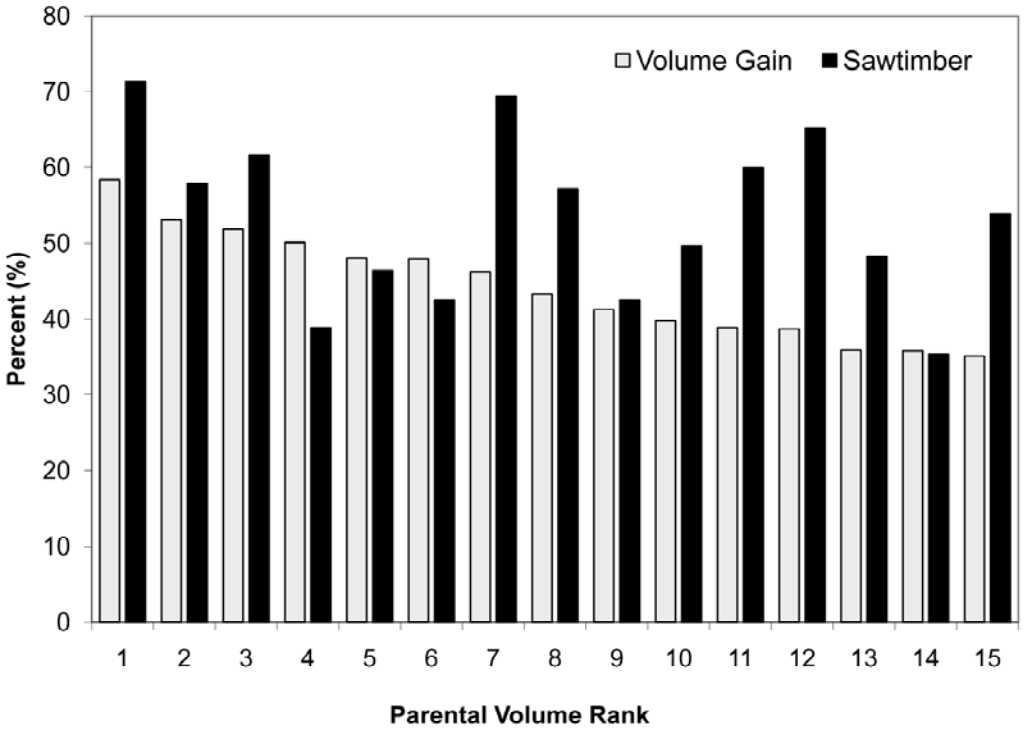

Figure 2. Volume gain and sawtimber proportion by half-sib breeding value for the 15 highest volume parents. Variation in the proportion of sawtimber quality trees can be seen among the parents with the highest breeding values for volume.

CHAPTER 3

Association genetics and marker effects to improve growth and stem quality in loblolly pine (*Pinus taeda* L.)

(*in the format appropriate for submission to Tree Genetics & Genomes*)

Association genetics and marker effects to improve growth and stem quality in loblolly pine (*Pinus taeda* L.)

W. Patrick Cumbie[1,] Fikret Isik[1], Ross Whetten[1], Bailian Li[1], Dahlia Nielsen[2],

John King[1], Andrew Eckert[2,3], Jill Wegrzyn[4], David Neale[4,5], Barry Goldfarb[1]

[1] Department of Forestry and Environmental Resources
North Carolina State University
Campus Box 8002
Raleigh, NC 27695-8002

[2]Department of Genetics
North Carolina State University
Raleigh, NC 27695

[3] Section of Evolution and Ecology
University of California at Davis
Davis, CA 95616, USA

[4]Center for Population Biology
University of California at Davis
Davis, CA 95616, USA

[5] Department of Plant Sciences
University of California at Davis
Davis, CA 95616, USA

[6]Institute of Forest Genetics
USDA Forest Service
Davis, CA 95616, USA

*Corresponding Author (email: patrick_cumbie@ncsu.edu)

Abstract

An association genetics approach was taken to identify single nucleotide polymorphisms (SNPs) associated with variation in growth and stem form traits in loblolly pine (*Pinus taeda* L.). Associations were tested between 4,200 SNPs and breeding values for a population of 200 largely unrelated selections of loblolly pine. We identified 13 SNP-phenotype associations for sawtimber (n=4), volume (n=5), and stem straightness (n=4) after multiple test correction. Five SNPs were identified in known sequences including a GRAS family transcription factor, a calmodulin family binding protein, CDPK adapter protein 1, and lysophosphatidic acid acyltransferase. Individual SNPs explained from 0% to 27% of the variance in breeding values used as phenotypes. The most significant SNPs (lowest p-value) were used to estimate genetic values for an independent population of 153 clonally replicated trees. The correlation between marker based estimated genetic values and the BLUP predictions for volume was highest when 10 to 25 SNP loci were used (r=0.27). Gain estimates from marker based selection scenarios were compared to seedling and clonal progeny testing scenarios to explore the needed reliability of marker-based estimates to assess the incorporation of marker based selection in loblolly pine breeding programs. Comparisons reveal that even low reliability values for marker-based selection may be more efficient in making genetic gain for within-family forward selections compared to traditional seedling progeny testing but not as efficient as clonal progeny testing.

Introduction

Estimating an individual tree's breeding value is essential to the success and improvement of forest tree breeding programs. Improvements in analytical methods and experimental designs have been utilized to remove environmental variation and efficiently estimate a parent's breeding value based on field-grown progeny in common garden experiments (Dutkowski et al. 2002; Gezan et al. 2006; Isik et al. 2004). However, the reliability of predicting an individual's breeding value based on phenotype from progeny testing remains low for within-family selection (Isik, 2004). Efforts to improve efficiency in making gains can be attained by shortening the breeding cycle, increasing the reliability of predicting breeding value, or both (Bernardo and Yu 2007; Hospital et al. 1997; Meuwissen et al. 2001). As genomic data become available for loblolly pine (*Pinus taeda* L.) the use of molecular markers to predict an individual's genetic value or breeding value is one potential method to increase the gain per unit time in tree breeding programs by reducing the need for extensive progeny testing and shortening the breeding cycle (Strauss et al. 1992).

In loblolly pine (*Pinus taeda* L.), measurements are often taken at after five or six growing seasons and selections are made for the next generation of improvement (McKeand 1988; McKeand and Bridgwater 1998). A comparison of seedling versus clonal progeny tested revealed substantial improvements for within-family selection using clonal replication, but at a substantial additional cost for testing and additional time to develop the clonally replicated trees (Isik et al. 2004). Previous research also demonstrated that the individual-

tree heritability estimates in loblolly pine are low (Balocchi et al. 1993; Baltunis et al. 2007; Lambeth et al. 1983). Marker-based selection could be an alternative to conventional field-testing if methods improve sufficiently to produce adequate reliability to predict the breeding value or genetic value of an individual without waiting for data from field-tested progeny.

Genomic resources are being developed for forest tree species, including loblolly pine, which provide opportunities to test for molecular markers associated with desirable phenotypes in breeding and natural populations (Neale and Ingvarsson 2008). The potential of marker-aided selection and breeding has been discussed for forest tree improvement programs but to date has not been widely implemented. Previous work to identify quantitative trait loci (QTL) explaining variation in forest tree populations was largely family based with limited application to large out-crossing breeding populations. Experiments in the genus *Pinus* have been successful in identifying significant QTL which explained variation in tree growth, wood density, and carbon isotope discrimination (Brendel et al. 2002; Emebiri et al. 1998; Kaya et al. 1999).

The more recent development of association mapping methods has provided a tool for breeders to identify polymorphic markers (single nucleotide polymorphisms or SNPs) associated with variation in economically important traits on a population level (Neale and Savolainen 2004). Association mapping experiments in forest trees have identified marker-trait associations with important traits such as cold tolerance, wood density and water use efficiency but these traits are not necessarily the focus of loblolly tree breeding programs

(Eckert et al. 2009b; González-Martínez et al. 2008; González-Martínez et al. 2007). Improvements in growth rate, stem form, and disease resistance are the major emphases in pine breeding programs in the southeastern United States (McKeand and Bridgwater 1998). If marker based selection and breeding is to be successful it will likely need to improve the gain and efficiency in one of these economic traits (Strauss et al. 1992). Associations identified with growth traits will help demonstrate the potential value or limitations for implementing marker-based information in loblolly pine breeding programs.

Nelson and Johnsen (2008) suggest that genomic resources will allow forest tree breeders to incorporate specific physiological traits associated with alleles of genes that regulate a plant function. However, experiments quantifying the physiological processes related to tree growth have not yet identified key physiological processes that explain overall growth differences in loblolly pine (Emhart et al. 2006; Martin et al. 2005; McKeand and Svensson 1997). An alternative to identifying markers associated with specific plant functions is to find associations between markers and the traditional traits measured such as height, volume, or diameter growth of trees. The association of molecular markers to economic traits may not provide insight regarding the genes related to an underlying physiological process, but could assist breeders in capturing genetic variation in breeding populations.

Validation of marker associations and the estimation of marker effects are critical for justifying the cost and effort to incorporate markers in breeding programs. Effects of alleles at individual loci can be difficult to estimate accurately and small numbers of markers which explain a small proportion of the total variance provide limited utility in

predicting quantitative traits. Traditional marker aided selection (MAS) in plant species has focused on few large effect QTL. In maize, response from selection on QTL resulted in varied levels of gain (Heffner et al. 2009) and Moreau (2004) reported no gain after implementing MAS in two selection cycles. Meuwissen (2001) proposed an approach called genomic selection in which markers are used across the entire genome to explain genetic variance. The availability of SNPs in sufficient density at cost-effective prices is making genomic selection feasible in some organisms. In dairy cattle populations, reliability of marker based breeding values ranged from 0.3 to 0.7 using experimental data and as high as 0.97 in simulations (de Roos et al. 2009), but these methods seek to include hundreds to thousands of markers across the genome of an organism. de Roos et al. (2009) simulated genomic selection in two diverging populations over a range of heritabilities and found that low heritability traits had predicted genetic values with low reliability compared to high heritability traits. Genomic selection has been applied in barley (Xu 2003a) and in theory shows great potential to use dense markers across the genome to account for many genes with small effects to predict the genetic value or breeding value for quantitative traits.

Genomic selection has currently not previously been posible in loblolly pine due to lack of a dense set of markers across the genome, and large genotyped populations. However, recent association studies provide a set of markers across multiple populations to quantify the potential reliability of predicting genetic value based on markers across populations. As a part of the USDA-CAP funded Conifer Translational Genomics Network we seek to develop marker associations that can be implemented in conifer breeding programs.

In this experiment we tested for SNP associations with an index trait for sawtimber quality improvement and the individual traits contributing to this composite phenotype (Chapter 2); and then used BLUP predictions to evaluate the potential reliability of predicting genetic values for volume using SNPs associated with quantitative traits for within-family selection in tree breeding programs to complement current breeding methods.

Materials and Methods

Phenotypic data

The phenotypes used in this analysis were breeding values for selections in the North Carolina State University – Cooperative Tree Improvement Program (www.treeimprovement.org). Traits included breeding values for tree height, volume, stem straightness, rust incidence, and stem forking. Briefly, breeding values for tree height and volume were estimated using Best Linear Unbiased Prediction using a mixed model with standard form:

$$[1] \qquad \mathbf{y} = \mathbf{Xb} + \mathbf{Zu} + \mathbf{e}$$

where **y** is the vector of response variable (observations); the **X** is the design matrix relating individual observations to the fixed effects in the model, **b** is the vector of fixed effect factors that includes the overall mean, site and replication within site effects; **Z** is the incidence matrix relating observations to random effects and **u** is the vector of random effects that includes general combining ability (GCA) effects, specific combining ability

(SCA) effects and their interactions with site; **e** is the vector of random residual terms. Prior to analysis, tree height and volume were standardized on individual site means.

For binary traits (stem straightness, stem forking and rust incidence) the following generalized linear mixed model was fit to data to obtain variance components:

[2] $$g^{-1}(\boldsymbol{\eta}) = g^{-1}(\mathbf{Xb} + \mathbf{Zu})$$

where $\boldsymbol{\eta}$ is the linear predictor ($\boldsymbol{\eta}$ = logit($\boldsymbol{\mu}$)), g is the link function, and $\boldsymbol{\mu}$ is the conditional mean. The remaining terms are as previously explained. (Schall 1991).

In addition to individual traits, an index trait (sawtimber index) was generated using weights adapted from a previous analysis of stem quality in loblolly pine (Chapter 2) where tree volume, rust incidence, stem straightness, and stem forking were weighted in a ratio of 3:-1:1:-1, respectively. A total of 200 individuals from two breeding regions (Coastal and Piedmont) were used in this study for association testing.

Genotypic Data

Genotypes for single nucleotide polymorphisms (SNPs) were obtained using the Illumina Infinium™ assay (Illumina, San Diego, CA). Smaller Illumina platforms have been shown to work well within the large and complex genome of conifers (Eckert et al. 2009b). The methods used in development of these SNPs are available online at http://dendrome.ucdavis.edu/adept2/. Briefly, SNPs were detected in 7508 resequenced amplicons generated from all available unique EST contigs representing all pine ESTs

known to date. From the resequenced amplicons, roughly 22,000 SNPs were discovered and 7216 were selected for genotyping using an Infinium™ genotyping chip. Based on quality, reliability of reads and polymorphism, 4200 SNPs were selected using the BeadStudio ver. 3.1.3.0 software (Eckert et al. 2009a).

Association Testing

Association tests were performed in TASSEL using the Generalized Linear Model option (Bradbury et al. 2007). Most selections in the population were plantation selections made on the basis of phenotypic performance. Although there were a few cases of sibs in the population of 200 these were not accounted for in the association testing. We used the positive false discovery rate approach to adjust p-values for multiple testing (Storey 2003), using the qvalue package in R with a false discovery rate of 0.10 (RDevelopmentCoreTeam 2005).

SNP effects

For each significantly associated SNP with three genotype classes present, we estimated the additive (a) and dominance (d) effects using ASReml 2.0 (Gilmour et al. 2006). SNPs were treated as fixed effects to test for significance and to generate best linear unbiased estimates (BLUEs) for genotype classes of each SNP. Additionally, variance estimates of significant SNPs were generated by adding a SNP effect into the mixed model analyzing phenotypic data, where the SNP effect was treated as a random effect with 1 or 2 degrees of freedom.

To obtain annotations for SNPs, flanking sequences of the corresponding EST contig were obtained from the Dendrome database (http://dendrome.ucdavis.edu/treegenes) for all SNPs associated with a trait after multiple testing correction. We performed a BLASTx query against the NCBI non-redundant protein database (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Gene and site annotations for the strongest hits (lowest e-value) for each sequence are reported.

Predicting Genetic Value from Associated SNPs

SNPs were ranked on the basis of p-value results from the association tests. For this analysis SNP loci were estimated individually, using the association population (n=200), as random effects in a mixed model and BLUP predictions were made for each genotype class at each SNP locus calculated as:

[3] $$\hat{g} = (X'X + \lambda I)^{-1} X'y$$

Where $\hat{\boldsymbol{g}}$ is the vector of effects of the genotype classes at the SNP locus; X is the design matrix with the number of rows equal to the number of records which allocates the SNP allele and the mean to the phenotype; y is the vector of phenotypes and λ is the shrinkage factor of σ^2_e / σ^2_{SNP} in which σ^2_e is the residual variance and σ^2_{SNP} is the variance among genotype classes for an individual SNP locus. Genotypes were coded as 0,1,or 2 where the minor allele was set to zero so that the SNP effect column reflects the number of copies of the major allele. The sum of the individuals' BLUP predictions (as a deviation from the mean) was used as the marker estimated genetic value:

[4] $$MEGV = \sum_{i}^{n} X_i \hat{g}_i$$

where MEGV is the marker estimated genetic value; $\boldsymbol{X}_i$ is design matrix allocating the individual tree to the genotype effect at SNP locus ***i***, and $\hat{\boldsymbol{g}}$ is the vector of effects of the genotype classes within the SNP locus ***i***.

Data from a second population consisting of 153 clones from 15 crosses, also genotyped with the same set of 4200 SNPs, was available for testing the value of the estimated marker effects on predicting volume. MEGV's were estimated using the volume breeding values from the association population of 200 progeny tested selections and were then applied to both the association population and the clonal validation population. MEGV was estimated for the best 5, 10, 15, 20, 25, 50, and 100 SNPs based on association test results for volume. Correlations between each individual's MEGV and the phenotypic volume for each population were calculated to explore the potential utility of marker based estimates for selection in loblolly pine breeding programs.

Gain and selection

Isik et al. (2004) reported scenarios for conventional seedling and clonal progeny testing for increasing within-family gains in tree improvement programs. We extended this scenario to include marker-based selection with estimated costs for field testing, vegetative propagation, and SNP genotyping. We estimated gain according to Falconer (1989):

[5] $$R = ih^2 \sigma_P$$

where R is the response to selection (Gain); i is the selection intensity calculated from the selection proportion; h^2 is the heritability (or square of the correlation in the case of marker based scenarios); and σ_P is the total phenotypic variance. For this comparison we assumed $\sigma_P = 1$ for all scenarios.

We used the assumptions for breeding and testing cycles from Isik et al. (2004), where selection of the current cycle requires five years, grafting and breeding new selections requires three years, seed collection and seedling rearing requires one year, and progeny testing the new selections using field-grown seedlings or clonal rooted cuttings requires six years. An additional two years is required for vegetative propagation for clonal testing. To evaluate scenarios we assumed 200 parents, 80 trees per parent tested in the field at four locations totaling 16,000 trees with a selection proportion of 2:100. For the clonal progeny testing scenario we tested 4000 clones with 4 ramets each with a selection proportion of 2:25. Marker-based scenarios required the same time for current selection, grafting and breeding, seed harvest and seedling growth in the greenhouse. Once seedlings were grown, DNA could be collected and individuals could be genotyped for SNPs associated with the trait of interest. Cost estimates for marker genotyping were taken from the Washington University Sequenom Center http://hg.wustl.edu/info/Sequenom_description.html) where 25 SNPs can be genotyped on 384 samples for $2304. Field testing and vegetative propagation costs were provided by the NCSU-Cooperative Tree Improvement Program where a single progeny test would cost $20,000 from test establishment through data

collection at age six years, and clonal propagation costs are approximately $20 per genotype (clone).

Results

Association testing

Significant associations were found following multiple testing correction for sawtimber index, volume, and straightness. All individual traits followed a normal distribution as did the sawtimber index (Figure 1). The sawtimber index score, a composite trait of volume, straightness, rust incidence, and forking, had overlapping associations with each trait that was a component of the index. The proportion of SNPs with p-values less than 0.05 ranged from 6.1% in Straightness to 7.0% in Height. The number of SNPs in common among a pair of traits was related to the correlation among phenotypic values ($r = 0.91$). Sawtimber index was most highly correlated with volume ($r = 0.85$) since it is the most highly weighted trait in the index and shared the greatest number of SNPs in common (165), thus nearly 60% of the SNPs associated with the sawtimber index were also significant for volume prior to multiple test corrections (Table 1). P-value distributions from association testing revealed an increase in the number of SNP loci with p-values on the low end of the distribution for the sawtimber index, volume, straightness, and rust incidence (Figure2a-f). Distributions for forking and height revealed fewer p-values below 0.05. The association population had an average minor allele frequency of 0.14 and its distribution reveals a large number of SNP loci with rare alleles ($MAF<0.05$). Application of the q-value adjustment

for multiple test correction (Figure 3) yielded 13 significant associations between SNPs and phenotypes (Table 2).

Minor allele frequencies for significantly associated SNPs ranged from 0.03 to 0.43 and individual SNPs explained 9 to 11% (r^2) of the variation in the linear model. No SNPs were significantly associated with height, rust disease incidence, or stem forking after correction for multiple testing. Distributions of p-values for height and forking revealed a reduced number of SNPs at the low end of the distribution (Figure 2). Sawtimber index and volume shared 2 SNPs in common after correction for multiple testing, 0_8448_02_252 and 0_13919_02_271.

Seven SNP loci included all three possible genotype classes which allowed the estimation of additive and dominance effects. Dominance effects were less than additive effects for all SNPs. When SNPs were treated as random effects in individual SNP models the percentage of variance explained by the SNP loci ranged from 0% up to 27% (Table 2). Standard errors were generally high reflecting the imprecision in estimating the effect of individual loci in this population.

Five SNPs were located in sequences similar to known genes or sequences. SNP 0_2222_02_74 was similar to a GRAS family transcription factor, SNP 0_12452_03_87 was similar to a CDPK adapter protein in Arabidopsis, SNP 0_15969_01_107 was similar to the at5g35570 k2k18_1 sequence in Arabidopsis, SNP 0_9918_01_641 was similar to a putative lysophosphatidic acid acyltransferase in rice and SNP 0_14875_01_63 was similar to a calmodulin-binding family

protein (Table2). The remaining SNPs were located in sequences with no similarity to known proteins.

MEGV for Volume

Marker estimated genetic values were estimated using the SNPs with highest significance for volume (ranked on p-value). Within the association population, MEGV was well correlated with the phenotypes from which it was estimated for volume (Figure 5). Correlations increased within the association population as additional markers are added to the MEGV. When MEGV was estimated for the validation population, the correlation between MEGV and phenotype was much lower. The correlation was as high as 0.25 to 0.27 when 5 to 10 loci were used to estimate MEGV, but then no further increase in correlation was observed through 25 loci and it decreased when more than 25 SNPs were used (Figure 5).

Gain Estimates

The marker-based selection cycles are six years shorter than seedling progeny test scenario and eight years shorter than clonal testing (Table 4). When the same phenotypic variance ($\sigma_P = 1$) is used in the scenarios, clonal testing results in the highest gain (122%) due to the high repeatability (or heritability) of within-family selection, while seedling progeny testing yielded the lowest gain (7%) with the lowest repeatability for these within-family scenarios. At higher levels of reliability modeled for marker-based selection, the higher reliability and shorter breeding cycle times lead to increases in the %Gain per year relative to conventional

seedling progeny testing. When assumptions on costs for testing and genotyping are included, all four marker-based scenarios yield gain at lower costs as compared to seedling tests (Table 5). Only the marker-based program with a reliability of 0.35 produced gain at a lower unit cost compared to clonal progeny testing.

Discussion

SNP associations

Significantly associated SNPs were variable in the proportions of total variance among breeding values they explained ranging from 0% up to 27%. For association testing we took a two-stage approach in which a large progeny test dataset was analyzed first and BLUP predictions for progeny tested parents were used as phenotypes in association tests with SNP genotypes in the parents. The breeding value estimates are based on many observations, and accuracy values are generally high for growth traits. The precision of phenotype in the association tests could be one reason higher percentages of variance are explained by individual SNPs but estimates still appear to be biased upwards when applied to the validation population. No significant SNPs were observed for height, rust incidence, or forking in this population. It is somewhat surprising that height, a component of volume estimation, did not have any significant associations while five SNPs showed associations with volume after correction for multiple testing. A previous study by Kaya et al. (1999) identified QTL explaining 7% to 30% of the phenotypic variation for height increment and 12.5% to 59.5% of the phenotypic variation for diameter growth. Growth measurements on the entire tree (height, diameter, volume) are composites of many plant functions that to

date, have not been successfully correlated to specific physiological differences (Martin et al. 2005). The identification of a small number SNPs associated with volume begins to dissect the complexity of growth, but further identification of SNPs associated with specific plant functions may provide greater insight into the underlying genetic architecture of complex quantitative traits in forest trees.

Five of the 12 SNP loci associated with traits in this study had homologs in other species (primarily Arabidopsis) based on similarities in the sequences flanking the SNP. For the sawtimber index SNP 0-8448-02-252 was similar to a GRAS family transcription factor. GRAS transcription factors are unique to plants and play various roles in plant growth and development. The GRAS family of transcription factors includes the SCARECROW-like group of genes identified in Arabidopsis, and this diverse family play various roles in GA signaling, root development, light signal transduction, and axillary shoot formation (Bolle 2004; Lee et al. 2008).

SNP 0_12452_03_87, associated with volume growth, is in a sequence similar to a CDPK Adapter Protein 1 in Arabidopsis. Calcium dependent protein kinases (CDPK's) are one of the major groups of proteins which decode calcium signals and are thought to be involved in a wide range of plant functions including abiotic stress response and adaptation, wound response, and transcriptional regulation (Patharkar and Cushman 2006). In a review by Kirkby and Pilbeam (1984), calcium was described as important in plants to maintain cell integrity and membrane stability, cell division and elongation. Underlying genetic variation in the ability to regulate or utilize Ca could have an impact on overall tree growth.

SNP 0_14875_01_63), associated with straightness, was in a sequence similar to a gene encoding a calmodulin-binding (CaM) family protein. CaM binding proteins are important in plant responses to biotic and abiotic stresses such as UV, ozone, drought, salt, cold, wind, and pathogens (Du and Poovaiah 2004). A review by Bowler and Fluhr (2000) discussed the importance of calcium and calmodulin for response to different stress factors on plants in multiple environments and its involvement in signaling abscisic acid (ABA).

While these similarities are not conclusive evidence that the SNPs identified in this study are causative, they are located in sequences similar to functional genes that could influence the associated trait. The 4,200 SNPs in the study reflect a small portion of genetic diversity present in the pine genome, which exceeds 24,000 MB (Ahuja and Neale 2005; Morse et al. 2009). The addition of more markers across the genome as well as larger populations will aid in further dissection of complex growth traits.

MEGV

The MEGV-phenotype correlation was high in the association population where the associations were detected and marker effects were estimated, as expected. The individual SNP locus effects are likely to be overestimated in the association population (Xu 2003b), which is consistent with observations of MEGV in the validation population. Marker effects explain substantially less variation in the independent validation population than in the association population. The estimates of marker effects would likely be more accurate if more individuals were added to the association population. The MEGV in the validation population remained steady from 0.25 to 0.27 as the number of markers increased from 5 to

10 suggesting that only the most significant SNPs (smallest p-values) in the association population actually explained variation in the validation population. Alternatively, the estimates of the effects of the additional SNPs may have lacked precision. The decrease in the correlation as more markers were added to estimate the MEGV most likely reflects the addition of non-significant SNPs which introduced error into the marker based prediction. Individual SNP effects were estimated using BLUP rather than a least squares approach so that the estimated effect would be shrunk towards the mean, but effects still lacked precision when estimated in a population of 200 individuals.

While this experiment was not genomic selection, the results from this study are consistent with simulations of genomic selection where a low number of markers and low heritability traits have poor correlations with true breeding values, and thus low reliability measures (de Roos et al. 2009). Simulated and real populations in dairy cattle reveal accuracy values ranging from 0.3 to 0.7 in real breeding populations and theoretically as high as 0.97 in simulated populations with higher heritability traits. De Roos (2009) showed through simulations that low heritability traits benefited from the addition of more phenotypes, which increase the precision of estimated effects, rather than the addition of more markers. As more individuals are genotyped in loblolly pine breeding populations and additional SNPs are identified across the genome, the opportunity to incorporate markers into breeding programs to enhance selection is plausible. The approach taken in this study was to rank markers based on associations between SNPs and high accuracy breeding values based on progeny testing. Future efforts will likely adopt an alternative strategy of

genotyping progeny and incorporating a genomic selection approach. If sufficient marker density is achieved, tree breeders may successfully capture sufficient genetic variation to increase the reliability of marker predicted genetic values. In lieu of a genomic selection approach we demonstrate that a small but useful portion of the phenotypic variation could be captured with a small number of SNPs. This approach could be implemented into existing breeding programs for much less expense than that required for genotyping of thousands of additional markers on many individuals for genomic selection.

Marker-based selection

In forest tree breeding programs backwards selection through progeny testing has been very effective, and the accuracy of parental breeding values from this method are high. Rogued second-generation seed orchards have yielded an estimated 26% to 35% gain in rotation-age volume (Li et al. 2000). The results of estimating MEGV in this study using the available populations and markers demonstrates that marker-based selection will not equal the reliability or accuracy of backward selection (and ultimately the genetic gain) if implemented at the present. However, there is potential to incorporate markers for within-family selection. The scenario of seedling testing with heritability or reliability at 0.03 is a low threshold to overcome. Using the square of the MEGV-phenotype correlation (r^2) from the validation population we estimate a reliability of 0.07 using 10 to 25 markers to predict volume growth. This reliability is higher than the seedling scenario we used in this comparison and is in the range of other individual tree heritability estimates for volume (Atwood et al. 2002; Baltunis et al. 2007; Cumbie 2002). A similar approach was discussed

by Kumar (2001) where QTL could be used to assist in making forward selections within full-sib families in radiata pine. Clonally replicated progeny tests generate more gain but do so at a significant time requirement compared to marker-based selection. Marker reliabilities above 0.25 provide % Gain per year exceeding that of the clonal progeny test scenario. A complementary approach using open-pollinated or polymix breeding as described by McKeand and Bridgewater (1998) could be used for backward selection to evaluate parents for deployment in open-pollinated seed orchards while a marker-based selection program could be used to save time and maintain a modest amount of gain through within-family forward selections.

From the viewpoint of the breeder, marker aided selection must increase gain or gain efficiency to be considered for application. The time and expense to install, maintain, and measure progeny tests is a substantial effort but has proven to be successful in loblolly pine (Li et al. 2000). Marker resources may soon offer a viable alternative to extensive progeny testing for within-family selection as genotyping costs continue to fall and high-throughput genotyping and sequencing technologies become available to the forest tree breeding community. In the scenarios we consider, 10 to 25 markers to replace within-family selection could greatly reduce breeding cycle time and provide similar gain at a lower cost than conventional seedling progeny testing.

Conclusions

The underlying genetic variation of economically important quantitative traits in loblolly pine is complex. Association tests revealed a few significant SNPs associated with a

sawtimber index trait, volume and straightness, after multiple testing correction. In this study we identified 13 SNP-phenotype associations and five of these were in sequences similar to those of other species. While a reductionist approach may eventually identify a sufficient number of genes to explain variation in specific physiological functions which impact overall growth in trees, genetic markers may be able to capture variation in traits without knowing the function of the sequence in which they are located. If sufficient linkage disequilibrium between the marker and the QTL exists to reliably capture variation by selecting on markers, within-family gains could be made in future tree breeding programs with significant reduction in the breeding cycle at greater efficiency than conventional seedling progeny testing.

Literature Cited

Ahuja, M.R., and D.B. Neale. 2005. Evolution of genome size in conifers. *Silvae genetica* 54(3):126-137.

Atwood, R.A., T.L. White, and D.A. Huber. 2002. Genetic parameters and gains for growth and wood properties in Florida source loblolly pine in the southeastern United States. *Can. J. For Res* 32(6):1025-1038.

Balocchi, C.E., F.E. Bridgwater, B.J. Zobel, and S. Jahromi. 1993. Age trends in genetic parameters for tree height in a nonselected population of loblolly pine. *For. Sci.* 39(2):231-251.

Baltunis, B.S., D.A. Huber, T.L. White, B. Goldfarb, and H.E. Stelzer. 2007. Genetic analysis of early field growth of loblolly pine clones and seedlings from the same full-sib families [electronic resource]. *Canadian journal of forest research* 37:195-205.

Bernardo, R., and J. Yu. 2007. Prospects for Genomewide Selection for Quantitative Traits in Maize [electronic resource]. *Crop science* 47(3):1082-1090.

Bolle, C. 2004. The role of GRAS proteins in plant signal transduction and development. *Planta* 218(5):683-692.

Bowler, C., and R. Fluhr. 2000. The role of calcium and activated oxygens as signals for controlling cross-tolerance. *Trends in plant science* 5(6):241-246.

Bradbury, P.J., Z. Zhang, D.E. Kroon, T.M. Casstevens, Y. Ramdoss, and E.S. Buckler. 2007. TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* 23(19):2633-2635.

Brendel, O., D. Pot, C. Plomion, P. Rozenberg, and J.M. Guehl. 2002. Genetic parameters and QTL analysis of delta13C and ring width in maritime pine. *Plant, cell and environment* 25(8):945-953.

Cumbie, W.P. 2002. Variation of wood density traits in rooted cuttings and seedlings of loblolly pine, North Carolina State University, Raleigh. 96 p.

de Roos, A.P.W., B.J. Hayes, and M.E. Goddard. 2009. Reliability of Genomic Predictions Across Multiple Populations. *Genetics* 183(4):1545-1553.

Du, L., and B.W. Poovaiah. 2004. A novel family of Ca2+ /calmodulin-binding proteins involved in transcriptional regulation: interaction with fsh/Ring3 class transcription activators. *Plant Molecular Biology.* 54(4):549-569.

Dutkowski, G.W., J. Costa E Silva, A.R. Gilmour, and G.A. Lopez. 2002. Spatial analysis methods for forest genetic trials. *Canadian journal of forest research* 32(12):2201-2214.

Eckert, A., J.v. Heerwaarden, J. Wegrzyn, C.D. Nelson, J. Ross-Ibarra, S.C. González-Martínez, and D. Neale. 2009a. Patterns of population structure and environmental associations with aridity across the range of loblolly pine (Pinus taeda L., Pinaceae). *Genetics* in press.

Eckert, A.J., A.D. Bower, J.L. Wegrzyn, B. Pande, K.D. Jermstad, K.V. Krutovsky, J.B. St. Clair, and D.B. Neale. 2009b. Association Genetics of Coastal Douglas Fir (Pseudotsuga menziesii var. menziesii, Pinaceae). I. Cold-Hardiness Related Traits. *Genetics* 182(4):1289-1302.

Emebiri, L.C., M.U. Slee, A.C. Matheson, and M.E. Devey. 1998. Interval mapping of quantitative trait loci affecting NESTUR, a stem growth efficiency index of radiata pine seedlings. *Theoretical and applied genetics* 97(7):1062-1068.

Emhart, V.I., D.A. Huber, T.L. White, and T.A. Martin. 2006. Genetic variation in basal area increment phenology and its correlation with growth rate in loblolly and slash pine families and clones [electronic resource]. *Canadian journal of forest research* 36(4):961-971.

Falconer, D.S. 1989. Introduction to quantitative genetics Burnt Mill, Harlow, Essex, England : Longman, Scientific & Technical ; New York : Wiley, 1989.

Gezan, S.A., T.L. White, and D.A. Huber. 2006. Achieving higher heritabilities through improved design and analysis of clonal trials [electronic resource]. *Canadian journal of forest research* 36:2148-2156.

Gilmour, A.R., B.J. Gogel, B.R. Cullis, and R. Thompson. 2006. ASReml User Guide Release 2.0. P. 320. VSN International Ltd, Hemel Hempstead, UK.

González-Martínez, S.C., D. Huber, E. Ersoz, J.M. Davis, and D.B. Neale. 2008. Association genetics in Pinus taeda L. II. Carbon isotope discrimination. *Heredity* 101(1):19-26.

González-Martínez, S.C., N.C. Wheeler, E. Ersoz, C.D. Nelson, and D.B. Neale. 2007. Association Genetics in Pinus taeda L. I. Wood Property Traits. *Genetics* 175(1):399-409.

Heffner, E.L., J.-L. Jannink, and M.E. Sorrells. 2009. Genomic Selection for Crop Improvement [electronic resource]. *Crop science* 49(1):1-12.

Hospital, F., A. Charcosset, A. Gallais, L. Moreau, and F. Lacoudre. 1997. More on the efficiency of marker-assisted selection. *Theoretical and applied genetics* 95(8):1181-1189.

Isik, F., B. Goldfarb, J. Frampton, and B. Li. 2004. Efficiency of seedlings and rooted cuttings for testing and selection in Pinus taeda. *Forest science* 50(1):44-53.

Kaya, Z., D.B. Neale, and M.M. Sewell. 1999. Identification of quantitative trait loci influencing annual height- and diameter-increment growth in loblolly pine (Pinus taeda L.). *Theoretical and applied genetics* 98(3/4):586-592.

Kirkby, E.A., and D.J. Pilbeam. 1984. Calcium as a plant nutrient. *Plant, cell and environment* 7(6):397-405.

Kumar, S., and D.J. Garrick. 2001. Genetic response to within-family selection using molecular markers in some radiata pine breeding schemes. *Canadian journal of forest research* 31(5):779-785.

Lambeth, C.C., R.B. McCullough, S.D. Duke, and J.P.v. Buijtenen. 1983. Early selection is effective in 20-year-old genetic tests of loblolly pine. *Silvae genetica* 32(5/6):210-215.

Lee, M.-H., C.E. Lim, S.O. Sohn, D.G. Kim, J. Lim, M.M. Lee, K.S. Chang, J.-O. Heo, S.-K. Song, B. Kim, M. Kim, S.A. Lee, and N.-I. Yu. 2008. Large-scale analysis of the GRAS

gene family in Arabidopsis thaliana [electronic resource]. *Plant Molecular Biology* 67(6):659-670.

Li, B., S. McKeand, and R. Weir. 2000. Impact of forest genetics on sustainable forestry--results from two cycles of loblolly pine breeding in the U.S. *Journal of Sustainable. Forestry* 10(1/2):79-85.

Martin, T.A., S.E. McKeand, and P.M. Dougherty. 2005. Strategies and case studies for incorporating ecophysiology into Southern pine tree improvement programs. *Southern journal of applied forestry* 29(2):70-79.

McKeand, S., and J. Svensson. 1997. Sustainable management of genetic resources. *Journal of forestry* 95(3):4-9.

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., and F.E. Bridgwater. 1998. A strategy for the third breeding cycle of loblolly pine in the Southeastern US. *Silvae genetica.* 47(4):223-234.

Meuwissen, T.H.E., M.E. Goddard, and B.J. Hayes. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157(4):1819-1829.

Morse, A.M., D.G. Peterson, M.N. Islam-Faridi, K.E. Smith, Z. Magbanua, S.A. Garcia, T.L. Kubisiak, H.V. Amerson, J.E. Carlson, C.D. Nelson, and J.M. Davis. 2009. Evolution of Genome Size and Complexity in *Pinus*. *PLoS ONE* 4(2):e4332.

Neale, D.B., and P.r.K. Ingvarsson. 2008. Population, quantitative and comparative genomics of adaptation in forest trees. *Current opinion in plant biology* 11:149-155.

Neale, D.B., and O. Savolainen. 2004. Association genetics of complex traits in conifers. *Trends in Plant Science* 9(7):325-330.

Nelson, C.D., and K.H. Johnsen. 2008. Genomic and physiological approaches to advancing forest tree improvement. *Tree physiology* 28(7):1135-1143.

Patharkar, O.R., and J.C. Cushman. 2006. A novel coiled-coil protein co-localizes and interacts with a calcium-dependent protein kinase in the common ice plant during low-humidity stress [electronic resource]. *Planta* 225(1):57-73.

RDevelopmentCoreTeam. 2005. R: A language and environment for statistical computing, reference index version 2.x.x. . Computing, R.F.f.S. (ed.), Vienna, Austria.

Schall, R. 1991. Estimation in generalized linear models with random effects. *Biometrika* 78(4):719-727.

Storey, J.D. 2003. The positive false discovery rate: A Bayesian interpretation and the q-value. *Ann. Stat.* 31(6):2013-2035.

Strauss, S.H., G. Namkoong, and R. Lande. 1992. Limitations of molecular-marker-aided selection in forest tree breeding. *Canadian journal of forest research* 22(7):1050-1061.

Xu, S. 2003a. Estimating Polygenic Effects Using Markers of the Entire Genome. *Genetics* 163(2):789-801.

Xu, S. 2003b. Theoretical basis of the Beavis effect. *Genetics* 165(4):2259-2268.

Table 1. Significant SNPs ($p<0.05$) for each trait (diagonal), SNPs shared in common by two traits (above), and correlations among traits (below). Saw = sawtimber index; Vol = Volume; Ht = Height; STRT = Straightness; Rust = Rust; Fork = Fork.

	Saw	**Vol**	**Ht**	**Strt**	**Rust**	**Fork**
Saw	278	165	121	29	94	39
Vol	0.85	269	151	20	34	15
Ht	0.73	0.85	296	15	35	19
Strt	0.29	-0.1	0.01	258	21	22
Rust	-0.58	-0.21	-0.14	-0.17	282	21
Fork	-0.36	-0.06	0.01	-0.27	0.26	275

Table 2. Significant marker-trait associations after multiple testing correction for the sawtimber index, volume, and straightness. Height, forking and rust incidence had no significant associations after multiple test correction.

Trait	SNP Locus	Annotation	SNP	MAF	r^2	p-value	q-value*
Saw	0_13919_02_271	NA	[A/G]	0.04	0.11	2.6E-06	0.01
Saw	0_16546_01_530	NA	[G/A]	0.11	0.10	8.5E-05	0.10
Saw	0_2222_02_74	GRAS family transcription factor	[A/T]	0.26	0.09	9.3E-05	0.10
Saw	0_8448_02_252	NA	[G/A]	0.18	0.09	1.0E-04	0.10
Volume	0_8448_02_252	NA	[G/A]	0.18	0.13	2.1E-06	0.01
Volume	0_13919_02_271	NA	[A/G]	0.04	0.10	4.3E-06	0.01
Volume	0_16729_01_535	NA	[A/G]	0.12	0.09	7.5E-05	0.07
Volume	0_16710_02_265	NA	[G/A]	0.43	0.10	8.5E-05	0.07
Volume	0_12452_03_87	af384822_1 cdpk adapter protein 1	[G/A]	0.03	0.08	9.5E-05	0.07
Straightness	0_15969_01_107	at5g35570 k2k18_1	[C/G]	0.05	0.11	8.7E-06	0.04
Straightness	0_9918_01_641	ac068923_13 lysophosphatidic acid acyltransferase	[A/G]	0.07	0.10	3.9E-05	0.06
Straightness	0_14875_01_63	calmodulin-binding family protein	[A/G]	0.04	0.10	5.3E-05	0.06
Straightness	CL2166Contig 1_01_105	NA	[C/A]	0.12	0.10	5.6E-05	0.06

*SNP = [major/minor nucleotide base], MAF = Minor allele frequency, r^2 = model r-square from association test, p-value = p-value from linear model, and q-value = adjusted p-value for multiple test correction.

Table 3. Estimated effects of associated SNPs for Sawtimber, Volume, and Straightness; *a* is the additive substitition effect, *d* is the dominance deviation, SE_a is the standard error of *a*, SE_b is the standard error of *d*, %Variance is the percentage of variance among breeding values explained by the individual SNP.

Trait	SNP Locus	*a*	SE_a	*d*	SE_d	*d/a*	%Var.
Sawtimber	0_13919_02_271	0.134	0.064	0.067	0.063	0.50	27%
Sawtimber	0_16546_01_530	-0.114	0.067	-0.101	0.068	0.89	0%
Sawtimber	0_2222_02_74	0.036	0.014	-0.020	0.017	-0.57	20%
Sawtimber	0_8448_02_252	0.088	0.053	0.029	0.053	0.33	17%
Volume	0_8448_02_252	0.763	0.417	0.197	0.422	0.26	23%
Volume	0_13919_02_271	0.458	0.507	0.000	0.518	0.00	15%
Volume	0_16729_01_535	-0.607	0.442	-0.518	0.486	0.85	0%
Volume	0_16710_02_265	-*	-	-	-	-	0%
Volume	0_12452_03_87	-	-	-	-	-	9%
Straightness	0_15969_01_107	-	-	-	-	-	0%
Straightness	0_9918_01_641	-	-	-	-	-	4%
Straightness	0_14875_01_63	-	-	-	-	-	<0.01 %
Straightness	CL2166Contig1_01_105	-	-	-	-	-	3%

*additive and dominance effects were not estimated for SNPs that did not have all 3 possible genotype classes in the population.

Table 4. Selection scenarios, cycle lengths, selection ratios and intensities, %Gain and %Gain per year for conventional seedling testing, clonal testing, and marker-based selection.

Method	Reliability	Years per Stage‡	Total Cycle Time	Reliability per year	%Gain*	%Gain per Year
Seedling Progeny Testing†	0.03	5+3+1+6	15	0.002	7%	0.5%
Clonal Progeny Testing†	0.70	5+3+1+2+6	17	0.041	122%	7%
Marker-based Selection	0.05	5+3+1	9	0.006	12%	1%
Marker-based Selection	0.15	5+3+1	9	0.017	35%	4%
Marker-based Selection	0.25	5+3+1	9	0.028	58%	6%
Marker-based Selection	0.35	5+3+1	9	0.039	81%	9%

† According to Isik et al 2004.
‡ Stages of a tree improvement cycle: 5 years for selection, 3 years to graft and breed, 1 year to collect seed and raise seedlings, 6 years to progeny test, and an additional 2 years to produce rooted cuttings for clonal tests.
*Gain calculated as $G = ih^2\sigma_p$ where h2 ~ reliability, I =2.328 with a selection proportion of 2:100 except for Clonal testing where I = 1.745 with a selection proportion of 2:25. We assume σ_P is 1 for every scenario.

Table 5. Cost comparisons for within-family selection using conventional seedling progeny tests, clonal tests, and marker-based selection at different levels of reliability for marker-based predictions. A total of 16,000 trees are tested or genotyped in each scenario.

Selection Method	Reliability	# of Genotypes	Field Testing Cost ($)	Vegetative propagation Cost ($)	Cost to Genotype 25 Markers ($)	Total Selection Cost ($)	% Gain*	Cost per % Gain ($)
Seedling-Progeny†	0.03	16,000	$80,000	-	-	$80,000	7%	$11,455
Clonal - Progeny†	0.70	4,000	$80,000	$80,000	-	$160,000	122%	$1,310
Marker-based	0.05	16,000	-	-	$96,042	$96,042	12%	$8,251
Marker-based	0.15	16,000	-	-	$96,042	$96,042	35%	$2,750
Marker-based	0.25	16,000	-	-	$96,042	$96,042	58%	$1,650
Marker-based	0.35	16,000	-	-	$96,042	$96,042	81%	$1,179

*Gain estimates reported in Table 4.

A testing scheme for 200 parents, 80 trees per parent, 4 sites. Cost of a single field trial is estimated to be $20,000 and the additional cost to vegetatively propagate is$20 per genotype (unpublished NCSU-CTIP data)

Genotyping costs reported from Washington University Sequenom Center. $25 SNPs costs 2304 per 384 samples. (http://hg.wustl.edu/info/Sequenom_description.html)

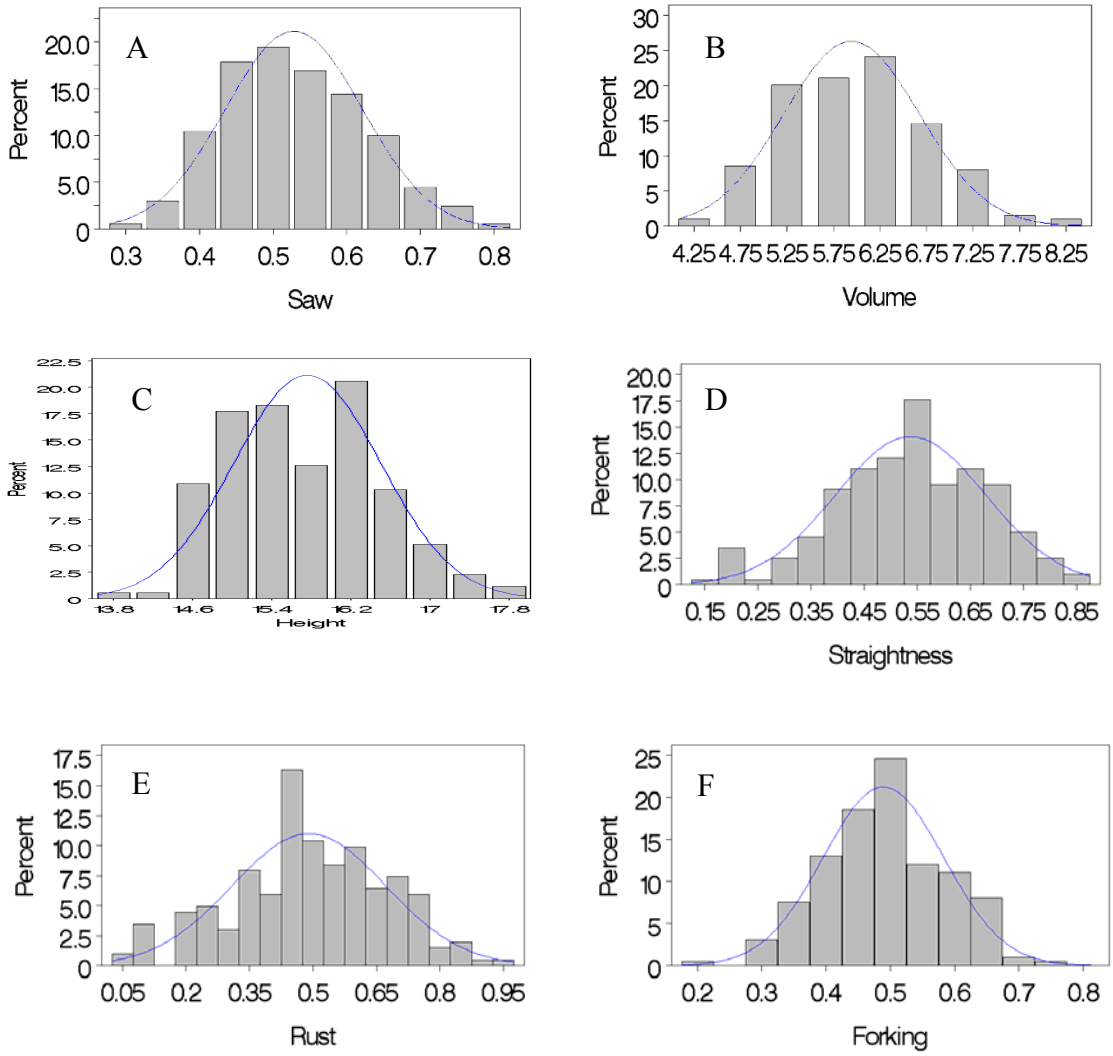

Figure 1. Distributions of trait phenotypes; a)Sawtimber Index (Saw), b) Volume, c) Height, d) Straightness,e) Rust Incidence, and f) Forking

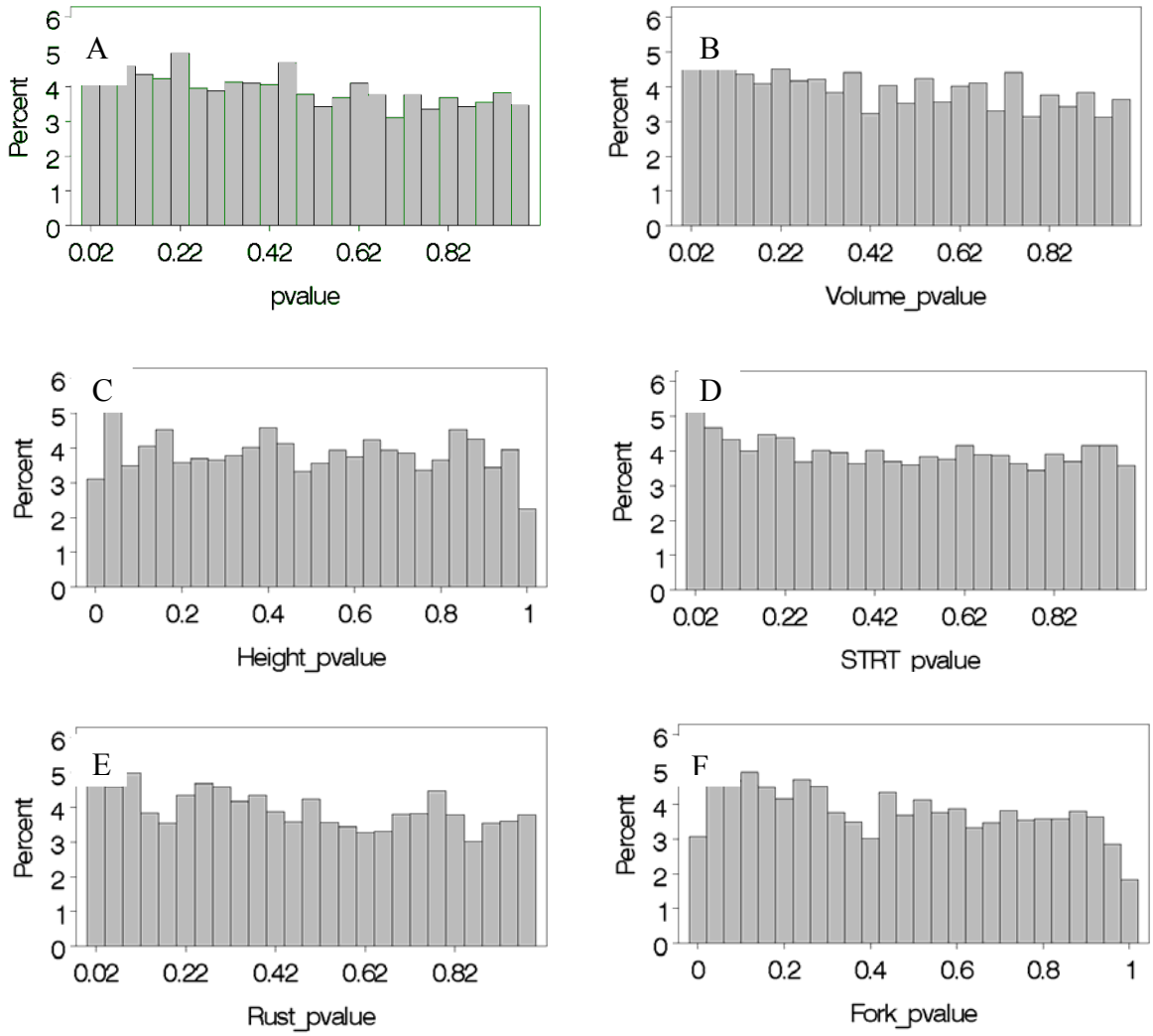

Figure 2. Pvalue distribution from association tests performed in TASSEL.4,205 SNPs were tested for associations with a) Sawtimber Index, b) Volume, c) Height, d)STRT (Straightness), e) Rust incidence, and f) Forking.

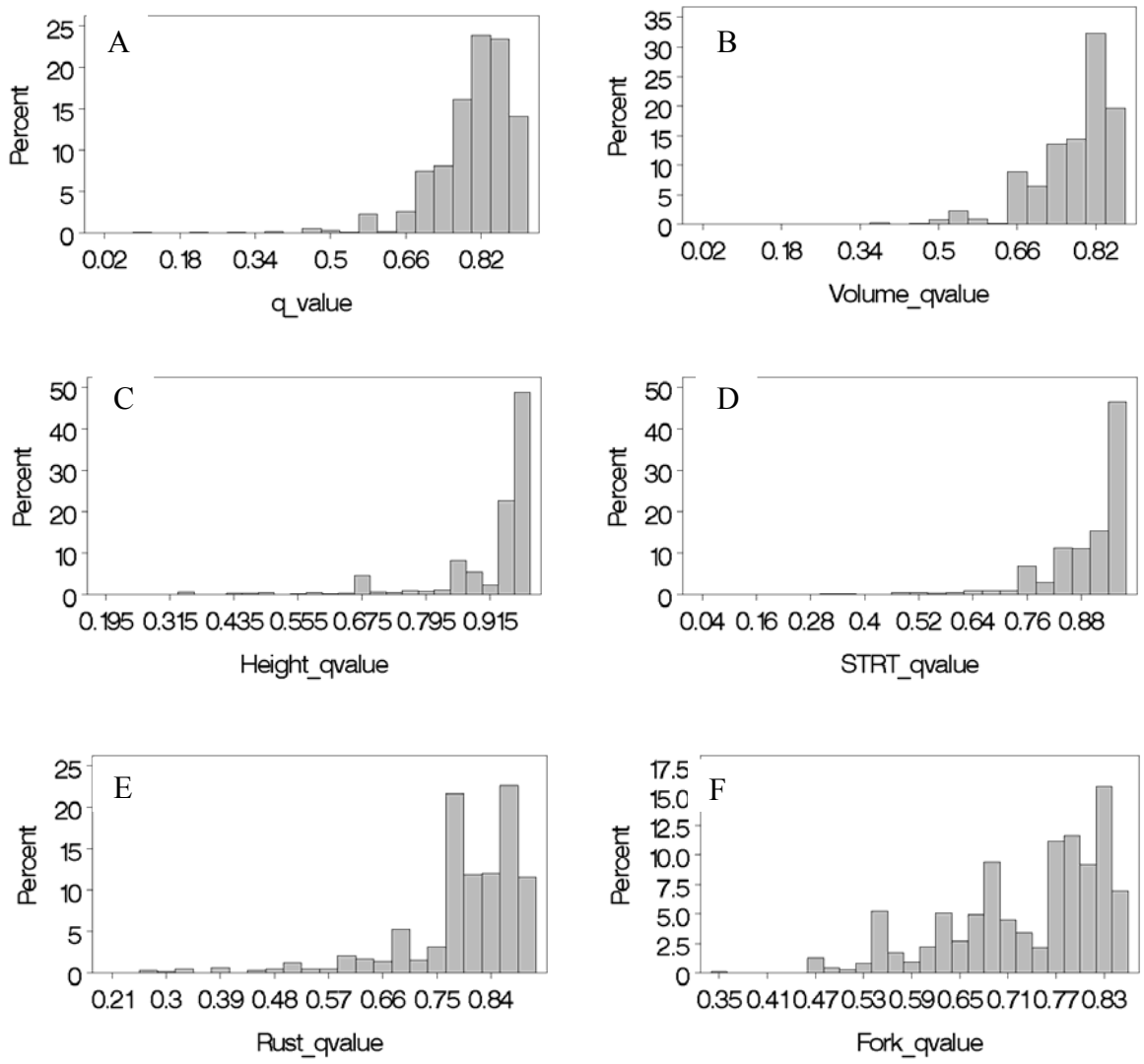

Figure 3. Qvalue distribution from association tests performed in TASSEL.4205 SNPs were tested for associations with a)Sawtimber Index, b) Volume, c) Height, d) STRT(Straightness), e) Rust incidence, and f) Forking.

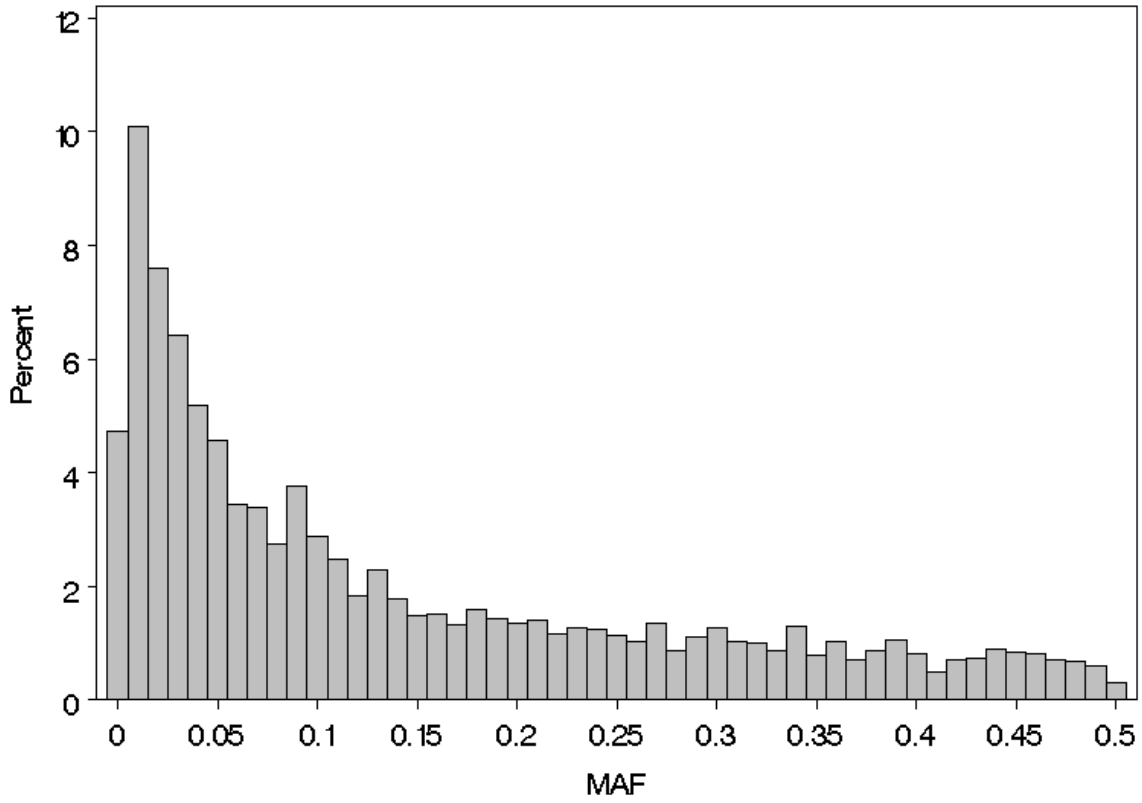

Figure 4. Minor Allele Frequences (MAF) for 4205 SNPs. The Mean MAF in the association population of 200 individuals was 0.14.

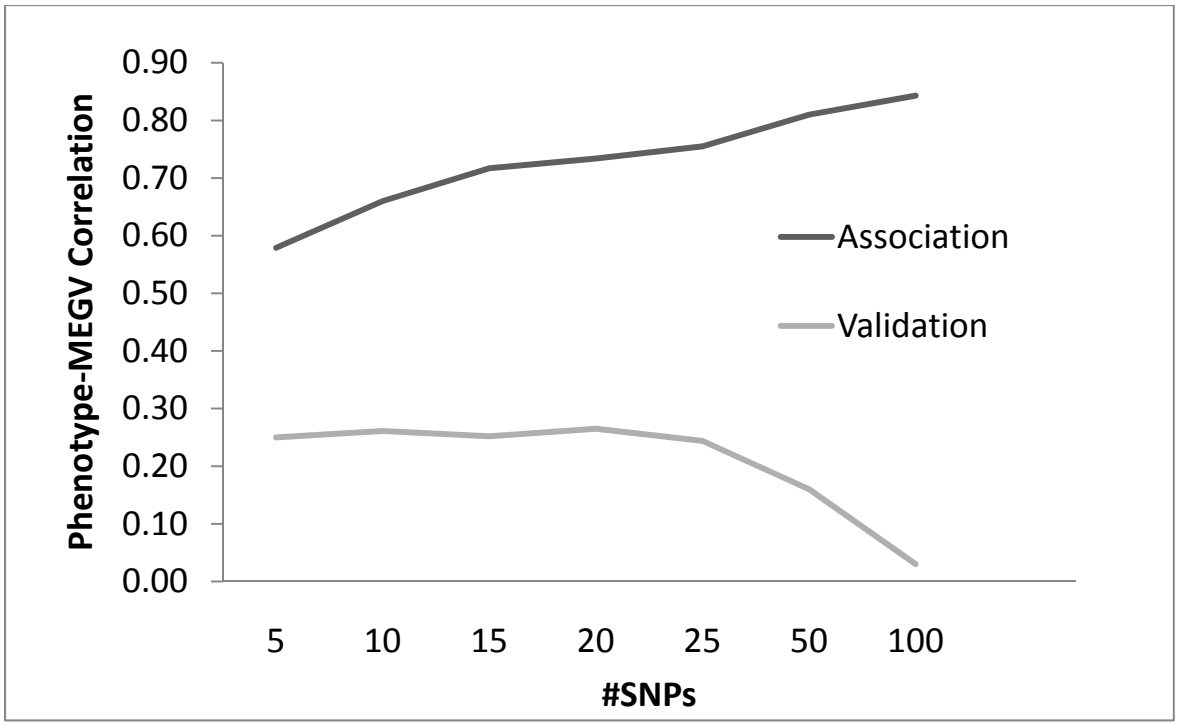

Figure 5. Correlations between phenotype and Marker Estimated Genetic Value (MEGV) in the association population (Association) and in the validation population (Validation) of 153 clonally replicated individuals by number of SNPs used to estimate MEGV..

SYNTHESIS

Genetic Variation in Loblolly Pine

Genetic variation in loblolly pine is substantial, and gains can be made through multiple approaches whether it is classical quantitative genetics, genomics, or a compliment of the two methods. Prior to beginning these experiments it was hypothesized, but not known, whether or not association genetics would yield positive results in forest trees. Candidate-gene based approaches used a small number of SNPs and results suggested potential associations, but larger numbers of markers had not been attempted. Association testing in these experiments revealed positive associations with potentially meaningful sequences for complex traits related to growth, water use efficiency, foliar nitrogen content, stem straightness, and an index trait assessing sawtimber potential. This list of traits suggests that there is potential in using genomic resources to identify the underlying genetic variation of quantitative traits. The results from these genetic association experiments do not begin to explain a majority of the variation, but suggest that larger, more powerful studies may be fruitful in identifying SNPs which explain variation in important forest tree traits.

Important limitations for association studies in loblolly pine are marker coverage across the genome and population size. At the time these experiments began it had been estimated that a population of 500 unrelated individuals would provide sufficient power to detect associations, but perhaps even larger populations would be helpful if rare alleles are

ultimately what is associated with specific traits. The SNPs used in these experiments represent 1 SNP per gene-like sequence publicly available for loblolly pine. Roughly 4000 SNPs were used in these experiments which represent approximately 10% of the estimated number of genes in loblolly pine. Being able to account for more variation within each gene-like sequence (more SNPs per gene) and having more gene sequences would provide greater opportunities to identify SNPs associated with traits of interest.

Future experiments in physiological traits, especially carbon isotope discrimination or other measures of water use efficiency, should be replicated in contrasting environments and on larger scales. This is a classical challenge for genetics experiments which attempt to measure physiological traits. The time and resources required for collecting physiology measurements on a large-scale is significant, but perhaps the forest research community is approaching a period where looking at multiple systems can be done. Concerns about adaptability, climate change, carbon sequestration, and sustainable forest productivity may provide funding opportunities for large scale genetics and physiology experiments.

If the association study in the first chapter of this thesis were repeated, hindsight would suggest greater replication of genotypes (perhaps 5 to 10 ramets) with greater control over watering regimes. Fortunately the second growing season (the season in which the needles sampled for isotopes grew) was under a severe drought, so at least moderate drought stress was placed on the experiment, but rainfall was not controlled.

Spatial models appear to be a useful tool in forestry trials. No matter how uniform a site or environment may appear there is likely some gradient or patchiness in the experiment that

will introduce variation and influence the experimental results. Capturing additional environmental variation with a spatial model in the nursery bed experiment increased the heritability by more than 10 percentage points. The spatial model removed a portion of environmental variation from the residual term which increased heritability estimates. If such tools are implemented in progeny test data analysis there could be significant improvements in genetic parameter estimates and gains.

Sawtimber Quality

The importance of sawtimber quality has been recognized by industrial forest products companies and has been improved in internal company populations, but until this quantitative study on sawtimber potential was conducted the cooperative breeding programs had not quantified sawtimber quality in breeding populations. The results from this study reveal that growth, rust resistance, and stem straightness are important factors in sawtimber quality. These individual traits have been the emphasis of breeding programs for loblolly pine but quantifying and estimating sawtimber gains had not been a priority. It is not surprising that these traits influence sawtimber quality, but until now the relative importance had not been quantified. For capturing value and maintaining relevance in the industry, breeding programs need to emphasize the improvement of the most economically valuable timber products. While product values may change over time a breeding program could maintain a broad base of germplasm and focus on the development of elite populations to address specific end products. If and when product emphasis needs to shift there is a genetic base from which to draw for future improvement.

The value of sawtimber is significantly higher than pulpwood. The straightforward comparison of value based on volume versus sawtimber potential demonstrated that additional value could be captured if tree form and stem quality is emphasized for seedling deployment. From the selection index comparison, a small increase in the sawtimber potential score would result in a substantial increase in value. Full-sib families in the Lower Gulf Elite population ranged widely in sawtimber percentage. Selecting the very best families for deployment will more than double the percentage of expected sawtimber in plantations as compared to unimproved material. These are gains which could readily be captured by selecting the appropriate open-pollinated seedlot or full-sib family seedlot for deployment.

The selection index for sawtimber potential improvement is somewhat limited in this study by the unfavorable correlation between forking and height. Sawtimber potential was not as influenced be forking (uncorrelated) because only low forks in the first log were considered a defect. The height of the fork was not measured but would likely add resolution to the issue. If this experiment were repeated the height of forking would be a valuable addition to the data set. Multiple regression showed the core traits to be what tree improvement programs currently measure, as previously stated. This could be bias in grading by the individual measuring trees, or the true relative importance of traits. Large diameter and steep branches have an impact on sawtimber quality, but not as much as volume, stem straightness or rust incidence. Branching traits had heritability estimates similar to growth

suggesting that specific branching characteristics could be improved through breeding if warranted.

Marker Applications in Loblolly Pine Breeding Programs

The potential reductions in breeding cycle time, increases in gain efficiency, and constantly falling genotyping costs require that tree breeders pay attention to possible applications of molecular markers to aid in selection and breeding. To date the evidence of marker-based gains has not reached a "critical mass" to start implementing marker aided selection (MAS) in tree breeding programs. As tree breeders explore genomic selection they may find a method which captures sufficient genetic variation to justify the cost of marker genotyping in breeding populations. Ways in which to incorporate MAS into current breeding strategies and populations will require considerable effort and should not be delayed. Marker-estimated values in this study were weakly correlated with phenotype in the validation population but still potentially useful. Within-family selection in seedling progeny tests yields a much lower amount of genetic gain than clonal tests, and marker-based estimates with low to moderate repeatability provide an intermediate level of gain compared to field testing options.

BIBLIOGRPAHY

Adams, G. W. and E. K. Morgenstern (1991). "Multiple-trait selection in jack pine." Canadian journal of forest research **21**(4): 439-445.

Ahuja, M. R. and D. B. Neale (2005). "Evolution of genome size in conifers." Silvae genetica **54**(3): 126-137.

Albaugh, T. J., K. H. Johnsen, P. M. Dougherty and H. L. Allen (2004). "Long term growth responses of loblolly pine to optimal nutrient and water resource availability." Forest ecology and management **192**(1): 3-19.

Allen, H. L., T. R. Fox and R. G. Campbell (2005). "What is ahead for intensive pine plantation silviculture in the South?" Southern journal of applied research. **29**: 62-69.

Amateis, R. L. and H. E. Burkhart (2005). "The influence of thinning on the proportion of peeler, sawtimber, and pulpwood trees in loblolly pine plantations." Southern journal of applied research. **29**: 158-162.

Anyia, A. O., J. J. Slaski, J. M. Nyachiro, D. J. Archambault and P. Juskiw (2007). "Relationship of Carbon Isotope Discrimination to Water Use Efficiency and Productivity of Barley Under Field and Greenhouse Conditions " Journal of agronomy and crop science **193**: 313-323.

Atwood, R. A., T. L. White and D. A. Huber (2002). "Genetic parameters and gains for growth and wood properties in Florida source loblolly pine in the southeastern United States." Canadian journal of forest research **32**(6): 1025-1038.

Balocchi, C. E., F. E. Bridgwater, B. J. Zobel and S. Jahromi (1993). "Age trends in genetic parameters for tree height in a nonselected population of loblolly pine." Forest science **39**(2): 231-251.

Baltunis, B. S., B. S. Baltunis, H. X. Wu and M. B. Powell (2007). "Inheritance of density, microfibril angle, and modulus of elasticity in juvenile wood of Pinus radiata at two locations in Australia." Canadian journal of forest research **37**(11): 2164-2174.

Baltunis, B. S., J. M. Davis, D. A. Huber and T. A. Martin (2008). "Inheritance of foliar stable carbon isotope discrimination and third-year height in Pinus taeda clones on

contrasting sites in Florida and Georgia [electronic resource]." Tree genetics & genomes **4**: 797-807.

Baltunis, B. S., D. A. Huber, T. L. White, B. Goldfarb and H. E. Stelzer (2007). "Genetic analysis of early field growth of loblolly pine clones and seedlings from the same full-sib families." Canadian journal of forest research **37**: 195-205.

Bernardo, R. and J. Yu (2007). "Prospects for Genomewide Selection for Quantitative Traits in Maize" Crop science **47**(3): 1082-1090.

Bolle, C. (2004). "The role of GRAS proteins in plant signal transduction and development." Planta **218**(5): 683-692.

Bouché, N. and H. Fromm (2004). "GABA in plants: just a metabolite?" Trends in Plant Science **9**(3): 110-115.

Bowler, C. and R. Fluhr (2000). "The role of calcium and activated oxygens as signals for controlling cross-tolerance." Trends in Plant Science **5**(6): 241-246.

Bradbury, P. J., Z. Zhang, D. E. Kroon, T. M. Casstevens, Y. Ramdoss and E. S. Buckler (2007). "TASSEL: software for association mapping of complex traits in diverse samples." Bioinformatics **23**(19): 2633-2635.

Brendel, O., D. Pot, C. Plomion, P. Rozenberg and J. M. Guehl (2002). "Genetic parameters and QTL analysis of delta13C and ring width in maritime pine." Plant, cell and environment **25**(8): 945-953.

Buckler, E. S., J. B. Holland, P. J. Bradbury, C. B. Acharya, P. J. Brown, C. Browne, E. Ersoz, S. Flint-Garcia, A. Garcia, J. C. Glaubitz, M. M. Goodman, C. Harjes, K. Guill, D. E. Kroon, S. Larsson, N. K. Lepak, H. Li, S. E. Mitchell, G. Pressoir, J. A. Peiffer, M. O. Rosas, T. R. Rocheford, M. C. Romay, S. Romero, S. Salvo, H. S. Villeda, H. Sofia da Silva, Q. Sun, F. Tian, N. Upadyayula, D. Ware, H. Yates, J. Yu, Z. Zhang, S. Kresovich and M. D. McMullen (2009). "The Genetic Architecture of Maize Flowering Time." Science **325**(5941): 714-718.

Burdon, R. D. (1977). Genetic correlation as a concept for studying genotype environment interaction in forest tree breeding. Silvae genet.ica 168-175.

Busby, C. L. (1983). "Crown-quality assessment and the relative economic importance of growth and crown characters in mature loblolly pine." Proceedings of the 17th southern. Forest tree improvement conference (39): 121-130.

Chan, E. K. F., H. C. Rowe and D. J. Kliebenstein (2009). "Understanding the Evolution of Defense Metabolites in Arabidopsis thaliana Using Genome-Wide Association Mapping." Genetics: genetics.109.108522.

Choi, J., R. L. Amateis and H. E. Burkhart (2008). "Modeling trends in stem quality characteristics of loblolly pine trees in unthinned plantations." Canadian journal of forest research **38**: 1446-1457.

Condon, A. G., G. D. Farquhar, G. J. Rebetzke and R. A. Richards (2004). "Breeding for high water-use efficiency." Journal of experimental botany **55**: 2447-2460.

Craig, H. (1954). "Carbon-13 in plants and the relationship between carbon-13 and carbon-14 variations in nature." Journal of Geology **62**: 115-149.

Craig, J. E., A. W. Hewitt, A. E. McMellon, A. K. Henders, L. Ma, L. Wallace, S. Sharma, K. P. Burdon, P. M. Visscher, G. W. Montgomery and S. MacGregor (2009). "Rapid inexpensive genome-wide association using pooled whole blood." Genome Research **19**(11): 2075-2080.

Cumbie, W. P. (2002). Variation of wood density traits in rooted cuttings and seedlings of loblolly pine. Forestry. Raleigh, North Carolina State University. **Master of Science:** 96.

de Roos, A. P. W., B. J. Hayes and M. E. Goddard (2009). "Reliability of Genomic Predictions Across Multiple Populations." Genetics **183**(4): 1545-1553.

Devey, M. E., S. D. Carson, M. F. Nolan, A. C. Matheson, C. T. Riini and J. Hohepa (2004). "QTL associations for density and diameter in *Pinus radiata* and the potential for marker aided selection." Theoretical and Applied Genetics **108**: 516-524.

Domec, J.-C., S. Palmroth, E. Ward, C. A. Maier, M. ThÉRÉZien and R. Oren (2009). "Acclimation of leaf hydraulic conductance and stomatal conductance of Pinus taeda (loblolly pine) to long-term growth in elevated CO_2 (free-air CO_2 enrichment) and N-fertilization." Plant, cell and environment **32**(11): 1500-1512.

Dougherty, P. M. and C. A. Gresham (1988). "Conceptual analysis of southern pine plantation establishment and early growth." Southern journal of applied forestry **12**(3): 160-166.

Druley, T. E., F. L. M. Vallania, D. J. Wegner, K. E. Varley, O. L. Knowles, J. A. Bonds, S. W. Robison, S. W. Doniger, A. Hamvas, F. S. Cole, J. C. Fay and R. D. Mitra (2009). "Quantification of rare allelic variants from pooled genomic DNA." Nature Methods **6**(4): 263-265.

Du, L. and B. W. Poovaiah (2004). "A novel family of Ca2+ /calmodulin-binding proteins involved in transcriptional regulation: interaction with fsh/Ring3 class transcription activators." Plant Molecular Biology **54**(4): 549-569.

Dutkowski, G. W., J. Costa E Silva, A. R. Gilmour and G. A. Lopez (2002). "Spatial analysis methods for forest genetic trials." Canadian journal of forest research **32**(12): 2201-2214.

Eckert, A., J. v. Heerwaarden, J. Wegrzyn, C. D. Nelson, J. Ross-Ibarra, S. C. González-Martínez and D. Neale (2009). "Patterns of population structure and environmental associations with aridity across the range of loblolly pine (Pinus taeda L., Pinaceae)." Genetics **in press**.

Eckert, A. J., A. D. Bower, J. L. Wegrzyn, B. Pande, K. D. Jermstad, K. V. Krutovsky, J. B. St. Clair and D. B. Neale (2009). "Association Genetics of Coastal Douglas Fir (Pseudotsuga menziesii var. menziesii, Pinaceae). I. Cold-Hardiness Related Traits." Genetics **182**(4): 1289-1302.

Emebiri, L. C., M. U. Slee, A. C. Matheson and M. E. Devey (1998). "Interval mapping of quantitative trait loci affecting NESTUR, a stem growth efficiency index of radiata pine seedlings." Theoretical and applied genetics **97**(7): 1062-1068.

Emhart, V. I., D. A. Huber, T. L. White and T. A. Martin (2006). "Genetic variation in basal area increment phenology and its correlation with growth rate in loblolly and slash pine families and clones." Canadian journal of forest research **36**(4): 961-971.

Falconer, D. S. (1989). Introduction to quantitative genetics Burnt Mill, Harlow, Essex, England : Longman, Scientific & Technical ; New York : Wiley, 1989. **3rd ed**.

Farquhar, G. D., J. R. Ehleringer and K. T. Hubick (1989). "Carbon isotope discrimination and photosynthesis." Annual review of plant physiology and plant molecular biology: 503-537.

Finkelstein, R. R., M. L. Wang, T. J. Lynch, S. Rao and H. M. Goodman (1998). "The Arabidopsis Abscisic Acid Response Locus ABI4 Encodes an APETALA 2 Domain Protein." Plant Cell **10**(6): 1043-1054.

Flint-Garcia, S. A., J. M. Thornsberry and E. S. B. IV (2003). "Structure of Linkage Disequilibrium in Plants." Annual reviews in plant biology **54**: 357-374.

Fox, T. R., H. L. Allen, T. J. Albaugh, R. Rubilar and C. A. Carlson (2007). "Tree Nutrition and Forest Fertilization of Pine Plantations in the Southern United States." Southern journal of applied forestry **31**: 5-11.

Gezan, S. A., T. L. White and D. A. Huber (2006). "Achieving higher heritabilities through improved design and analysis of clonal trials [electronic resource]." Canadian journal of forest research **36**: 2148-2156.

Gilmour, A. R., R. D. Anderson and A. L. Rae (1985). "The analysis of binomial data by a generalized linear mixed model." Biometrika **72**(3): 593-599.

Gilmour, A. R., R. D. Anderson and A. L. Rae (1987). "Variance components on an underlying scale for ordered multiple threshold categorical data using a generalized linear mixed model." Journal of animal breeding and genetics **104**(3): 149-155.

Gilmour, A. R., B. J. Gogel, B. R. Cullis and R. Thompson (2006). ASReml User Guide Release 2.0. Hemel Hempstead, UK, VSN International Ltd**:** 320pp.

Gocke, M. H. (2006). Production system influences the survival and morphology of rooted stem cuttings of loblolly pine (Pinus taeda L.) and sweetgum (Liquidambar styraciflua L.). Forestry. Raleigh, North Carolina State University. **Master of Science**.

Goebal, N. B. and J. R. Warner (1966). Volume Tables for Small Diameter Loblolly, Shortlead, and Virginia Pine in the Upper South Carolina Piedmont. Clemson, SC, Clemson University.

González-Martínez, S. C., D. Huber, E. Ersoz, J. M. Davis and D. B. Neale (2008). "Association genetics in Pinus taeda L. II. Carbon isotope discrimination." Heredity **101**(1): 19-26.

González-Martínez, S. C., N. C. Wheeler, E. Ersoz, C. D. Nelson and D. B. Neale (2007). "Association Genetics in Pinus taeda L. I. Wood Property Traits." Genetics **175**(1): 399-409.

Heffner, E. L., J.-L. Jannink and M. E. Sorrells (2009). "Genomic Selection for Crop Improvement [electronic resource]." Crop science **49**(1): 1-12.

Hill, W. G., M. E. Goddard and P. M. Visscher (2008). "Data and Theory Point to Mainly Additive Genetic Variance for Complex Traits." PloS Genetics **4**(2): e1000008.

Hirschhorn, J. N. and M. J. Daly (2005). "Genome-wide association studies for common diseases and complex traits." Nature Reviews Genetics **6**(2): 95-108.

Hospital, F., A. Charcosset, A. Gallais, L. Moreau and F. Lacoudre (1997). "More on the efficiency of marker-assisted selection." Theoretical and applied genetics **95**(8): 1181-1189.

Hunt, R. and P. S. Lloyd (1987). "Growth and partitioning." New phytologist **106**(1,suppl.): 235-249.

Ingvarsson, P. K., M. V. Garcia, V. Luquez, D. Hall and S. Jansson (2008). "Nucleotide polymoirphism and phenotypic associations within and around the phytochrome B2 locus in European aspen (Populus tremula, Salicaceae)." Genetics **178**(4): 2217-2226.

Isik, F., D. D. Boos and B. Li (2005). "The distribution of genetic parameter estimates and confidence intervals from small disconnected diallels." Theoretical and applied genetics **110**: 1236-1243.

Isik, F., B. Goldfarb, J. Frampton and B. Li (2004). "Efficiency of seedlings and rooted cuttings for testing and selection in Pinus taeda." Forest science **50**(1): 44-53.

Isik, F., B. Goldfarb, X. Sun, M. Gumpertz and B. Li (2008). "Analysis of cellulose microfibril angle using a linear mixed model in Pinus taeda clones." Canadian journal of forest research **38**(6): 1676-1689.

Isik, K. and F. Isik (1999). "Genetic variation in Pinus brutia Ten. in Turkey. II. Branching and crown traits." Silvae genetica **48**(6): 293-302.

Ivkovic, M., H. X. Wu, T. A. McRae and M. B. Powell (2006). "Developing breeding objectives for radiata pine structural wood production. I. Bioeconomic model and economic weights." Canadian journal of forest research **36**: 2920-2931.

Jayawickrama, K. J. S. (2001). "Genetic parameter estimates for radiata pine in New Zealand and New South Wales: a synthesis of results." Silvae genetica **50**(2): 45-53.

Jayawickrama, K. J. S. and C. B. Low (1999). "Pinus radiata selections from different regions in New Zealand differ in branch habit, form, and growth rate." New Zealand journal of forest research **29**(1): 3-24.

Jing, H.-C., D. Kornyukhin, K. Kanyuka, S. Orford, A. Zlatska, O. P. Mitrofanova, R. Koebner and K. Hammond-Kosack (2007). "Identification of variation in adaptively important traits and genome-wide analysis of trait marker associations in Triticum monococcum." Journal of experimental botany **58**(13): 3749-3764.

Johnsen, K. H., L. B. Flanagan, D. A. Huber and J. E. Major (1999). "Genetic variation in growth, carbon isotope discrimination, and foliar N concentration in Picea mariana: analyses from a half-diallel mating design using field-grown trees." Canadian journal of forest research **29**(11): 1727-1735.

Kang, H. M., N. A. Zaitlen, C. M. Wade, A. Kirby, D. Heckerman, M. J. Daly and E. Eskin (2008). "Efficient Control of Population Structure in Model Organism Association Mapping." Genetics **178**(3): 1709-1723.

Kaya, Z., D. B. Neale and M. M. Sewell (1999). "Identification of quantitative trait loci influencing annual height- and diameter-increment growth in loblolly pine (Pinus taeda L.)." Theoretical and applied genetics **98**(3/4): 586-592.

Kirkby, E. A. and D. J. Pilbeam (1984). "Calcium as a plant nutrient." Plant, cell and environment **7**(6): 397-405.

Kumar, S. and D. J. Garrick (2001). "Genetic response to within-family selection using molecular markers in some radiata pine breeding schemes." Canadian journal of forest research **31**(5): 779-785.

Lambeth, C. (2000). "Realized genetic gains for first generation improved loblolly pine in 45 tests in coastal North Carolina." Southern journal of applied forestry. **24**(3): 140-144.

Lambeth, C. C., R. B. McCullough, S. D. Duke and J. P. v. Buijtenen (1983). "Early selection is effective in 20-year-old genetic tests of loblolly pine." Silvae genetica **32**(5/6): 210-215.

Lebude, A. V., B. Goldfarb, F. A. Blazich, F. C. Wise and J. Frampton (2004). "Mist, substrate water potential and cutting water potential influence rooting of stem cutting of loblolly pine." Tree physiology **24**: 823-831.

Lee, M.-H., C. E. Lim, S. O. Sohn, D. G. Kim, J. Lim, M. M. Lee, K. S. Chang, J.-O. Heo, S.-K. Song, B. Kim, M. Kim, S. A. Lee and N.-I. Yu (2008). "Large-scale analysis of the GRAS gene family in Arabidopsis thaliana [electronic resource]." Plant Molecular Biology **67**(6): 659-670.

Li, B., S. McKeand and R. Weir (2000). "Impact of forest genetics on sustainable forestry--results from two cycles of loblolly pine breeding in the U.S." Journal of sustainable forestry **10**(1/2): 79-85.

Li, B., S. E. McKeand and R. J. Weir (1996). Genetic parameter estimates and selection efficiency for the loblolly pine breeding in the southeastern U.S. Tree improvement for sustainable tropical forestry. Proceedings of the QFRI-IUFRO conference, Caloundra, Queensland, Australia, 27 October-1 November 1996. Volume 2., Gympie; Australia, Queensland Forestry Research Institute.

Lin, C. Y. (1978). "Index selection for genetic improvement of quantitative characters." Theoretical and applied genetics **52**(2): 49-56.

Lopes, U. V., D. A. Huber and T. L. White (2000). "Comparison of methods for prediction of genetic gain from mass selection on binary threshold traits." Silvae genetica (1): 50-56.

Lowe, W. J., T. D. Byram and F. E. Bridgwater (1999). "Selecting loblolly pine parents for seed orchards to minimize the cost of producing pulp." Forest science. **45**(2): 213-216.

Mantysaari, E. A., R. L. Quaas and Y. T. Grohn (1991). "Simulation study on covariance component estimation for two binary traits in an underlying continuous scale." Journal of dairy science **74**(2): 580-591.

Martin, T. A., S. E. McKeand and P. M. Dougherty (2005). "Strategies and case studies for incorporating ecophysiology into Southern pine tree improvement programs." Southern journal of applied forestry **29**(2): 70-79.

Martinez-Vilalta, J., A. Sala and J. Pinol (2004). "The hydraulic architecture of Pinaceae - a review." Plant ecology **171**: 3-13.

McKeand, S. and J. Svensson (1997). "Sustainable management of genetic resources." Journal of forestry **95**(3): 4-9.

McKeand, S., T. White, T. Byram and T. Mullin (2003). "Deployment of genetically improves loblolly and slash pine in the South." Journal of forestry **101**(3): 32-37.

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. 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**: 352-358.

McKeand, S. E. and F. E. Bridgwater (1998). "A strategy for the third breeding cycle of loblolly pine in the Southeastern US." Silvae genetica **47**(4): 223-234.

McKeand, S. E., E. J. Jokela, D. A. Huber, T. D. Byram, H. L. Allen, B. Li and T. J. Mullin (2006). "Performance of improved genotypes of loblolly pine across different soils, climates, and silvicultural inputs [electronic resource]." Forest ecology and management **227**: 178-184.

McKeand, S. E., B. Li, J. E. Grissom, F. Isik and K. J. S. Jayawickrama (2008). "Genetic parameter estimates for growth traits from diallel tests of loblolly pine throughout the southeastern United States." Silvae genetica **57**(3): 101-110.

Meuwissen, T. H. E., M. E. Goddard and B. J. Hayes (2001). "Prediction of total genetic value using genome-wide dense marker maps." Genetics **157**(4): 1819-1829.

Morris, E. R. and J. C. Walker (2003). "Receptor-like protein kinases: the keys to response." Current opinion in plant biology **6**(4): 339-342.

Morse, A. M., D. G. Peterson, M. N. Islam-Faridi, K. E. Smith, Z. Magbanua, S. A. Garcia, T. L. Kubisiak, H. V. Amerson, J. E. Carlson, C. D. Nelson and J. M. Davis (2009). "Evolution of Genome Size and Complexity in *Pinus*." PLoS ONE **4**(2): e4332.

Neale, D. B. and P. r. K. Ingvarsson (2008). "Population, quantitative and comparative genomics of adaptation in forest trees." Current opinion in plant biology **11**: 149-155.

Neale, D. B. and O. Savolainen (2004). "Association genetics of complex traits in conifers." Trends in plant sciences **9**(7): 325-330.

Nelson, C. D. and K. H. Johnsen (2008). "Genomic and physiological approaches to advancing forest tree improvement." Tree physiology **28**(7): 1135-1143.

Patharkar, O. R. and J. C. Cushman (2006). "A novel coiled-coil protein co-localizes and interacts with a calcium-dependent protein kinase in the common ice plant during low-humidity stress [electronic resource]." Planta **225**(1): 57-73.

Prasolova, M. V., Z. H. Xu, G. D. Farquhar, P. G. Saffigna and M. J. Dieters (2000). "Variation in branchlet delta13C in relation to branchlet nitrogen concentration and growth in 8-year-old hoop pine families (Araucaria cunninghamii) in subtropical Australia." Tree physiology(15): 1049-1055.

Prasolova, N. V., N. V. Prasolova, K. Lundkvist and Z. H. Xu (2005). "Genetic variation in foliar nutrient concentration in relation to foliar carbon isotope composition and tree growth with clones of the F1 hybrid between slash pine and Caribbean pine." Forest ecology and management **210,**: 173-191.

Prestemon, J. and R. C. Abt (2002). Timber products supply and demand. The southern forest assessment final technical report. Asheville, NC, USDA

RDevelopmentCoreTeam (2005). R: A language and environment for statistical computing, reference index version 2.x.x. . R. F. f. S. Computing. Vienna, Austria.

Rebetzke, G. J., R. A. Richards, A. G. Condon and G. D. Farquhar (2006). "Inheritance of Carbon Isotope Discrimination in Bread Wheat (Triticum Aestivum L.)." Euphytica **150**: 97-106.

SAS (1989). SAS/STAT User's Guide, Version 6. Cary, NC, SAS Institute Inc.

Schall, R. (1991). "Estimation in generalized linear models with random effects." Biometrika **78**(4): 719-727.

Sierra-Lucero, V., S. E. McKeand, D. A. Huber, D. L. Rockwood and T. L. White (2002). "Performance differences and genetic parameters for four coastal provenances of loblolly pine in the Southeastern United States." Forest science. **48**(4): 732-742.

Smith, J. G. and C. Newton-Cheh (2009). "Genome-wide association study in humans." Methods in Molecular Biology (Totowa, NJ, United States) **573**: 231-58.

Stanturf, J. A., S. B. Jones, F. S. Broerman and R. C. Kellison (2003). "Productivity of the southern pine plantations: where are we and how did we get here?" Journal of forestry **101**(3): 26-31.

Stich, B., J. Mohring, H.-P. Piepho, M. Heckenberger, E. S. Buckler and A. E. Melchinger (2008). "Comparison of Mixed-Model Approaches for Association Mapping." Genetics **178**(3): 1745-1754.

Stich, B., H.-P. Piepho, B. Schulz and A. Melchinger (2008). "Multi-trait association mapping in sugar beet (Beta vulgaris L.)." Theoretical and applied genetics **117**(6): 947-54.

Storey, J. D. (2003). "The positive false discovery rate: A Bayesian interpretation and the q-value." Annals of Statistics **31**(6): 2013-2035.

Strauss, S. H., G. Namkoong and R. Lande (1992). "Limitations of molecular-marker-aided selection in forest tree breeding." Canadian journal of forest research **22**(7): 1050-1061.

Tang, W., R. Newton, C. Li and T. Charles (2007). "Enhanced stress tolerance in transgenic pine expressing the pepper CaPF1 gene is associated with the polyamine biosynthesis." Plant cell reports **26**(1): 115-124.

"Timber Mart-South."(2007). Retrieved June 1, 2007, from www.TimberMart-South.com/tmart/.

Xiong, J. S., F. Isik, S. McKeand and R. Whetten (2009). "Genetic Variation of Stem Forking in Loblolly Pine." Submitted to Forest Science.

Xu, S. (2003). "Estimating Polygenic Effects Using Markers of the Entire Genome." Genetics **163**(2): 789-801.

Xu, S. (2003). "Theoretical basis of the Beavis effect." Genetics **165**(4): 2259-68.

Yamada, Y. (1962). "Genotype by environment interaction and genetic correlation of same trait under different environments." Japanese journal of genetics **37**(6): 498-509.

Yu, J., G. Pressoir, W. H. Briggs, I. Vroh Bi, M. Yamasaki, J. F. Doebley, M. D. McMullen, B. S. Gaut, D. M. Nielsen, J. B. Holland, S. Kresovich and E. S. Buckler (2006). "A unified mixed-model method for association mapping that accounts for multiple levels of relatedness." Nature genetics **38**(2): 203-8.