1994

Molecular Marker Linkage Mapping in Southern Pine (Longleaf Pine and Slash Pine).

Thomas Lester Kubisiak
Louisiana State University and Agricultural & Mechanical College

MOLECULAR MARKER LINKAGE MAPPING IN SOUTHERN PINE (LONGLEAF PINE AND SLASH PINE)

A Dissertation

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

in

The School of Forestry, Wildlife, and Fisheries

by
Thomas Lester Kubisiak
B.S., Michigan State University, 1988
M.S., University of Minnesota, 1990
December 1994

ACKNOWLEDGMENTS

I would especially like to thank my mother, Elizabeth Anne Kubisiak, whose understanding and help allowed me to pursue an education in the natural resources. Her invaluable financial and mental support allowed my early interests to germinate, take root, and grow into a passion for forestry.

I would also like to thank the members of my graduate committee, Professors Michael Stine, C. Dana Nelson, James H. Oard, Robert J. Tempelman, Quang V. Cao, and Manjit S. Kang for their guidance and ever-stimulating discussion concerning the use of molecular markers for genetic mapping, quantitative trait locus searching, and potential application for marker-aided selection.

A special thanks to Michael Stine, C. Dana Nelson, Warren L. Nance and Ronald C. Schmidtling for sharing their invaluable knowledge regarding the genetics of southern pine species, as well as to the research staff at the USDA Forest Service, Southern Institute of Forest Genetics (specifically, Lori Nelson, Melissa Crawford, and Glen Johnson) for all their technical assistance. A special thanks to Mary Bowen for all of her help, which is too numerous to mention succinctly!

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

The goal of this work was to develop molecular markers for use in a backcross breeding program to speed the introgression of genes influencing rapid early height growth (EHG) from slash pine (*Pinus elliottii* Engelm. var. *elliottii*) into longleaf pine (*Pinus palustris* Mill.). The efficacy of molecular markers for genetic mapping in the *Pinaceae* was determined in segregating haploid and diploid populations. Initial screening for genetic polymorphisms was conducted using the random amplified polymorphic DNA (RAPD) technique. Using DNAs obtained from haploid megagametophytes, a RAPD-based genetic map for longleaf pine clone 3-356 [16 linkage groups and 6 pairs (133 markers) covering 1,635 cM] was constructed. Concern regarding the efficacy of RAPD data lead to a series of computer simulations investigating the effects of missing and mis-scored data on linkage group construction. Given the parameters investigated, levels as high as 15% missing data and 2% mis-scored data still provided accurate low- to medium-density map construction. Individual parental maps were constructed with F_1 progeny from a slash pine H-28 (♂) × longleaf pine 3-356 (♀) cross. The longleaf pine 3-356 map consisted of 18 groups and 3 pairs (122 markers) covering 1367.5 cM, and the slash pine H-28 map 13 groups and 6 pairs (91 markers) covering 952.9 cM. Orders and distances of loci in common between the two maps constructed for longleaf pine 3-356 were compared. Orders were found to be conserved for those groups containing three or more loci. However, genetic distance estimates varied considerably, but not in any systematic manner. RAPD and allozyme loci identified as being heterozygous

in both parents were utilized to combine the parent-specific maps constructed for the slash pine × longleaf pine cross. Five RAPD loci and one allozyme locus suggested homology between the otherwise parent-specific linkage groups. Substantial phenotypic variation for EHG was observed in the F_1 population, therefore the parent-specific markers and maps were used to localize putative EHG QTL. Using simultaneous marker models (multiple regression), marker loci were found to be significantly associated with QTL influencing hypocotyl length, total height, brown spot resistance and root collar diameter.

CHAPTER 1

INTRODUCTION

Tree improvement programs have made slow progress due to several characteristics of the growth and development of trees. In contrast to our more-familiar crops where two to three generations of selections may be made in a single year, trees take 3-10 years per cycle and the most advanced tree breeding programs are only in their first few generations of selection. Also, little is known about the mode of inheritance for most traits of interest in tree species. This is additionally complicated by changes that occur during phase change (maturation). In addition, traits of economic interest show variable heritabilities as the tree matures. Methods to identify juvenile planting stock capable of superior rotation-age (adult) character expression has been a principal focus for tree breeders, and the use of polymorphic molecular markers to aid in early selection has been a subject of interest for many years.

Molecular Markers

Advances in methods for assaying polymorphisms [e.g. protein variants, restriction fragment length polymorphisms (RFLPs) and more recently random amplified polymorphic DNAs (RAPDs)] have made it possible to greatly increase the density of genetic linkage maps of many plant species. These maps have changed quantitative genetics by creating a base necessary for mapping genes underlying quantitative traits (Beckmann and Soller 1986; Jensen 1989; Lander and Botstein 1989; Simpson 1989; Knapp et al. 1990; Knapp 1991; Weller 1992; Jansen and Stam

1994; Zeng 1994). Classical quantitative genetic methods are useful for making inferences about population parameters (e.g. genetic variances and heritabilities). Hypotheses tested by these methods describe the characteristics of a population, but not of the specific genes influencing the trait of interest. In other words, classical methods do not lead to an understanding of the location and effects of genes underlying quantitative traits, which is essential for implementing marker-aided selection (MAS) (Tanksley et al. 1989; Young and Tanksley 1989; Lande and Thompson 1990).

One marker system that has contributed greatly to research on the genetics of tree species has been various protein-based markers (specifically, allozymes). Allozymes are allelic variants of an enzyme that are transcribed from a single genetic locus. These alternative enzymatic alleles can be identified by staining for enzyme activity in tissue extracts subjected to starch gel electrophoresis (reviews of protocols are available in Shaw and Prasad 1970; Conkle et al. 1982). Polymorphisms are visualized as differences in banding patterns between or among two or more individuals.

Allozyme markers have been used extensively to investigate genetic variability within and among tree species (Hamrick et al. 1992; Loveless 1992; Moran 1992; Muller-Starck et al. 1992), to examine mating systems (Mitton 1992), gene dispersal (Adams 1992; Ellstrand 1992), genetic structure within populations (Epperson 1992), to investigate associations between levels of heterozygosity and growth factors (Bush and Smouse 1992) and to examine associations among markers and quantitative trait

variation (El-Kassaby and Sziklai 1982; Furnier et al. 1991). The inheritance and linkage of allozymes has been described for numerous pine species (Guries et al. 1978; Rudin and Ekberg 1978; Adams and Joly 1980; Conkle 1981; King and Dancik 1983; Cheliak and Pitel 1985; Furnier et al. 1986; Harry 1986; Strauss and Conkle 1986; Muona et al. 1987; Niebling et al. 1987; Perry and Knowles 1989; Adams et al. 1990; Xie et al. 1991). However, the paucity of available isozyme markers and their low level of polymorphism have limited the scope of such analyses.

A rather recent technique, two-dimensional (2-D) electrophoresis and staining of total proteins has significantly increased the number of protein polymorphisms that can be identified (Anderson et al. 1985; Bahrman and Damerval 1989; Gerber et al. 1993), however their numbers are still limited. Additional problems such as poorly understood modes of inheritance and developmental instability (Bahrman and Damerval 1989) would limit the practical use of total protein markers for linkage map construction, QTL detection, and subsequent MAS.

The development and application of new DNA-based marker techniques have now made it feasible to conduct large-scale genetic mapping projects (Lalouel 1992). One commonly used technique to collect marker data for such applications has been RFLP analysis. RFLPs are a class of mutation that reflect DNA sequence variation (Beckmann and Soller 1983). They are generated by restriction endonucleases of bacterial origin that recognize and cut DNA at specific base sequences. RFLPs can result from: changes in the DNA sequence at the restriction site, insertions or deletions of DNA within a region bordered by two restriction sites, or by

rearrangements in DNA (inversions and/or translocations). They are identified by hybridizing labeled single-stranded probes to membranes containing immobilized, restricted, denatured genomic DNA samples. Polymorphisms are visualized as differences in banding patterns between or among two or more individuals by autoradiography (Figure 1.1). RFLPs have numerous advantages over morphological mutants or protein markers for genetic linkage analyses as they are potentially unlimited in number, can be detected in coding as well as non-coding DNA, are detectable in most tissue types and developmental stages, and are usually codominant genetic markers.

RFLPs have recently been used to construct moderate- to highly-saturated genetic linkage maps for a number of annual crop species such as tomato (*Lycopersicon esculentum*) [Bernatzky and Tanksley 1986], maize (*Zea mays*) [Helentjaris et al. 1986], lettuce (*Lactuca sativa*) [Landry et al. 1987]; rice (*Oryza sativa*) [McCouch et al. 1988], potato (*Solanum tuberosum*) [Gebhardt et al. 1989], and soybean (*Glycine max*) [Keim et al. 1990], as well as for various other species (O'Brien 1993). Despite the fact that genetic marker mapping and quantitative trait loci (QTL) searching strategies are still being developed, mapped marker information is successfully being used to construct saturated maps and identify QTL (Paterson et al. 1988; Keim et al. 1990; Stuber 1992; Goldman et al. 1993; Komatsuda et al. 1993; Mansur et al. 1993; Leonards-Schippers et al. 1994; Wang et al. 1994; and many others).

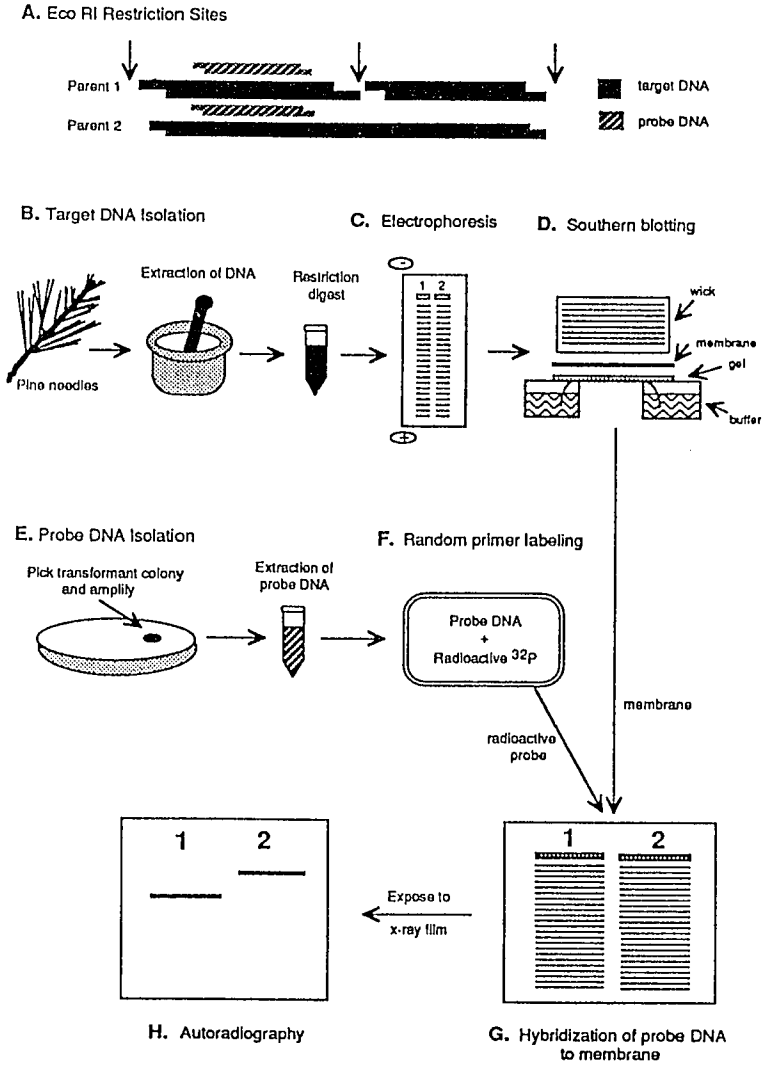

Figure 1.1. Identification of RFLPs[1]. A. RFLP generated by restriction enzyme *Eco*RI. Parent #1 contains a restriction site lacking in parent #2. Methods to identify this RFLP outlined in steps B - H. Target DNA purified from pine needles (B) and cut with restriction enzyme. Electrophoresis (C) of DNA through an agarose gel separates the DNA fragments based on size (largest fragments remaining near the origin (-), smallest moving further (+). DNA is transferred to support membrane via Southern blotting (D). Probe DNA is cloned into bacterial vector and purified (E). Addition of radioactive ^{32}P to the probe by random primer labeling (F) allows for detection of the probe DNA. Probe DNA is hybridized to the membrane (G) binding to the target DNA that it matches. The probe binding site is identified by autoradiography (H) when the radioactive probe DNA exposes x-ray film.

---

[1]Used by permission of Dr. Michael Stine and Mary Bowen from Louisiana Education Quality Support Fund Proposal: "Restriction fragment length mapping of longleaf pine hybrids".

The use of RFLPs to map conifer genomes has recently been proposed (Nance and Nelson 1989; Neale and Williams 1991), and is currently underway for loblolly pine (*Pinus taeda* L.) (Devey et al. 1991). Although RFLPs have been extensively and successfully used in a wide array of genetic studies, the technique suffers from some limitations inherent in the process. The procedure requires well-trained personnel, expensive equipment, and numerous steps (Figure 1.1) which are fairly time consuming. Several additional factors have delayed the widespread use of RFLPs in genetic linkage studies involving pines. Pine species have been found to possess rather large physical genomes (2C = 33 to 57 pg, Ohri and Khoshoo 1986) complicating the autoradiographic detection of single-copy genes (Ausubel et al. 1987; Weeden 1994). Exposure periods of one to two weeks are not uncommon (Devey et al. 1991). Due to the long generation times inherent with perennial tree species there is a lack of well-defined pedigrees (beyond two generations) with which to work. This can often confound the correct determination of allelic phase for some parental gene combinations (Ott 1991).

Possibly as a result of such complications, only a limited number of tree species have been mapped using RFLPs. Two RFLP-based genetic maps for citrus (*Citrus spp.*) (Durham et al. 1992; Jarrell et al. 1992), one for apple (*Malus domestica*) (Weeden and Hemmat 1993), and one for quaking aspen (*Populus tremuloides* Michx.) (Liu and Furnier 1993) have recently been published. In comparison to pine species, these species have relatively small genomes (for citrus 2C = 1.24 pg (Guerra 1984); and for aspen 2C = 1.6 pg (Dhillon et al. 1984)).

Quaking aspen also has the added advantage of early floral development [minimum seed-bearing age ~ 5 years] (USDA 1974), facilitating pedigree advancement.

A technique applying the polymerase chain reaction (PCR) amplification of template DNA by short primers of arbitrary sequence has recently been used to identify polymorphic markers (Welsch et al. 1990; Williams et al. 1990; Caetano-Anolles et al. 1991). These markers have most commonly been referred to as random amplified polymorphic DNA (RAPD) markers (Williams et al. 1990). Template DNA, in the presence of a heat stable DNA polymerase, and an arbitrarily sequenced primer are subjected to repeated rounds of denaturation, annealing, and extension (see Figure 1.2 I., and 1.2 II.), resulting in the exponential increase of various DNA fragments which can be visualized after electrophoretic separation in agarose gels and staining with ethidium bromide.

Theoretically, any short sequence of nucleotides is likely to find a number of complementary sites in a eukaryotic genome. When two sites of complementarity are nearby (usually within 2000 bp, depending upon extension time) in opposite orientation on both sense and anti-sense template DNA strands, PCR can amplify the DNA between the sites. If one or more target sites are absent (presumably due to base pair mutation), a null phenotype is indicated by the absence of a band. Polymorphisms can be identified as DNA fragments that amplify in one individual and not another. The majority of RAPDs identified are inherited as dominant Mendelian markers (i.e. it is not possible to distinguish between a heterozygous band-present individual from a homozygous band-present individual).

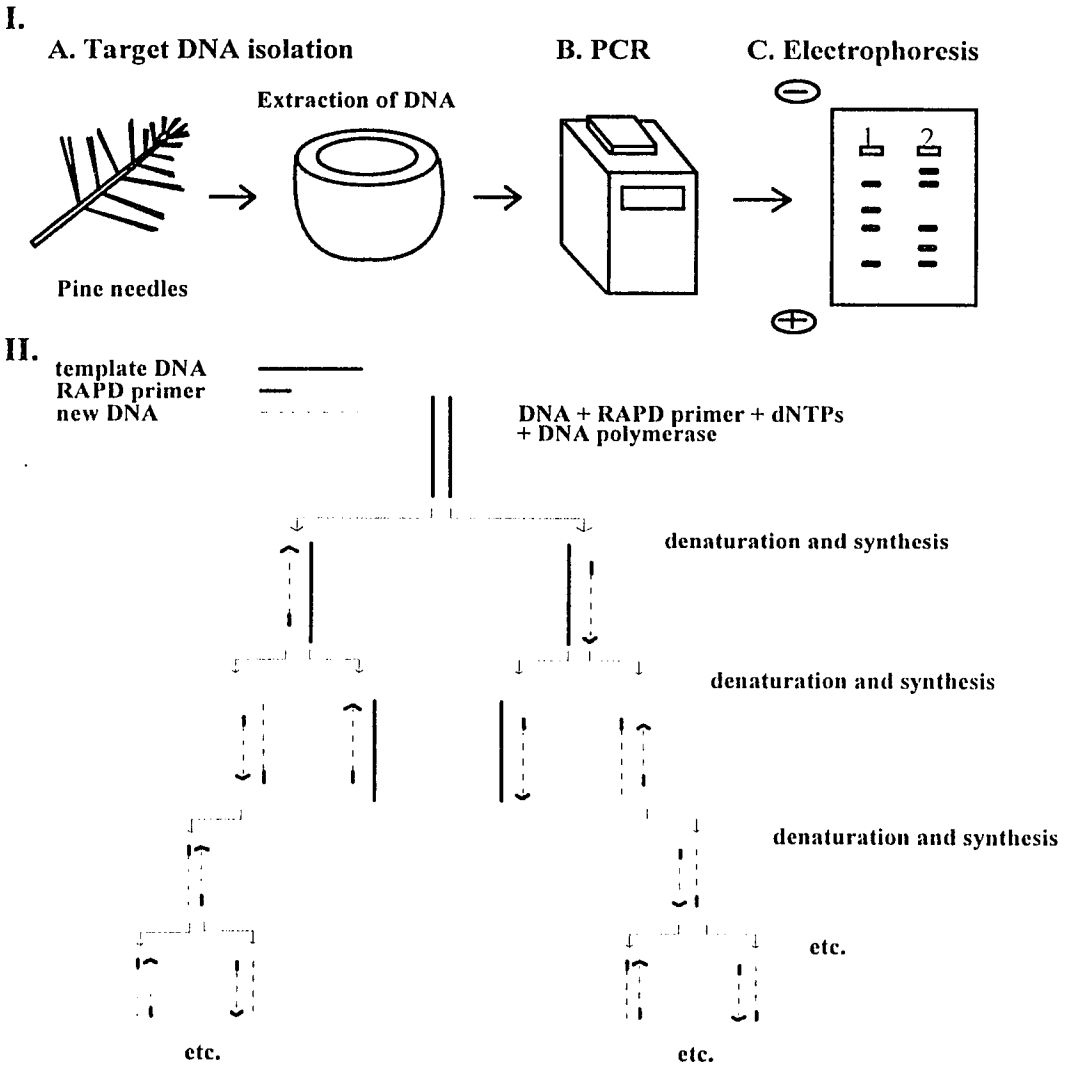

Figure 1.2 I. Identification of RAPDs. Methods to identify RAPDs are outlined in steps a - c. The template DNA is extracted from needles or megagametophytes (a). An appropriate amount of template DNA is amplified in the presence of an oligonucleotide (10 base pair) primer via the polymerase chain reaction (PCR) technique (b). After amplification, the reaction products are separated in an agarose gel stained with ethidium bromide and visualized under UV light (c). 1.2 II[2]. PCR amplification of template DNA.

[2]Modified from Ausubel et al. 1987.

Despite their dominant inheritance, RAPDs share many of the same benefits associated with RFLPs. They have turned out to be virtually unlimited in number and can be detected in coding as well as non-coding DNA. RAPD analysis has a number of advantages over RFLP analysis. RAPDs require small amounts of template DNA, bypass the need for the Southern transfer of DNA, the handling of radioactive isotope, and have a significantly increased throughput due to their potential for automation (Grattapaglia et al. 1992). RAPDs are of considerable interest as the procedure is far less complex than the RFLP procedure and can be developed and applied with minimal cost of equipment and labor, which is often a major concern to forestry organizations (public or private).

Mapping Theory

The large number of polymorphisms that can be assayed (far greater than the haploid number of chromosomes) means that many of these polymorphisms must exist on the same chromosome. The idea that chromosomes behave as units of segregation in meiosis led to the expectation that the alleles of all genes located on the same chromosome should be transmitted as a unit and therefore show linkage (Sturtevant 1913). Linkage between two genes on the same chromosome is usually never complete, resulting from the exchange of DNA between homologous chromosomes. The occurrence of such an exchange event between homologous chromosomes, called cross-over, results in recombination between gene pairs on chromosomes. The frequency of cross-over between any two genes serves as a measure of genetic distance between them (Haldane 1919; Kosambi 1944). Thus, the

measure of recombination between different genes on a chromosome provides the basis for construction of a genetic map of the chromosome.

In linkage analyses the main problem is parameter estimation, specifically the estimation of recombination frequencies between markers. The hypothesis to be tested is that of free recombination (independent assortment $\theta = 50\%$, or linkage $\theta < 50\%$). One of the simplest means of performing such a test has been referred to as the direct approach (Ott 1985). This simply entails counting recombinants and non-recombinants and performing statistical tests (for example a χ^2 test). The smaller the recombination frequency the tighter the linkage. A number of other likelihood-based approaches have been taken to detect linkage between markers such as Bernstein's *Y* statistics, the sib-pair method, and Fisher's *U* statistics (briefly discussed in Ott 1985). However, most of these approaches do not take into account full use of the available data, and uninformative mating configurations only add complications such as loss of data.

A commonly used approach towards linkage analysis developed from work conducted by Haldane and Smith (1947) has been the likelihood approach. Morton (1955) developed a sequential test procedure for linkage analyses, since he was typically interested in detecting linkage in several human pedigrees (families) each with a limited number of offspring. This approach, however, is not applicable to larger family pedigrees (Ott 1985). Elston and Stewart (1971) proposed an iterative algorithm widely used in statistics, and very productively in linkage analyses. This approach allows the representation of the likelihood, or log of the odds (LOD) score,

for pedigrees of many complexities. In addition to being able to handle both qualitative (categorical) as well as quantitative data it also easily deals with missing data. One downfall of this approach is that it is not very well suited to handling the number of loci involved with constructing saturated genetic maps, as computation time increases exponentially with the number of markers added. To handle some of the problems inherent with the analysis of large data sets and often complex pedigrees was the development of the Lander and Green algorithm (Lander and Green 1987). The advantage of this algorithm is that its computation time increases linearly instead of exponentially as markers are added. This fact alone has made it very desirable for genomic mapping studies.

The idea of using qualitative markers to search for genes responsible for quantitative trait variation goes back to Sax (1923). Since this time, three principal methods have been developed and applied to detect and localize such variation; least squares estimation, method of moment estimation, and maximum likelihood estimation. Although standard linear model methodology has the desirable properties of minimum residual variance and unbiasedness (Montgomery 1984; Weisberg, 1985) this methodology has several disadvantages over the latter two methods. Hypothesis testing using linear model analysis is based on assumptions which do not hold under certain circumstances (dependent variable is non-normally distributed, residuals are non-normally distributed or have unequal variances). Also, it is not possible to estimate the genetic distance between a single marker and QTL, therefore it is not possible to distinguish between the tight linkage of a small effect QTL or the loose

linkage of a major effect QTL. To overcome these problems two methods have been effectively applied; method of moment estimation (Darvasi and Weller 1992) and maximum likelihood estimation (Luo and Kearsey 1989).

The main concept behind the method of moments is to take repeated product sums of various defined equations (Darvasi and Weller 1992; Weller 1992). These equations estimate the QTL variable mean, variance, skewness (asymmetry), and kurtosis. The method of moments technique is non-parametric; it is not dependent on the nature of the distribution of the variable. This non-parametric approach should be used given that the assumptions of the parametric method (normality of the QTL distribution or residuals) are clearly violated. The principle behind mapping a QTL by the method of moments is that incomplete linkage will result in a skewed distribution of the individual marker-genotype samples. Since skewness is measured by the third central moment, a difference in skewness can be used to estimate linkage distance between the QTL and the genetic marker. One problem encountered in this approach is that estimates often fall out of the parameter space (e.g. negative recombination estimates or variances) (Weller 1992).

The concept behind the maximum likelihood method is to find the parameter estimates that best fit the data (Weller 1986). It differs from both of the previously mentioned methods. Compared to least square estimates, estimates may be biased, and it differs from the moment of methods in that the distribution of the data needs to be known (it is parametric). Unfortunately, the introduction of bias means that the estimates do not necessarily yield least squares estimates, however, maximum

likelihood is able to take advantage of all the data and can be applied in both linear and non-linear situations with most estimates falling within the parameter space. Maximum likelihood is therefore very applicable in many situations (Ott 1991).

The ability to generate genetic linkage maps is dependent upon the levels of heterozygosity detected in the particular pedigree chosen (constructed). Levels of heterozygosity in tree species are known to vary greatly both within and among populations (Hamrick et al. 1992). Large variation in the levels of heterozygosity among pine species has been documented [e.g. low levels have been recorded for red pine (Fowler and Morris 1977) and torrey pine (Ledig and Conkle 1983), whereas much higher levels have been recorded for other pine species (Conkle 1992)]. In general, Hamrick et al. (1992) found gymnosperms to contain rather high percentages of polymorphic loci, P_s levels approaching 71% (P_s = number of polymorphic loci in at least one population divided by the total number of loci analyzed).

The proper choice of parents and development of informative pedigrees is also critical in terms of QTL searching and overall marker-aided improvement of tree species. Most current QTL mapping theory is based on the inbred-line approach where each parent is inbred and hence homozygous at all loci. Parents at the phenotypic extremes of a particular trait of interest are chosen for mapping. One parent is assumed to have all "positive effect alleles" for the trait and the other parent all "negative effect alleles". For traits of interest displaying substantial variation, selection of parents at the phenotypic extremes would create large disequilibrium

during hybridization maximizing segregation in subsequent generations (F_2 or BC_1), greatly aiding efforts to map QTL (Simpson 1989; Lande and Thompson 1990).

Genetic marker and QTL mapping theory in outcrossed tree species is not as simple as for the inbred-line approach. Inbred populations or even three generation pedigrees are generally not available and are difficult to obtain due to a significant genetic load and time constraints (Zobel and Talbert 1984). Available pedigrees for the majority of tree species generally involve maternal half-sib families, or two parents and their full-sibs. To incorporate MAS strategies into forest tree breeding it is imperative to explore alternative approaches for the construction of linkage maps that make use of pedigree structures already existing and commonly generated in tree breeding programs. Breeding programs of pine species are based on simple or recurrent selection, interspecific hybridization, and clonal selection. Linkage disequilibrium generated by hybridization coupled with the possibility of capturing non additive genetic variance through clonal propagation are conditions that would greatly enhance the potential use of MAS strategies in tree species.

Species Selection

Longleaf pine used to be a major component of the forests of the lower south-eastern coastal plain, it is now only a minor timber species (Schmidtling and White 1989). In the 8 states where longleaf pine presently occurs, only about 3% of the total timberland area supports a longleaf pine forest type (Kelly and Bechtold, 1989). There is also an unfavorable outlook for the remaining resource as stands are

maturing and young stands are not replacing those being harvested (Donner and Hines 1987).

Longleaf pine possesses many desirable qualities such as a straight, clear bole, high naval stores content, moderate to high wood specific gravity, and fusiform rust resistance (USDA 1965). Despite these qualities, its often poor survival rate following planting and extended phase of juvenile development referred to as the "grass stage" have limited its use in artificial regeneration programs (Schmidtling and White 1989).

Longleaf pine seedlings characteristically do not initiate height growth until their second year at the earliest, and depending upon growing conditions may not for 20 years or longer (USDA 1965). During this period seedlings develop extensive root systems, and apical meristems increase in diameter but not in length (Harlow and Harrar 1978). The resulting appearance is that of a mound of grass. The grass stage has been found to be under both genetic and environmental control, with all longleaf pine manifesting this trait (Hare 1984; Layton and Goddard 1982).

The physiological processes involved with grass stage regulation are poorly understood. However, early height growth is known to be closely tied to photosynthetic capacity as artificial defoliation due to brown spot infection, artificial defoliation, and shading studies have all been shown to reduce height growth (USDA 1965). The unpredictability of the grass stage greatly reduces the popularity of managing for longleaf pine. Under intensive short rotation systems, the length of the grass stage can have profound impacts on the economics of growing longleaf pine.

Breeding objectives for longleaf pine have primarily concentrated on brown-spot needle blight resistance, and early height growth (reviewed in Schmidtling and White 1989). Significant variation among and within families for these traits has been reported (Snyder and Derr 1972; Wells and Snyder 1976; Byram and Lowe 1985). Despite this variation, improvement efforts have been limited when compared to those in other southern pine species. For example, as of 1989 only 545 acres of longleaf pine seed orchards had been established compared to over 5,000 acres for loblolly pine and 3,000 acres for slash pine (USDA 1982; Schmidtling and White 1989).

Interspecific hybrids show promise for addressing the problem of delayed height growth in longleaf pine. Intermediate height growth has been observed in various families of longleaf pine × loblolly pine or slash pine (Brown 1964; Derr 1966). Variation in early height growth (EHG) was also found to be significant among and within families in several field tests of longleaf pine × slash pine hybrids (Derr 1966; Derr 1969). Analysis of F_2 hybrids of longleaf pine and loblolly pine yields an estimate of at least 10 loci controlling the grass stage trait (Brown 1964; Wright [in Castle 1921]). More recent work (C. D. Nelson unpublished data) with F_2 and BC_1 hybrids indicates that the minimum number of loci maybe fewer than 5 (using methods in Lande 1981; Cockerham 1986). However, recent theoretical work (Zeng et al. 1990; Zeng 1992) shows that violating any of the several assumptions implicit in the employed analytical methodology can result in large underestimates of the minimum number. Thus, estimates calculated to date may be considerably low,

but they do suggest that the grass stage character is a quantitative trait controlled by a finite number of genes (oligogenic vs polygenic). Numerous QTL mapping studies have found various traits to be controlled by genes with both large and small effects on the phenotype (Paterson et al. 1988; Keim et al. 1990; Stuber 1992; Goldman et al. 1993; Komatsuda et al. 1993; Mansur et al. 1993; Wang et al. 1994; Leonards-Schippers et al. 1994; and many others). The possibility that EHG is also influenced by genes with large effects on the phenotype will greatly facilitate their detection and localization. Longleaf pine and slash pine exhibit phenotypic extremes for EHG, therefore large disequilibrium should be generated during hybridization and substantial phenotypic variance should be segregating in either F_2 or BC_1 generations (Lande and Thompson 1990).

RAPD markers will eventually be scored in two backcross populations, one with longleaf pine (BC_L) as the recurrent parent and one with slash pine (BC_S) as the recurrent parent. Use of the same hybrid F_1 male parent to construct these populations should provide an opportunity to study the segregation of both "positive effect" EHG alleles from slash pine and "negative effect" EHG (grass stage) alleles from longleaf pine in the most comparable genetic backgrounds. With this approach a more complete understanding of the genetic control and inheritance of EHG in both longleaf pine and slash pine will be obtained.

The main objective of this dissertation was to develop, and determine the efficacy of molecular markers for use in genetic linkage mapping in both a maternal half-sib longleaf pine family and a full-sib slash pine × longleaf pine family. The

first matters to be addressed were: (1) could the RAPD technique be used to amplify various southern pine DNAs, (2) if so, what were the optimal reactions conditions required for reproducible amplification, (3) how much interspecific variation could the RAPD technique detect, and (4) how much variation was the RAPD technique able to detect within a "typical" southern pine (CHAPTER 2). Once these matters were addressed, the RAPD technique was used to construct a genetic linkage map for longleaf pine based on the inheritance of fragments in a haploid population of megagametophytes (Nelson et al. in press). Concern regarding the efficacy of this data lead to a series of computer simulations investigating the effect of missing and mis-scored data on linkage group construction (CHAPTER 3). Interest in expanding the RAPD technique to diploid tissues lead to the construction of individual parental maps from a slash pine H-28 (♂) × longleaf pine 3-356 (♀) cross (CHAPTER 4). Subsequently, a comparison of loci (orders and distances) common to both of the maps constructed for longleaf pine 3-356 was conducted (CHAPTER 5). Codominant allozyme markers were employed in an attempt to combine the parent-specific maps constructed from the slash pine × longleaf pine cross (CHAPTER 6). Finally, as substantial phenotypic variation for EHG was observed in the F_1 population, the parent-specific linkage maps were used to localize putative EHG QTL (CHAPTER 7).

CHAPTER 2

RAPD OPTIMIZATION, PARENTAL SCREENING AND LEVEL OF HETEROZYGOSITY

Introduction

The success and efficiency of genetic linkage mapping and quantitative trait loci (QTL) searching studies can be markedly increased by careful consideration and utilization of proper pedigrees, focusing on traits amenable to marker-aided improvement, and properly ascertaining the potential of the marker systems currently available to quickly furnish reliable data. Due to the extensive number of marker assays required for genetic mapping and quantitative trait locus (QTL) searching studies, the proper choice of a marker system that will allow for speedy and efficient collection of reliable data is paramount.

Recently, random amplified polymorphic DNA (RAPD) markers have shown tremendous potential as genetic markers for genome mapping (Williams et al. 1990). RAPDs possess a number of advantages over restriction fragment length polymorphisms (RFLPs), most particularly the technical simplicity of the assay, the lower cost per data point, and the generally higher level of polymorphism resolved. For these reasons RAPD markers appear to be more applicable to plant breeding programs than RFLPs. Despite its conceptual simplicity (see Figure 1.2 II. in CHAPTER 1), the kinetics of the RAPD reaction are quite complex. Template DNA purity and concentration, Mg^{+2} concentration, annealing temperature, degree of sequence similarity at priming sites, and primer competition all can affect the

successful and reproducible amplification of RAPD markers (Weeden et al. 1992; Williams et al. 1993).

To determine the efficacy of the RAPD assay for genetic linkage mapping in the *Pinaceae*, several matters were initially addressed. First, the quality (purity) of pine DNAs was assessed. The quality of template DNA has a great affect on the generation and resolution of amplified products. DNAs isolated with significant quantities of polysaccharides (e.g. DNA from cotton) or phenolics (e.g. DNA from pecan) do not make suitable template, and DNAs containing moderate levels of impurities often produce blurred or faint RAPD phenotypes. Secondly, the optimal reaction conditions needed for producing complex yet reproducible amplification of pine template DNAs was determined. Of primary concern were the proper template DNA and Mg^{+2} concentrations. Thirdly, the level of interspecific variation resolved by RAPD primers was investigated. Since the long-term goal of this research is to speed the introgression of genes influencing rapid early height growth from slash pine or loblolly pine into longleaf pine, the level of interspecific variation was of primary interest. Finally, preliminary screening of RAPD primers against a segregating haploid population was conducted to provide insight on the levels of heterozygosity within a "typical" outcrossed southern pine genotype.

Materials and Methods

Plant Material

Parent trees originally selected for this study came out of an existing pine breeding program being conducted and maintained by the USDA Forest Service,

Southern Institute of Forest Genetics in Gulfport, Mississippi. A total of ten trees were selected. Four longleaf pine of southeastern Louisiana origin (clones 27-168, 3-356, 5-77, and 15-366) were chosen based on early height growth and superior brown-spot needle resistance. Two loblolly pines also of southeastern Louisiana origin (clones A-1-14 and A-1-9) were chosen based on early height growth. Finally, four slash pine (clones 9-2, H-28, W-1-7, and J-1-5) of north Florida origin were selected for height growth and fusiform rust resistance. Seed for investigating the level of heterozygosity within a single genotype were obtained from wind-pollinated cones of longleaf pine (clone 27-168).

DNA Isolation and Purification

In the case of the parental trees, total nucleic acids were isolated from needles as described in Wagner et al. (1987) except that spermine and spermidine were omitted from both the extraction and wash buffers. The same (scaled-down) technique was also used to isolate total DNAs from megagametophytes of wind-pollinated seed of longleaf pine clone 27-168.

RAPD Procedures

RAPD primers were obtained from Operon Technologies (Alameda, CA) [Oligo Sets A, B, E, G, W, X, and Y]. RAPD reactions were based on the protocol reported by Williams et al. (1990), with modifications suggested by C. D. Nelson (USDA Forest Service). The reaction consisted of the following in 25 μl total volume: varying concentrations of template DNA (1.0 ng to 1.0 μg) , 5 pmoles primer DNA (0.3125 μM), 3.2 nmoles each dNTP (200 μM each), 2.5 μl 10X *Taq*

DNA polymerase reaction buffer (Promega, Madison, WI)[500 mM KCl, 100 mM Tris-HCl (pH 9.0 at 25°C), 1% Triton X-100], varying concentrations of $MgCl_2$ (0.0 to 5.0 mM), and 1.0 U *Taq* DNA polymerase (Promega, Madison, WI). Reactions were loaded in 250 μl microcentrifuge tubes, overlaid with 25 μl of mineral oil, and amplified in a COY temperature cycler (Coy Laboratory Inc., Ann Arbor, MI). The RAPD reactions were driven to completion using the following thermal profile: 45 cycles (1 min at 94°C, one min at 36°C, and 2 min at 72°C), followed by an indefinite hold at 4°C.

The completed RAPD reactions were electrophoresed in 1.4% agarose gels and TBE buffer (90 mM Tris-borate, 2 mM EDTA pH 8.0) containing ethidium bromide for approximately 5.0 hr at 93 V. Prior to gel electrophoresis, 5.0 μl of loading buffer (40% sucrose, 0.5% bromophenol blue) was added to each reaction. Upon completion, the gels were washed in distilled water for 45 min, and photographed over UV light using a Polaroid 545 Land camera and Polaroid Type 57 instant film.

The optimal range of template DNA concentration was determined by varying this parameter and holding all other reaction parameters constant (as described above, with 2.0 mM $MgCl_2$). The concentrations of template DNA examined were 1.0 ng, 2.5 ng, 5.0 ng, 10 ng, 25 ng, 50 ng, 100 ng, and 1.0 μg per reaction. The experiment was performed twice to ensure repeatability. Determination of the optimal range of Mg^{+2} was conducted in a similar manner (using optimal level found

above). The various Mg^{+2} concentrations investigated were 0.0 mM, 0.5 mM, 1.0 mM, 2.0 mM, 3.0 mM, 4.0 mM, and 5.0 mM $MgCl_2$ per reaction.

The level of heterozygosity within a single southern pine genotype was determined by screening a total of 133 oligonucleotide primers against DNAs obtained from longleaf pine 27-168 and a subset of wind-pollinated megagametophytes (11 to 12 megagametophytes). Those RAPD bands which were amplified in the parental DNA and in only some (but not all) of the megagametophytic DNAs were scored as heterozygous.

Results and Discussion

RAPD Optimization

After obtaining successful amplification from all parental DNAs [DNA concentration = 25 ng; $MgCl_2$ concentration 2.0 mM], one genotype was focused on to determine the sensitivity of fragment amplification to minor changes in the composition of the reaction mixture. If the RAPD phenotype produced was sensitive to minor fluctuations in the concentration of either template DNA or $MgCl_2$ then reproducibility among laboratories might be seriously compromised.

One of the more difficult parameters to accurately estimate is template DNA concentration. By keeping all other parameters constant [$MgCl_2$ concentration = 2.0 mM], the affect of various template DNA concentrations on the complexity of the RAPD profile was investigated. This provided a range of template DNA concentrations at which amplification products could consistently be obtained. Specific products were amplified at all template DNA concentrations examined

except 1.0 μg, which appeared as a non-specific smear of DNA after amplification. The most complex and reproducible profiles occurred over a ten-fold range between 2.5 ng to 25 ng. This demonstrated that some error in DNA concentration could be tolerated without compromising reproducibility.

Based on the previous experiment, a DNA concentration of 10 ng was chosen for future amplifications. Keeping all other parameters equal, the effect of various $MgCl_2$ concentrations on the complexity of the RAPD profile was also investigated. The most complex RAPD profiles were consistently obtained at the 2.0 and 3.0 mM $MgCl_2$ concentrations. In general, higher levels of $MgCl_2$ can be tolerated (at least up to 5.0 mM), however, lower concentrations of Mg^{+2} (0.0 and 0.5 mM) markedly decreased amplification. This was not surprising as *Taq* DNA polymerase requires Mg^{+2} as a necessary cofactor (Ausubel et al. 1987).

Parental RAPD Screening

In order to determine whether the RAPD technique was able to detect polymorphism among different southern pine species, all ten parental DNA samples were screened with 15 primers from Operon Technologies Oligo Kit A. As the long-term goal of this research is to speed the introgression of genes influencing rapid early height growth from slash pine or loblolly pine into longleaf pine, the presence and absence of RAPD bands between/among species was of primary interest. Of the 15 primers screened, 5 primers produced faint or unscorable amplification. Of the 10 primers which produced scorable amplification products, an average of 13.6 bands per primer were amplified, ranging in size from approximately 400 bp to 2000 bp.

Seven of the 10 primers which amplified parental DNAs identified species-specific RAPDs. A total of eleven species-specific polymorphisms were identified (see Table 2.1). Four primers amplified two polymorphisms each and three amplified one polymorphism.

Although only 10 primers produced scorable amplification products, the RAPD technique appeared to be able to detect a great deal of variation between/among the various southern pine species. Although caution must be taken when interpreting these data, as only a very limited number of individuals per species was investigated, it does appear that the RAPD technique will provide a suitable number of polymorphisms for mapping in an interspecific backcross program. Unfortunately, the heterozygous or homozygous condition of the band-present phenotype (in any one genotype) could not be determined by this analysis.

Table 2.1. Operon Technologies Oligo Kit A primers which amplified polymorphisms between/among species based on screening DNAs from four longleaf pines, two loblolly pines, and four slash pines. Plus (+) refers to band present in at least one individual of a species, minus (-) refers to band absent in all individuals of a species.

Primer	~Band Size (bp)	*P. palustris*	*P. taeda*	*P. elliottii*
A01	0550	+	-	+
A04	0950	+	-	-
A07	1000	-	+	-
A07	1275	-	-	+
A08	0600	-	+	-
A08	1800	+	-	-
A09	0600	+	+	-
A09	0800	-	-	+
A10	0650	-	+	-
A10	1000	-	+	-
A13	0900	-	-	+

Level of Heterozygosity

In order to investigate the level of heterozygosity within a "typical" outcrossed southern pine genotype, a total of 133 primers from Operon Technologies Inc. Oligo Kits A, B, E, G, W, X, and Y were screened against megagametophytic DNAs obtained from wind-pollinated seed of longleaf pine clone 27-168. An average of 8.3 bands was amplified per primer screened, ranging in size from approximately 300 bp to 2700 bp. The 133 primers amplified a total of 85 polymorphisms (0.64 polymorphisms per primer screened). Twenty-eight primers amplified one polymorphism, 14 amplified two polymorphisms, seven amplified three polymorphisms, and two amplified four polymorphisms. The level of heterozygosity detected by the RAPD technique within a "typical" longleaf pine genotype was found to be intermediate to those found for a segregating maize population [1.0 polymorphism per primer screened] and soybean population [0.5 polymorphism per primer screened] (Williams et al. 1990). The RAPD technique appears to be able to detect a great deal of variation within a single southern pine genotype. At the current level of detection, it should be possible to identify roughly 200 to 300 polymorphisms for genetic linkage analyses by screening a total of 300 to 450 primers.

CHAPTER 3

A SIMULATION STUDY OF LINKAGE MAP CONSTRUCTION WITH MISSING AND MIS-SCORED RAPD DATA[1]

Introduction

The relative ease and speed with which large numbers of RAPD markers can be generated makes them extremely appealing for use in constructing primary genetic linkage maps. RAPD markers are generated by the use of single, randomly sequenced oligonucleotide primers and the polymerase chain reaction (Williams et al., 1990). A segment of DNA is amplified whenever two nucleotide sequences with high degrees of complementarity to that of the primer occur within 2-3 Kb of one another on opposite strands of the template DNA. Repeated cycles of denaturation and extension result in the exponential amplification of the segment. Despite its conceptual simplicity the kinetics of the RAPD reaction are quite complex. Annealing temperature, degree of sequence similarity at priming sites, and primer competition all can affect the amplification of RAPD markers. In addition, when large numbers of RAPD reactions are being run on a daily basis failed reactions are not uncommon. As a result, amplification inconsistencies could produce spurious data in the form of mis-scored individuals, and unless re-amplified, failed reactions would have to be recorded as missing data. The goal of this research was to investigate what effects various levels of missing and mis-scored RAPD data have on recombination estimates and linkage group construction.

---

[1]Published in the Proceedings of The 22nd Southern Forest Tree Improvement Conference, see APPENDIX B for permission and proof of authorship.

Materials and Methods

Pedigree Construction

Ten known marker maps and corresponding data sets were constructed using the software GREGOR version 1.3 (Nick Tinker, McGill University). A configuration to model 10 pairs of chromosomes, 160 possible loci per chromosome and 1% recombination between adjacent loci, was chosen. A marker list consisting of 100 randomly distributed loci was defined for each map. The parents used for generating the mapping population were defined as follows: parent 1 was heterozygous (complete coupling) for all 100 marker loci, parent 2 was defined as being homozygous recessive for all 100 marker loci (tester). This coding arrangement would be similar to that used when constructing maps from haploid megagametophyte data. The mapping population consisted of 80 individuals.

Missing Data Sets

In order to investigate the effects of missing data, five MAPMAKER II-compatible data sets were produced from each GREGOR data set. One represented the true data set and four represented various levels of missing data (5%, 10%, 15%, and 20%), for a total of 50 data sets. To determine how missing data should be targeted, we looked at the distribution of missing entries in actual RAPD data sets. Based on data generated for 2 different slash pines (Nelson et al., 1992; vanBuijtenen et al., 1992) and a longleaf pine (Kubisiak et al., 1992), missing data appear to be exponentially distributed. Most markers have no, or a few missing entries, with considerably fewer markers being found as levels of missing data increase. By

randomly sampling from the function describing this distribution, the study targeted each marker to receive a specified number of missing entries, so that when averaged over all markers, the data set-wide levels were equal to 5%, 10%, 15%, or 20% missing.

Mis-scored Data Sets

Five MAPMAKER II-compatible data sets containing various levels of mis-scoring (0%, 1%, 2%, 4% and 8%) were also produced from each GREGOR data set, for a total of 50 data sets. When RAPD loci were scored, markers were categorized based on a confidence score (Kubisiak et al., 1992). A putative polymorphic locus was given a lower confidence rating if the locus of interest was only faintly amplified or bands of similar molecular weight as the locus of interest were present. If mis-scoring were to result from one of these two sources, most errors should be occurring within specific loci, versus being random over the entire data set. Therefore, in order to produce the overall levels of 1%, 2%, 4%, and 8% mis-scoring, 20% of the markers were randomly chosen to receive 5%, 10%, 20%, or 40% mis-scoring.

Mapping Strategy

The data sets were entered into the computer package MAPMAKER II (version 1.9), and recombination estimates and linkage group information were obtained. The mapping strategy was similar to that suggested by Lander et al. (1987). To determine all two-point groupings, a log of the odds ratio (LOD) of 5.0 and a recombination frequency of 0.25 were chosen. To determine marker order

within a particular linkage group, a LOD score of 3.0 was chosen. These markers, and their respective orders, were designated as framework groupings.

Data Analysis

In order to evaluate the effects of missing or mis-scored data, an analysis of variance was used to determine if there were significant differences among treatment means for various descriptive measures. These included the standard deviation of the departure of two-point recombination estimates from their "true" or known values [std(|r-θ|)], number of framework markers mapped, number of linkage groups obtained, number of markers per linkage group, and number of marker order changes. The std(|r-θ|) had the following form:

$$\sqrt{\frac{\sum_{1}^{p}\left[(|r_1-\theta_1|)-\frac{\sum_{1}^{p}(|r_1-\theta_1|)}{p}\right]^2}{p}}$$

r = pairwise recombination estimate
θ = "true" or known pairwise distance
p = number of pairwise comparisons
(For 100 loci p=4950)

Results

Given no mis-scored data, std(|r-θ|) was found to increase with the level of missing data (Table 3.1). A significant difference among treatment mean groupings occurred at 15%. With no missing data, std(|r-θ|) did not appear to increase until a level of 4% mis-scoring was attained (Table 3.1). However, it was not until 8% that a significant difference among treatment mean groupings was detected. The number

Table 3.1. Effect of missing and mis-scored data on the standard deviation of two-point recombination estimates [std(|r-θ|].

Missing Data		Mis-scored Data	
Treatment (% missing)	Mean std(|r-θ|)	Treatment (% mis-scored)	Mean std(|r-θ|)
0	0.056045 A*	2	0.055173 A
5	0.056627 A	1	0.055358 A
10	0.057566 A	0	0.056045 A
15	0.059567 B	4	0.056638 A
20	0.061033 B	8	0.060773 B

*Those means with the same letter are not significantly different at alpha = 0.05 using Tukey's Studentized Range Test. Means based on 10 replicate data sets.

of markers placed into framework groupings was found to decrease as the level of missing or mis-scored data increased (Table 3.2). A significant difference among treatment mean groupings was not detected until a level of 20% missing data was attained. For the mis-scored data sets, a significant difference among treatment mean groupings was not detected until a level of 4% was attained (Table 3.2). Missing

Table 3.2. Effect of missing and mis-scored data on the number of framework markers mapped (F.M.).

Missing Data		Mis-scored Data	
Treatment (% missing)	Mean # F.M.	Treatment % mis-scored)	Mean # F.M.
0	69.8 A*	0	69.8 A
5	68.6 A	1	66.5 A,B
10	65.0 A	2	60.7 B
15	64.6 A	4	49.8 C
20	57.7 B	8	48.3 C

*Those means with the same letter are not significantly different at alpha = 0.05 using Tukey's Studentized Range Test. Means based on 10 replicate data sets.

data did not appear to affect the number of linkage groups obtained in any sort of a predictable manner (Table 3.3). No significant differences among treatment means were detected. However, as the levels of mis-scoring increased the number of linkage groups obtained was found to decrease (Table 3.3). A significant difference among

Table 3.3. Effect of missing and mis-scored data on the number of linkage groups (L.G.).

Missing Data		Mis-scored Data	
Treatment (% missing)	Mean # L.G.	Treatment (% mis-scored)	Mean # L.G.
0	14.6 A*	0	14.6 A
5	14.0 A	1	14.6 A
10	14.3 A	2	13.9 A
15	14.8 A	4	11.9 B
20	13.8 A	8	11.5 B

*Those means with the same letter are not significantly different at alpha = 0.05 using Tukey's Studentized Range Test. Means based on 10 replicate data sets.

treatment mean groupings occurred at 4%. The average number of markers per linkage group was, generally found to decrease as the levels of missing or mis-scored data increased (Table 3.4). However, no significant differences among treatment mean groupings was detected. Finally, the number of marker order changes was not found to be affected by missing data in any sort of a predictable manner, and no significant differences among treatment means was detected (Table 3.5). The number of marker order changes was found to increase up to the level of 2% mis-scoring, after which numbers decreased again (Table 3.5). Treatment means were not determined to be statistically different.

Table 3.4. Effect of missing and mis-scored data on the average number of framework markers per linkage group (M./L.G.).

Missing Data		Mis-scored Data	
Treatment (% missing)	Mean # M./L.G.	Treatment (% mis-scored)	Mean # M./L.G.
5	4.948 A*	0	4.811 A
0	4.811 A,B	1	4.568 A,B
10	4.591 A,B	2	4.388 A,B
15	4.410 A,B	8	4.216 B
20	4.216 B	4	4.194 B

*Those means with the same letter are not significantly different at alpha = 0.05 using Tukey's Studentized Range Test. Means based on 10 replicate data sets.

Table 3.5. Effect of missing and mis-scored data on the number of marker order changes (O.C.).

Missing Data		Mis-scored Data	
Treatment (% missing)	Mean # O.C.	Treatment (% mis-scored)	Mean # O.C.
20	0.50 A*	2	1.20 A
10	0.30 A	4	0.80 A,B
0	0.20 A	1	0.50 A,B
15	0.10 A	0	0.20 A,B
5	0.00 A	8	0.10 B

*Those means with the same letter are not significantly different at alpha = 0.05 using Tukey's Studentized Range Test. Means based on 10 replicate data sets.

The interaction between missing and mis-scored data was also investigated and not found to be statistically significant for any of the variables investigated.

Discussion

Prior to the analysis, we hypothesized that as levels of missing or mis-scored data increased within a data set the standard deviation of the departure of pairwise recombination estimates from their "true" or known values, std($|r-\theta|$) would increase;

likewise, the accuracy with which the genetic distance between two markers can be estimated decreases. The results seem to support our hypothesis. The std($|r-\theta|$) was found to increase as the levels of missing or mis-scored data increased (significance at 15% and 8%, respectively).

We hypothesized that levels of mis-scored data would have a more pronounced effect on std($|r-\theta|$) than would comparable levels of missing data. Missing data only indirectly affect linkage estimation by reducing the effective mapping population within particular markers. However, mis-scored data would tend to confound the linkage relationship between markers and hence directly affect recombination estimates. Interestingly, these simulations indicate that, in terms of std($|r-\theta|$), the overall effects of a 5% level of missing data are comparable to a 4% level of mis-scoring.

The differences among mean values for the various levels of missing data and mis-scored data appear to be quite small in terms of genetic distance, however, this is due to the fact that a majority of the markers which are unlinked are still estimated to be unlinked even when harboring missing or mis-scored data. Therefore, a large number of the comparisons are not contributing to the sum of squared deviations in the calculation. The levels at which significant differences among mean groupings were detected does appear to be indicative of a problem threshold. In other words, the levels of missing or mis-scored data at which significant differences among treatment means were detected for std($|r-\theta|$) are similar to the levels found to cause

significant group discrimination in other measures such as the number of framework markers, number of linkage groups, and number of markers per linkage group.

It makes sense that the number of markers placed into framework groupings would decrease as levels of missing and mis-scored data increase. These results indicate just such an inverse relationship (Table 3.2). Given no mis-scored data, at levels of 20% missing data, 17.3% fewer markers could be placed into framework groupings. Given no missing data, at levels of 4% mis-scoring, 28.6% fewer markers could be placed into framework groupings. In terms of the number of framework markers placed, lower levels of mis-scoring appear to have a more profound effect than do comparable levels of missing data.

Developing *a priori* hypotheses regarding how missing and mis-scored data might affect the number of linkage groups and average number of markers per linkage group was a more problematic situation. It could be hypothesized that missing or mis-scored data might cause whole linkage groups to fall apart, resulting in fewer mapped groups. Alternatively, mis-scored or missing data might cause larger groups to be broken into two or more smaller groups, resulting in a larger number of mapped groups, each having fewer markers. In terms of missing data, no apparent trends were found in the number of linkage groups obtained. However, as levels of mis-scoring increased, the number of linkage groups obtained decreased. This would seem to indicate that mis-scoring is primarily causing entire linkage groups to be lost. Although the number of framework markers per linkage group

appears to decrease with increased missing or mis-scored data, no significant difference among treatment mean groupings was found.

Prior to the analysis, we hypothesized that marker ordering would be adversely influenced by increased levels of missing and mis-scored data. We also felt that, at comparable levels, mis-scored data would have more of an influence on marker ordering than would missing data. Over all simulated data sets, surprisingly few marker order changes occurred (38 or 0.38 per data set). Consistent with our *a priori* expectations, 27 (71%) of these were found to occur in mis-scored data sets. There does not appear to be any apparent trend in the number of marker order changes for the various levels of missing data. However, for the mis-scored data, the number of marker order changes increased up to the 2% level, beyond which they decreased. Low levels of mis-scoring tend to cause marker order changes, whereas higher levels confound linkages, causing markers to be dropped from the map.

Interestingly, there does not appear to be an additive effect when both missing and mis-scored data are included in the same data set. For example, with 10% missing data, the mean number of markers placed into framework groupings was 65, that is 4.8 fewer than with no missing, and with 2% mis-scored, the mean number placed was 60.7, 9.1 fewer than with no mis-scored. When both were included, the number of markers placed was 64.2. We would have expected this mean to be somewhat lower if the effect of both missing and mis-scored data was additive.

In addition to gaining a better understanding of how missing and mis-scored data effect primary map construction, we were also interested in how trends in the

simulated data sets compared with those seen in actual RAPD data sets. Due to the fact that we do not know the true distances between markers or the level and distribution of mis-scoring in actual RAPD data sets, some of the measures investigated in this paper are not directly comparable. However, we do know the level and distribution of missing values. In a RAPD data set generated for longleaf pine, levels of missing data were found to approach 5%. When compared with simulated data sets with 5% missing data, some general trends appear. For the simulated data sets, on average, 68.6% of the markers could be placed into framework groupings. For longleaf pine, 64.3% of the markers (121 out of 188) were placed into framework groupings. In the simulated data sets, the majority of the markers that are lost as a result of 5% missing data are those harboring the highest levels of missing entries. Only 52.6% of the markers with greater than 15% missing entries mapped, whereas 71.4% of those markers with less than 15% missing entries mapped. For the longleaf data set, only 41.2% of the markers with greater than 15% missing entries mapped, whereas 66.1% of the markers with less than 15% missing entries mapped.

Although the amount of mis-scoring that occurs within an actual RAPD data set is not known, we felt that if mis-scoring were occurring it would be primarily concentrated in the markers with lower confidence scores. If the percentage of markers mapped per confidence grouping is used as an indicator of potential mis-scoring, this would appear to be the general trend for the longleaf data set. Of the markers that were entered into MAPMAKER II, 70.6% of the markers classified as

good were mapped, whereas only 52.2% of the markers classified as fair were mapped. Mis-scoring might possibly be responsible for the difference between these percentages.

Based on the variables analyzed in this simulation study, it appears as if levels of missing and mis-scored data as high as 15% and 2%, respectively, can be tolerated during primary genetic map preparation, since they do not significantly affect recombination estimates or linkage group construction. The genetic system simulated in this analysis, 100 markers distributed randomly over a 1600 cM genome, is fairly representative of the situation encountered during the early phases of linkage mapping. We caution, however, that the results from this study are only applicable to low density mapping situations (i.e. early map construction). When more saturated linkage conditions exist the effects of missing and mis-scored data will have more of an impact, particularly on marker ordering.

CHAPTER 4

RAPD LINKAGE MAPPING IN A LONGLEAF PINE × SLASH PINE F_1 FAMILY[2]

Introduction

Longleaf pine possesses many desirable qualities such as excellent timber form, high naval stores content, moderate to high wood specific gravity, and fusiform rust resistance (USDA 1965). Despite these qualities, complications associated with an extended phase (2 to 20+ years) of juvenile development referred to as the grass stage, have limited its use in artificial regeneration programs (Schmidtling and White 1989). During the grass stage, seedlings develop extensive root systems, and apical meristems increase in diameter but not in length (Harlow et al. 1978). The grass stage greatly increases the opportunity for brown-spot needle blight infection [caused by the fungus *Scirrhia acicola* (Dearn.)] (Siggers 1944). This disease can greatly prolong the grass stage and, if severe enough, can kill seedlings. Applications of fungicides to the roots of longleaf pine prior to planting can significantly reduce brown-spot disease and promote early height growth (Kais 1975; Kais et al. 1981). Regardless of these efforts, increased seedling mortality (compared to the other southern pines) and the unpredictability of the duration of the grass stage still make planting longleaf pine a risky investment under intensive management systems.

Inter-specific hybrids of longleaf pine have shown promise for addressing the problem of delayed height growth. Intermediate height growth has been observed in

[2]Accepted for publication in THEORETICAL AND APPLIED GENETICS.

various families of longleaf pine crossed to either loblolly pine or slash pine (Brown 1964; Derr 1966; 1969). Analysis of F_2 and BC_1 hybrids of longleaf pine and loblolly pine yielded an estimate of at least 10 loci controlling EHG (Brown 1964 using methods of Wright [in Castle 1921]). More recent work with hybrids of longleaf pine and loblolly pine indicates that there may be fewer than 5 loci affecting EHG (C. D. Nelson unpublished data, using methods of Lande 1981 and Cockerham 1986). In light of recent theoretical work (Zeng et al. 1990; Zeng 1992) and the fact that current estimates of the number of loci influencing EHG are based on only a few hybrid families, estimates calculated to date may be low, but they do suggest that the grass stage character is a quantitative trait controlled by a finite number of genes (oligogenic vs. polygenic).

Recent advances in DNA-based marker technology have made it possible to conduct efficient genetic mapping and quantitative trait loci (QTL) searching experiments. Both restriction fragment length polymorphisms (RFLPs) and random amplified polymorphic DNAs (RAPDs) have been used to construct genetic linkage maps for a number of annual crop species (e.g. Bernatzky and Tanksley 1986; Helentjaris et al. 1986; Landry et al. 1987; McCouch et al. 1988; Gebhardt et al. 1989; Keim et al. 1990; Paran et al. 1991), as well as for several perennial tree species (Tulsieram et al. 1992; Faure et al. 1993; Liu and Furnier 1993; Nelson et al. 1993; Nelson et al. in press; Hemmat et al. 1994). Due to the large number of marker assays required for genetic mapping and QTL searching experiments,

automated approaches afforded by the RAPD technique offer an enormous benefit in terms of time and labor (Grattapaglia et al. 1992; Nelson et al. in press).

Low- to medium-density RAPD maps have recently been published for several tree species such as white spruce (Tulsieram et al. 1992), slash pine (Nelson et al. 1993), longleaf pine (Nelson et al. in press), and apple (Hemmat et al. 1994). In the case of conifers, their unique reproductive biology provides a source of segregating haploid tissue, the megagametophyte (USDA 1974). The megagametophyte is derived from repeated mitotic divisions of a single meiotic product, and has the same maternal genetic complement as the embryo contained in the same seed. Since the megagametophyte is haploid (of maternal origin), segregation and recombination can be evaluated in a sample of seeds from a single tree without the need for controlled pollinations (Guries et al. 1978; Rudin and Ekberg 1978; Conkle 1981). Sufficient DNA for several hundred to several thousand RAPD reactions can be extracted from one megagametophyte, making this an excellent system for evaluating the efficacy of the RAPD technique for constructing genetic linkage maps.

Despite the advantage afforded by the megagametophyte system (uni-parental segregation), this system could potentially be inefficient in terms of QTL analyses and marker-aided selection (MAS) applications. The genotypic data scored/collected on megagametophytes can be used to search for QTL (of maternal origin) in plants derived from these same seeds (Grattapaglia et al. 1992; Liu et al. 1993). However, the paternal genetic contribution to these plants might potentially confound QTL

analyses. The genetic component of an individual's phenotype results from specific alleles received from both its maternal and paternal parents. If precautions are not taken to control (randomize) the paternal genetic complement (for example, by using polymix pollen sources) this approach to QTL mapping might prove to be mis-leading.

In terms of genetic mapping (and subsequent QTL searching), a more efficient strategy would be to simultaneously map each parent of a specific cross by using their progeny as the mapping population (Carlson et al. 1991; Hemmat et al. 1994). For dominant RAPD markers, the practicality of such an approach is limited by the number of loci found to be in testcross configuration between parents. Using eight different RAPD primers, Carlson et al. (1991) identified 10 loci segregating in at least one of three Douglas-fir F_1 families. Most of these loci (70%) were found to be in testcross configuration between the parents. Using a total of 64 primers or primer combinations, Hemmat et al. (1994) identified a total 367 RAPD loci segregating in the F_1 progeny of a single cross between the apple cultivars "White Angel" and "Rome Beauty". Greater than 90% of these loci were in testcross configuration between the parents. Loci at which both parents were heterozygous (segregating 3:1) were useful in defining homologous counterparts between the two parental maps. These results suggest that the RAPD technique is well-suited to genetic mapping in highly heterozygous outcrossed species.

In the present study, we used RAPD markers to simultaneously construct linkage maps for the parents of a longleaf pine × slash pine F_1 family. The long-

term goal of this research is to employ these markers in a backcross breeding program to accelerate the introgression of positive effect early height growth (EHG) alleles from slash pine into longleaf pine. In this paper, we present RAPD linkage maps for the parents of an F_1 family and discuss how this information might be used in backcross populations to map EHG loci and apply genomic selection.

Materials and Methods

Plant Material

A longleaf pine × slash pine inter-specific F_1 family was used as the mapping population. The family was produced by crossing slash pine H-28 (♂) to longleaf pine 3-356 (♀). The longleaf pine parent had previously been mapped using RAPD markers and the megagametophyte system (Nelson et al. in press). No mapping information was available for slash pine H-28. Both parents were selected for disease resistance and growth rate at the Southern Institute of Forest Genetics (SIFG) near Gulfport, Mississippi. The cross was completed in the spring of 1990. Seeds were extracted from mature cones in the fall of 1991 and sown into containers in February of 1992. The seedlings were grown in a greenhouse for four months and then transplanted to a nursery bed. A total of 98 progeny were available for use in this study.

DNA Isolation and Purification

Total nucleic acids were isolated from needles as described in Wagner et al. (1987) except that spermine and spermidine were omitted from both the extraction and wash buffers. The RNA component of these extracts was removed by incubation

in the presence of RNase A as described in Ausubel et al. (1987). Approximately 2.0 μg of DNA was further purified using the Prep-A-Gene™ DNA Purification Kit (BioRad, Hercules, CA) as described by the manufacturer.

Primer Selection and DNA Amplification

Primer DNAs were obtained from either Operon Technologies (Alameda, CA) or J.E. Carlson (Univ. of British Columbia, Vancouver, B.C., Canada). A total of 288 primers were chosen for this study. Of these 288 primers, 102 (35.4%) were chosen because they had been used previously to amplify mappable loci in our longleaf pine parent (clone 3-356) using the haploid megagametophyte system (Nelson et al. in press); another 148 primers (51.4%) were chosen because they had previously amplified mappable loci in one of several other pine species (Nelson and Nance unpubl. data). The remaining 38 primers (13.2%) were randomly selected from a set of over 250 additional candidate primers. DNA amplification followed the protocol outlined in Nelson et al. (in press). The only modification consisted of a doubling of the template DNA to 6.25 ng per reaction, to compensate (in theory) for the use of diploid versus haploid material.

Primer Screening and Marker Scoring

To identify primers which amplified polymorphisms, primers were screened against both parents and six F_1 progeny. Three different parent-progeny RAPD banding patterns were scored as putatively polymorphic. When a RAPD band was present in only one parent, and in at least one of the six progeny, the parent was classified as potentially heterozygous for that locus (referred to as testcross loci).

RAPD bands which were present in both parents and absent in at least one of the progeny, classified both parents as potentially heterozygous (referred to as 3:1 loci). Bands which were present in only one parent and all of the progeny, tentatively classified each parent as homozygous for alternate alleles (referred to as non-segregating loci). A subset of primers (162) that maximized the number of polymorphisms in testcross configuration was selected and segregation scored in an additional 80 progeny. The 80 progeny were divided into four template sets (three sets consisting of 22 progeny each, and a fourth set of an additional 14 progeny). Each template set was amplified, along with both parents, on different temperature cyclers. Those polymorphisms that could confidently be scored across all four template sets were included in our analyses. In the case of testcross loci, presence of a band was scored as an 'H' (heterozygous) while absence of a band was scored as 'A' (homozygous band absent). Those cases in which the presence or absence of bands was unclear, were recorded as missing data.

Marker Naming

Each polymorphism was assigned a two-part name according to Nelson et al. (in press). The first part corresponding to the primer with which the polymorphism was amplified (a letter followed by a two-digit number corresponds to an Operon Technologies Inc. primer, and a three-digit number corresponds to a University of British Columbia primer), followed by a letter corresponding to consecutively identified polymorphisms within the same primer. When only a single polymorphism was amplified by a primer, it was given the letter A. When a primer amplified

multiple polymorphisms, the polymorphisms were assigned consecutive letter designations from the smallest molecular weight band to the largest (for example X19_A, X19_B, X19_C).

Segregation Analysis

Each RAPD band was tested for goodness of fit to a 1:1 or 3:1 Mendelian segregation ratio by chi-square (χ^2) analysis ($\alpha = 0.05$). Those loci that appeared to be experiencing segregation distortion were excluded from initial mapping analyses. The testcross data were entered into the computer package MAPMAKER/EXP (version 3.0) and analyzed using a modified backcross format (Nelson et al. 1993). The mapping strategy used was similar to that suggested in Lincoln et al. (1992). Significant associations between 3:1 loci and testcross loci were determined by χ^2 analysis. By considering only the homozygous band-absent genotypes (-/-) at a 3:1 locus, linkage with a testcross locus was implied if the testcross locus genotypes significantly deviated from 1:1 in this subset ($p < 0.001$).

Linkage Group Designations

Linkage groups were assigned three letter names. The first two letters designate species (Pp = *Pinus palustris*, Pe = *Pinus elliottii*), and the third designates linkage groups in descending size (A = largest). The longleaf pine linkage groups were assigned names according to Nelson et al. (in press). Linkage groups with letter designations from PpA to PpP are homologous to those identified in Nelson et al. (in press). Longleaf pine linkage groups that currently show no homology between maps were assigned additional letter designations (PpQ-PpT).

Results

We screened 288 oligonucleotide primers against the parents and six F_1 progeny of a longleaf pine × slash pine cross. Results of primer pre-screening are summarized in Table 4.1. Of the 288 primers initially screened, 172 primers

Table 4.1. Results of pre-screening 288 primers against both parents and six F_1 progeny of a longleaf pine × slash pine cross.

Primer Group[1]	# of Primers Screened	# of Primers Revealing Polymorphism	# of Putative Marker Loci
longleaf	102	79	162
other pine	148	80	138
random	38	13	18
Total	288	172	318

[1]Primer group: longleaf - these primers previously amplified mappable loci in longleaf 3-356; other pine - these primers previously amplified mappable loci in one of several other pine genotypes; random - these primers were randomly selected.

amplified a total of 318 putatively segregating polymorphisms. Of these 318 polymorphisms, 298 were tentatively classified as testcross loci (possibly segregating 1:1), and the other 20 appeared to be loci heterozygous in both parents (possibly segregating 3:1). Of the 298 putative testcross loci, 167 loci were heterozygous in the longleaf pine parent and the other 131 were heterozygous in the slash pine parent. Based on these results, 162 primers were chosen to further characterize these polymorphisms in an additional 80 progeny.

With the 162 pre-screened primers, a total of 281 segregating loci and 87 polymorphic (between parents) but non-segregating loci were scored on an additional four template sets of the mapping population. Of the 281 segregating loci, 267 were

in testcross configuration, and the other 14 were heterozygous in both parents. One hundred fifty-two of the 267 testcross loci were heterozygous in the longleaf pine parent, the other 115 loci were heterozygous in the slash pine parent. Of the 87 potentially polymorphic but non-segregating loci, 43 loci were unique to the longleaf pine genome and 44 loci were unique to the slash pine genome.

One hundred fifteen of the 162 primers chosen to run against the mapping population identified a total of 152 testcross loci in the longleaf pine parent (86 primers identified one RAPD locus, 21 identified two loci, and eight identified three loci). Ninety-one primers identified a total of 115 testcross loci in the slash pine parent (69 primers identified one RAPD locus, 15 identified two loci, two identified three loci, one identified four loci, and one identified six loci). Thirty-six of the 162 primers identified segregating RAPD loci in both parents.

Chi-square analysis indicated that 132 of the 152 longleaf pine testcross loci were segregating at the expected 1:1 Mendelian ratio ($n = 86$, $\alpha = 0.05$), as were 101 of the 115 slash pine testcross loci. Those loci that appeared to be experiencing segregation distortion were excluded from initial mapping analyses. Two-point analysis of the longleaf pine testcross loci classified 129 of the 132 loci into 18 groups (three or more loci) and three pairs. Ninety-one of the 101 slash pine testcross loci were classified into 13 groups and seven pairs. Orders of loci that were consistent for all three-point tests were taken as framework orders. This analysis resulted in the ordering of 64 framework loci within 13 longleaf pine linkage groups, and 41 framework loci within nine slash pine linkage groups (see markers with prefix

*, Figure 4.1). Those loci which were clearly linked to a particular group, but could not be confidently ordered using framework thresholds, were placed in their most likely positions. An additional 52 loci were placed in longleaf pine and 38 in slash pine. Linkage of those loci initially ungrouped (based on two-point analyses), as well as linkage between all possible pairs of groups was tested. Only one further linkage was suggested. The LOD for linkage between the grouped slash pine markers (268_A, E02_A, 169_A) and (608_A, 561_A) was 3.1, corresponding to a genetic distance of 38.3 cM (See linkage group PeG, Figure 4.1).

Linkage analysis of the 132 testcross loci heterozygous in the longleaf pine parent suggested a genetic map consisting of 18 groups and 3 linked pairs (122 markers) spanning a total of 1367.5 cM (See Figure 4.1). The weighted-average distance between markers within the 21 longleaf pine linkage groups is 13.0 cM (13.6 cM in the groups and 9.3 cM in the pairs). Using methods described in Hulbert et al. (1988), genome size estimates for longleaf pine were 2373, 2348, and 2392 cM for LOD scores of 2.0, 3.0, and 4.0, respectively. Assuming that each unlinked marker accounts for 30 cM, and that the 24 of the 42 ends of our 21 linkage groups and pairs cover true telomeric regions (15 cM/unaccounted end), the total map coverage is estimated at 1937.5 cM, or 80.9% of our largest genome size estimate.

Linkage analysis of the 101 testcross loci identified in the slash pine parent suggested 13 groups and 6 linked pairs (91 markers) spanning a total of 952.9 cM,

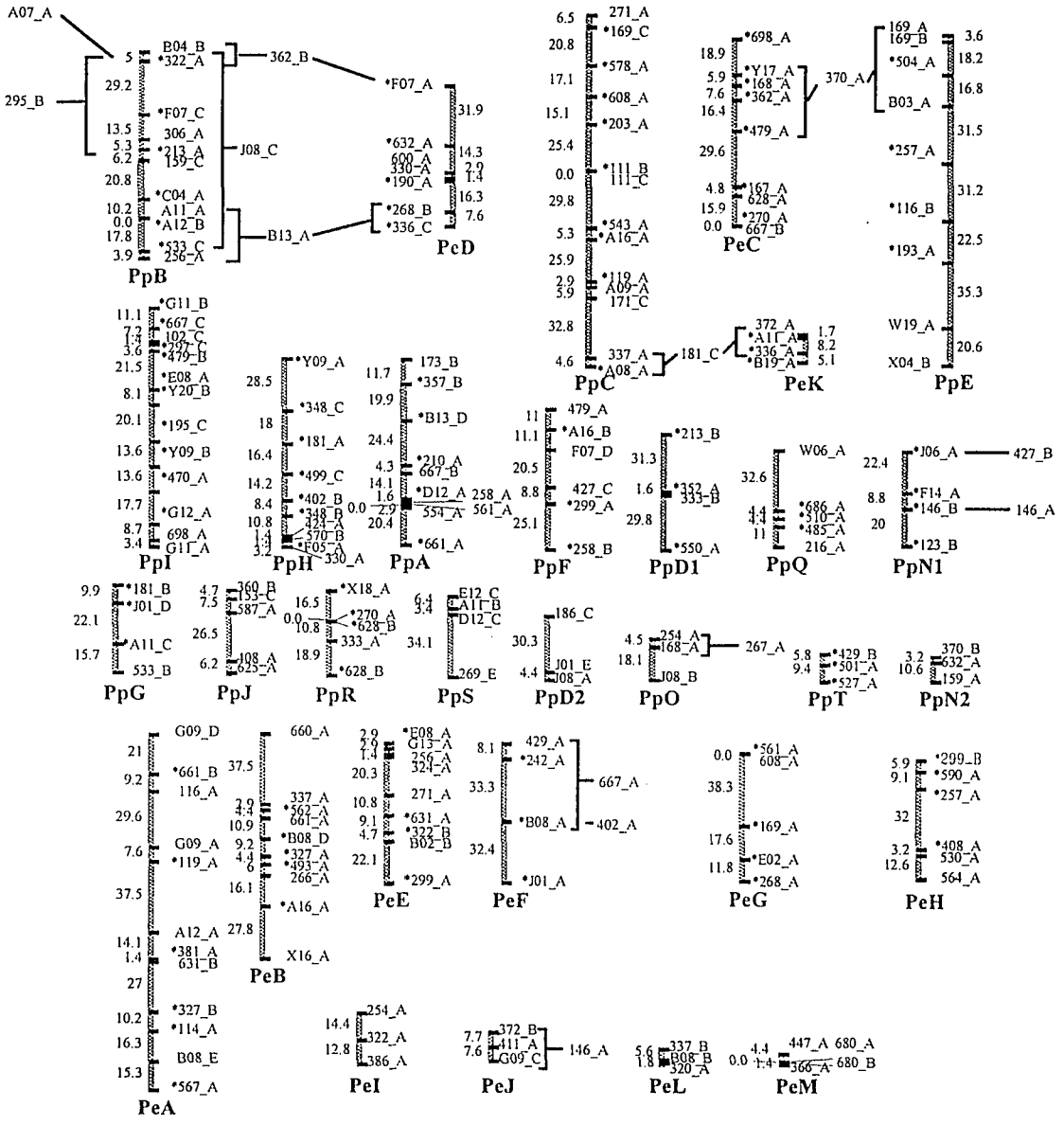

Figure 4.1. Genetic linkage maps of longleaf pine clone 3-356 (linkage groups Pp_) and slash pine clone H-28 (linkage groups Pe_). Linkage groups are arranged in descending order (A = largest). Longleaf pine linkage groups assigned letter designations from PpA to PpP are homologous to those identified in Nelson et al. (in press). Longleaf pine linkage groups which currently show no homology between maps were assigned additional letter designations (PpQ-PpT). Primer names and Haldane centiMorgan (cM) distances are provided. Framework markers (LOD 3.0, distance 35 cM) are indicated by an asterisk. Association of 3:1 loci is indicated by heavy lines.

with a weighted-average distance between markers of 16.1 cM (12.9 cM in the groups and 20.2 cM in the pairs). Genome size estimates for slash pine were 2292, 2342, and 2372 cM for LOD scores of 2.0, 3.0, and 4.0, respectively. Again, using a 30 cM map scale we estimate the total map coverage to be 1462.9 cM (or approximately 61.7% of the slash pine genome).

Chi-square analyses performed between 3:1 loci and testcross loci suggested a number of significant associations (See Table 4.2). Most 3:1 loci were found to be

Table 4.2. Results of chi-square (χ^2) analyses between 3:1 loci and testcross loci. By considering only the homozygous band-absent genotypes (-/-) at a 3:1 locus, linkage with a testcross locus was implied if the testcross locus genotypes significantly deviated from 1:1 in this subset ($p < 0.001$).

3:1 Loci	Testcross Loci	chi-square	p-value	Linkage Group
146_A	146_B	10.7143	~0.001	PpN1
146_A	411_A	13.7619	<0.001	PeJ
146_A	372_B	12.8000	<0.001	PeJ
146_A	G09_C	12.8000	<0.001	PeJ
181_C	337_A	11.0000	0.0	PpC
181_C	A08_A	9.8000	<0.01	PpC
181_C	336_A	18.0000	0.0	PeK
181_C	A11_A	16.2000	<0.001	PeK
181_C	372_A	15.2105	<0.001	PeK
267_A	254_A	12.0000	0.0	PpO
267_A	168_A	8.3333	<0.01	PpO
295_B	322_A	15.0000	0.0	PpB
295_B	A07_B	11.2667	<0.001	unordered(PpB)
295_B	F07_C	10.2857	<0.01	PpB
295_B	213_A	9.0000	<0.01	PpB

table con'd.

3:1 Loci	Testcross Loci	chi-square	p-value	Linkage Group
362_B	B04_B	22.0000	0.0	PpB
362_B	322_A	15.6957	<0.001	PpB
362_B	C04_A	15.6957	<0.001	Pair
362_B	F07_A	12.5652	<0.001	PeD
370_A	169_B	19.1739	<0.001	PpE
370_A	B03_A	18.1818	<0.001	PpE
370_A	169_A	15.6957	<0.001	PpE
370_A	504_A	14.7273	<0.001	PpE
370_A	173_B	12.5652	<0.001	unordered (PpE)
370_A	362_A	21.0000	0.0	PeC
370_A	168_A	15.6957	<0.001	PeC
370_A	479_A	15.6957	<0.001	PeC
370_A	Y17_A	12.5652	<0.001	PeC
402_A	B08_A	11.8421	<0.001	PeF
427_B	J06_A	11.2667	<0.001	PpN1
427_B	299_C	16.0000	0.0	Pair
427_B	116_B	11.2667	<0.001	Pair
667_A	242_A	13.00	0.0	PeF
667_A	429_A	13.000	00.0	PeF
667_A	B08_A	12.000	00.0	PeF
A07_A	322_A	11.2667	<0.001	PpB
B13_B	256_A	16.0000	0.0	PpB
B13_B	533_C	12.2500	<0.001	PpB
B13_B	590_A	12.2500	<0.001	unordered (PpB)
B13_B	A11_A	12.2500	<0.001	PpB
B13_B	336_C	16.0000	0.0	PeD
B13_B	268_B	12.2500	<0.001	PeD
J08_C	213_A	17.0000	0.0	PpB
J08_C	306_A	17.0000	0.0	PpB
J08_C	F07_C	16.0000	0.0	PpB
J08_C	A11_A	13.2353	<0.001	PpB
J08_C	322_A	13.2353	<0.001	PpB
J08_C	533_C	11.2667	<0.001	PpB
J08_C	B13_B	13.2353	<0.001	PeP

associated with more than one testcross locus per linkage group ($p < 0.001$), providing evidence against chance associations. Ten 3:1 loci were found to be associated with testcross loci on five different longleaf pine linkage groups (see Figure 4.1). Seven of the 3:1 loci were also found to be significantly associated with testcross loci on five different slash pine linkage groups. Those 3:1 loci that showed significant association to markers in both parents were used to infer homologous linkage groups between maps. Of the ten 3:1 loci that were found to be associated with longleaf pine testcross loci, six also suggested significant association to slash pine testcross loci (Refer to Table 4.2, and Figures 4.1 & 4.2). Four of the longleaf pine linkage groups appear to be potentially homologous counterparts to five different slash pine linkage groups (PpB-PeD,PeP; PpC-PeK; PpE-PeC; PpN1-PeJ).

Discussion

The high level of DNA polymorphism detected by the RAPD technique in the genus *Pinus* allowed us to simultaneously construct medium-density linkage maps for the parents of a longleaf pine × slash pine cross using their F_1 progeny as the mapping population. High heterozygosity in conifers has been documented by several investigators (Conkle 1981; Carlson et al. 1991; Devey et al. 1991; Tulsieram et al. 1992; Gerber et al. 1993; Nelson et al. 1993; Nelson et al. in press), and numerous loci might be expected to be segregating in the progeny of any conifer cross. In this inter-specific cross, longleaf pine 3-356 had about 22.3% more heterozygous loci than did slash pine H-28 (166 versus 129 loci, respectively). This excess can be partially explained by the criteria used for initially choosing candidate

primers. Approximately 35% of the primers used in this study were chosen because they were already known to amplify heterozygous loci in longleaf pine 3-356 using a segregating haploid population of megagametophytes (Nelson et al. in press). Preliminary comparisons between the megagametophyte- and F_1-based maps indicates that a number of marker loci were lost in the F_1-based map due to either slash pine H-28 being homozygous band-present for these loci (ie: not segregating in the progeny), or due to poor amplification from the diploid tissues.

The dominant nature of RAPD markers does not appear to present a problem when mapping in highly heterozygous outcrossed species, due to the large number of loci found to be in testcross configuration between parents (Carlson et al. 1991; Grattapaglia et al. 1992; Hemmat et al. 1994). Since two classes of loci are identified (one set heterozygous in one of the parents, and a second set heterozygous in the other parent) parent-specific linkage maps are produced. By taking advantage of the information provided by loci heterozygous in both parents (segregating 3:1) it is possible to infer homologous linkage groups between parents (Hemmat et al. 1994). Combining previously parent-specific linkage groups would greatly increase the number of markers associated with specific groups and the overall map in general. In this study, the number of loci found to be heterozygous in both the longleaf pine and slash pine parents were limited (14 of 281 loci, or 5.0%). A similar study in apple (Hemmat et al. 1994) found only 8.7% of the segregating markers to be heterozygous in both parents (39 of 448 loci). The limited number of loci found to be heterozygous in both parents may be an artifact of the rather diverse

origin of the parents used in these studies, as a preliminary study involving intra-specific crosses of yellow birch (*Betula alleghaniensis*), found five of 14 loci (35.7%) to be heterozygous in both parents (Roy et al. 1992).

Currently, our maps are incomplete as they include 18 groups and 3 pairs (122 markers) for longleaf pine and 13 groups and 6 pairs (91 markers) for slash pine. Based on karyotype analyses, pine species are known to contain 12 (similar sized) pairs of homologous chromosomes (Saylor 1972; Kormutak 1975). The number of linkage groups identified in each of our maps (~ 20) is comparable with previously published findings. Using the megagametophyte system, Nelson et al. (1993) mapped 73 RAPD markers to 13 groups and 9 pairs in slash pine, and in a similar study involving longleaf pine mapped 133 markers to 16 groups and 6 pairs (Nelson et al. in press). Obviously, more markers are required to bridge the gaps between current groupings and expand the coverage towards 100%. Estimates of the number of markers required to obtain 90% coverage of the pine genome (average spacing of 20 to 30 cM) suggest that approximately 200 to 300 markers will be required (Neale and Williams 1991; Nelson et al. 1993). At the current levels of genomic coverage (60-85%), screening more primers for additional polymorphisms will not be efficient. Approaches such as increasing the sample size for terminal and unlinked markers or applying bulked segregant analysis (Michelmore et al. 1991) to search for polymorphisms near the terminal and unlinked markers may prove to be the most efficient means to achieve such ends.

Prior to analyses, we anticipated that selective forces might cause markers to deviate from their expected segregation ratios, most likely as a result of pollen-ovule incompatibility (Richards 1986). There appears to be a lack of consensus among researchers as to whether markers experiencing segregation distortion should be used in linkage studies (Tulsieram et al. 1992; Faure et al. 1993; Nelson et al. 1993). Regardless of the fact that various genetic models (single locus and multi-locus) have been proposed to explain the existence of segregation distortion (Richards 1986; Lin and Ikehashi 1993; Pham and Bougerol 1993), and that the use of loci whose alleles deviate markedly from their expected ratio has no effect on estimates of recombination (Ott 1991), in general, distorted markers have not been included in linkage analyses. Our initial mapping efforts focused only on those loci segregating at their expected Mendelian ratios. However, further linkage analyses were performed which included distorted markers. By including distorted markers we hoped to map additional regions of the parental genomes (those experiencing selective drag), possibly allowing us to further converge our maps towards 12 linkage groups. Of the 20 testcross loci in longleaf pine which were not segregating at the expected 1:1 ratio (n=86, α= 0.05), nine mapped (see markers with a prefix d in the revised linkage groups, Figure 4.2). Likewise, eight of the 14 slash pine testcross loci, displaying significant distortion, mapped (Figure 4.2).

Initial mapping analyses (those excluding distorted markers) identified a total of 21 linkage groups in longleaf pine and 19 linkage groups in slash pine. Including the distorted markers did not allow further convergence of either map. In fact, the

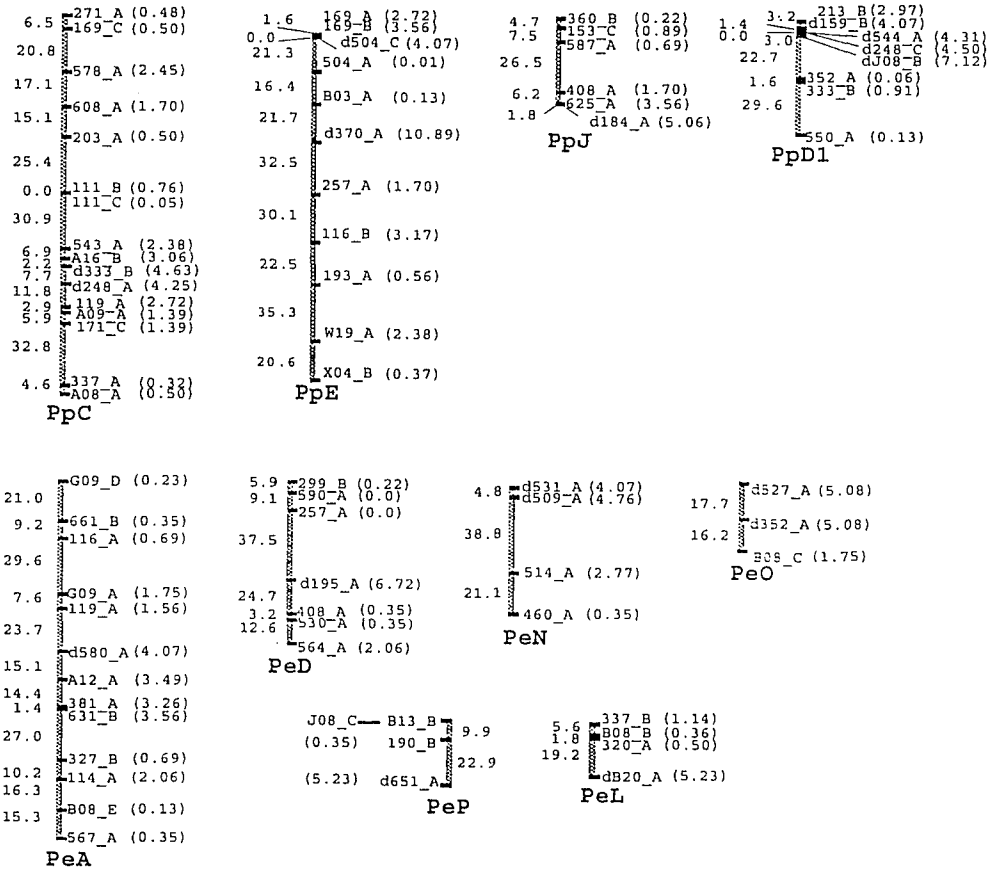

Figure 4.2. Revised maps showing linkage of marker loci displaying significant segregation distortion ($\alpha \leq 0.05$). Marker names and Haldane cM distances are provided. Distorted markers are indicated by a prefix d. Chi-square (χ^2) values are provided in parentheses.

number of linkage groups increased (from 19 to 22 in slash pine) as some of the distorted markers were found to be associated with previously unlinked markers. Inspection of the segregation ratios of markers found to be linked within 10 cM of the distorted markers shows them to be slightly distorted (see χ^2 values, Figure 4.2). The apparent gradation in the level of distortion as one moves along a particular linkage group would seem to reinforce the validity of the suggested linkages.

In terms of genetic mapping, QTL searching, and applying genomic selection in BC_1 populations, we plan to focus on primers that amplify more than one heterozygous locus (within a specific parent, or between parents) as these primers would maximize the efficiency of such applications. Results from this study indicate that multiple RAPD loci amplified by the same primer (within a specific parent) tend to be unlinked. For example, of the 23 primers that amplified more than one heterozygous locus in longleaf pine, five primers amplified loci that mapped to the same linkage group, only two of which were linked at less than 20 cM (loci 111_B-111_C at 0 cM, PpC; loci 169_A-169_B at 3.6 cM, PpE). Of the 19 primers that amplified more than one heterozygous locus in slash pine, only one amplified loci linked at less than 20 cM (680_A-680_B at 0 cM, PeM). These results appear to be similar to those obtained in other studies (Tulsieram et al. 1992; Nelson et al. 1993; Hemmat et al. 1994). Utilization of such "multi-locus" primers would simultaneously maximize the number of informative loci available for linkage applications, while minimizing the total number of RAPD reactions required.

In our mapping population, we found the number of heterozygous loci per F_1 individual to range from a low of 96 to a high of 130. Therefore, we might expect to have segregation information for a maximum of 96-130 mapped loci available for use in BC_1 populations. In addition, assuming that each RAPD band represents a single genetic locus, the 87 polymorphic (between parents) but non-segregating loci should be heterozygous in the F_1 progeny. Of course, the actual number of loci useful for genetic mapping and QTL searching in BC_1 populations will depend upon the F_1 individual chosen, and the genotype of the recurrent parents. Our plans are to focus on those F_1 individuals (intermediate for early height growth) that are heterozygous for a maximum number of marker loci. In order to avoid possible complications that might arise as a result of inbreeding (poor seed set and reduced seedling vigor), we plan to use unrelated recurrent parents in our backcross pedigrees (Bernatzky and Mulcahy 1992). Potential recurrent parents will be genotyped at all loci known to be heterozygous in the selected F_1 parent. Those recurrent parents that are homozygous band-absent at the most loci, will be selected.

To identify loci influencing EHG we plan to produce and test two divergent backcross families. In one family, a longleaf pine will be used as the recurrent parent, and in the other family, a slash pine will be used as the recurrent parent. Use of the same F_1 individual as the male parent in each of the backcross families should allow for the identification of EHG loci in the most comparable genetic background (with positive effect EHG alleles coming from the hybrids' slash pine parent and negative effect (grass stage) EHG alleles coming from the hybrids' longleaf pine

parent). By analyzing both backcross families, we expect to find "real" (not false positive) EHG loci as they should map to the same marker loci in both families.

Successful completion of this research should result in the development of longleaf pine hybrid genotypes that exhibit vigorous early height growth as a result of their harboring high numbers of positive effect EHG alleles. These genotypes could be selected for further backcrossing, clonally propagated and used directly in production plantings, or crossed to fix loci influencing EHG. Selected progeny, fixed at EHG loci, could then be used as parents in seed orchards as well as in further backcrosses.

CHAPTER 5

COMPARISON OF RAPD LINKAGE MAPS CONSTRUCTED FOR A SINGLE LONGLEAF PINE FROM BOTH HAPLOID AND DIPLOID MAPPING POPULATIONS

Introduction

Linkage studies in the *Pinaceae* have traditionally been conducted with isozyme markers (Guries et al. 1978; Rudin and Ekberg 1978; Adams and Joly 1980; Conkle 1981; King and Dancik 1983; Cheliak and Pitel 1985; Furnier et al. 1986; Harry 1986; Strauss and Conkle 1986; Muona et al. 1987; Niebling et al. 1987; Perry and Knowles 1989; Adams et al. 1990; Xie et al. 1991). However, the paucity of available isozyme markers and their low level of polymorphism, have precluded the construction of detailed (saturated) genetic maps. Two-dimensional (2-D) electrophoresis and staining of total proteins has significantly increased the number of protein polymorphisms that can be identified (Anderson et al. 1985; Bahrman and Damerval 1989). Using 2-D techniques, Gerber et al. (1993) placed 65 protein loci identified in megagametophytic tissues of maritime pine (*Pinus pinaster* Ait.) into 10 groups (3+ loci) and 7 pairs, covering 530 cM. Karyological studies have revealed that pine species contain 12 similar-sized pairs of homologous chromosomes (Saylor 1972; Kormutak 1975), and current genome size estimates based on partial genetic linkage data vary from 2300-3300 cM (Neale and Williams 1991; Nelson et al. in press). Although 2-D electrophoretic techniques have significantly increased the number of available protein loci (approximately three-fold), protein markers alone will not allow for effective coverage of the entire pine genome.

In contrast to protein markers, DNA markers are potentially unlimited in number as they allow direct access to coding and non-coding regions of the genome (Soller and Beckmann 1983). Several methods have been developed to detect DNA polymorphisms. The most common approach applies restriction fragment length polymorphisms (RFLPs) which use conserved or variable regions of the genome as DNA probes. While genetic maps based on RFLPs have been constructed for a number of agronomic crop species (for example: Bernatzky and Tanksley 1986; Helentjaris et al. 1986; Gebhardt et al. 1989; Landry et al. 1987; McCouch et al. 1988; Havey and Muehlbauer 1989; Keim et al. 1990), progress in long-lived perennial tree species lags behind. The use of RFLPs to map conifer genomes has recently been proposed (Nance and Nelson 1989; Neale and Williams 1991), and is currently underway for loblolly pine (*Pinus taeda* L.) where over 100 RFLP markers have been developed and are starting to be used for linkage analyses (Devey et al. 1991; Colby et al. 1993; Groover et al. 1993).

Several factors have delayed the widespread use of RFLPs in genetic linkage studies in the *Pinaceae*. The large physical size of pine genomes (2C = 33 to 57 pg, Ohri and Khoshoo (1986)), can complicate the autoradiographic detection of single-copy genes (Ausubel et al. 1987; Neale and Williams 1991; Shattuck-Eidens et al. 1992). Due to the long generation times associated with long-lived tree species, there is also a lack of well-defined pedigrees (beyond two generations) with which to work. Possibly as a result of such complications, only a limited number of tree species have been mapped. Two RFLP-based genetic maps for citrus (*Citrus spp.*)

(Durham et al. 1992; Jarrell et al. 1992), one for apple (*Malus domestica*) (Weeden and Hemmat 1993), and one for quaking aspen (*Populus tremuloides* Michx.) (Liu and Furnier 1993) have recently been published. In comparison to pines, these species have relatively small genomes [for citrus 2C = 1.24 pg (Guerra 1984); and for aspen 2C = 1.6 pg (Dhillon et al. 1984)]. Quaking aspen also has the added advantage of early floral development (USDA 1974) facilitating pedigree advancement.

A new technique for generating DNA markers, commonly referred to as the random amplified polymorphic DNA (RAPD) technique, the arbitrary primed polymerase chain reaction (AP-PCR), or the DNA amplification fingerprinting (DAF) technique, has recently been developed (Williams et al. 1990; Welsch et al. 1990; Caetano-Anolles et al. 1991). Due to their rapidity and simplicity of detection, RAPDs are making quick advances in genetic mapping possible. The method applies the polymerase chain reaction (PCR) with a short oligonucleotide primer, randomly amplifying fragments of template DNA (Williams et al. 1990). The procedure requires only very small amounts of template DNA, and can rapidly amplify sequences that are inherited in a Mendelian fashion. The RAPD procedure holds great promise for quickly placing markers on linkage groups, even for species with large genomes such as conifers (Grattapaglia et al. 1992; Tulsieram et al. 1992; Nelson et al. 1993).

An advantage of the RAPD technique with conifers is the use of haploid DNA, which is available in the megagametophyte of conifer seeds. The

megagametophyte is derived from repeated mitotic divisions of a single meiotic product (USDA 1974), and sufficient DNA for several hundred to several thousand RAPD reactions can be extracted from a single megagametophyte (Tulsieram et al. 1992; Nelson et al. 1993). The use of haploid tissues is beneficial because the RAPD technique does not readily differentiate between heterozygotes (+/-; + = band present, - = band absent) and dominant homozygotes (+/+) in diploid individuals. Low- to medium-density RAPD maps have recently been constructed for several conifer species such as white spruce (*Picea glauca* Voss) (Tulsieram et al. 1992), loblolly pine (*Pinus taeda* L.)(Grattapaglia et al. 1992), slash pine (*Pinus elliottii* Engelm.) (Nelson et al. 1993), and longleaf pine (*Pinus palustris* Mill.) (Nelson et al. in press) using the haploid megagametophyte system.

In contrast to the haploid system, a strategy employing diploid progeny from a specific cross provides data on both parents simultaneously (Grattapaglia and Sederoff 1994; Hemmat et al. 1994) The efficiency of such an approach, however, is limited by the number of loci found to be in testcross configuration between parents (+/- × -/- or -/- × +/-). Testcross locus configurations appear to be quite common in matings between highly-heterozygous outcrossed tree species (Carlson et al. 1991; Roy et al. 1993), and have been used to construct maps for the parents of crosses between the banana (*Mus acuminata*) cultivars "SF265" × "Banksii" (Faure et al. 1993), the apple (*Malus domestica*) cultivars "Rome Beauty" × "White Angel" (Hemmat et al. 1994), by selfing F_1 progeny of the peach (*Prunus persica* (L.) Batsch) cultivars "NC174RL" × "Pillar" (Chaparro et al. 1994), *Eucalyptus grandis* ×

Eucalyptus urophylla (Grattapaglia and Sederoff 1994), and *Pinus elliottii* × *Pinus palustris* (Kubisiak et al. in press).

Preliminary comparisons among species-specific RAPD maps suggest that bands of similar molecular weight amplified by the same primer often map to discrete, unlinked locations in different genotypes (Johns 1992; Nance et al. 1992a; Kesseli et al. 1994). This is not surprising as RAPD amplification products can often contain (or are contained within) repetitive DNA sequences (Williams et al. 1993; Kazan et al. 1992; Paran and Michelmore 1993; Kesseli et al. 1994). Although some RAPD loci of the same molecular weight amplified by the same primer may not map to the same location in different genotypes, the linkage relationship of loci within a single genotype should be conserved regardless of the mapping population used for linkage analyses (haploid tissues or F_1 progeny). In the present study, we examined the linkage relationship (orders and distances) among heterozygous RAPD loci identified in longleaf pine 3-356 based on their inheritance in both a haploid megagametophyte mapping population (Nelson et al. in press) as well as in a slash pine H-28 (♂) × longleaf pine 3-356 (♀) diploid F_1 mapping population (Kubisiak et al. in press).

Materials and Methods

Plant Material

Megagametophytes of longleaf pine clone 3-356 were dissected from wind-pollinated seeds, and total DNA prepared as outlined in Nelson et al. (1994). Needle tissue from F_1 progeny of a cross between slash pine H-28 (♂) × longleaf pine 3-356

(♀) were collected, and total DNA prepared as outlined in Kubisiak et al. (in press). In total, the haploid mapping population included 88 megagametophytes (80 used in Nelson et al. (in press), plus 8 additional megagametophytes), and the diploid population included the 86 F_1 progeny originally used in Kubisiak et al. (in press).

Primer selection and DNA amplification

A total of 71 primers that had previously identified heterozygous loci in longleaf pine 3-356 (using both the haploid megagametophyte- and diploid F_1-based systems (Nelson et al. in press; Kubisiak et al. in press) were selected for this study. Primer DNAs were obtained from either Operon Technologies Inc. (Alameda, CA) or J.E. Carlson (Univ. of British Columbia, Vancouver, B.C., Canada). In order to ensure that the same size fragments were being scored in both populations, an additional 8 megagametophytes and 14 F_1 progeny [fourth template set employed in Kubisiak et al. (in press)] were amplified and subsequently separated in the same agarose gel. Amplification of template DNAs followed the protocol outlined in Nelson et al. (in press), with the only modification consisting of a doubling of the diploid template DNAs to 6.25 ng per reaction.

Segregation analysis

Linkage analysis was performed on the full haploid data set (original data plus the additional eight megagametophytes) using the program MAPMAKER/EXP version 3.0. The mapping procedures followed were the same as those outlined in Kubisiak et al. (in press), except that a slightly reduced log of the odds (LOD)

threshold was tolerated (LOD < 2.0) in order to maximize the number of marker loci available for comparative purposes.

Results

Amplification and separation of the additional haploid DNAs in the same agarose gel facilitated the scoring of nine additional loci not previously included in the haploid data set (Nelson et al. in press). This increased the total number of heterozygous loci available for linkage analyses to 183. The additional megagametophytic data increased the size of the mapping population by approximately 10% (from 80 to 88 megagametophytes).

Current linkage analyses suggest a genetic map consisting of 14 groups and two pairs (162 loci) spanning a total of 2023.3 cM. Twenty-nine additional markers were mapped, increasing the coverage by 388 cM. Of the 29 additional loci mapped, 6 were new marker loci and 23 were previously unordered in Nelson et al. (in press) as they were found to have two or more likely (LOD < 2.0) positions. As long as inconsistencies in linkage were not detected (e.g. a locus could just as likely be placed in two or more distantly separated locations) loci were placed in their most likely position. Although some incorrect marker orders might be declared (especially in the case of tight linkage), this comprehensive approach allowed us to increase the number of marker loci available for comparative purposes.

The 18 groups and 6 pairs identified in Nelson et al. (in press) coalesced to 14 groups and 2 pairs. As a result of the additional data, linkage was now suggested between the previously designated groups A and O, H and P, I and K, and an

additional linkage group (previously a linked pair, now designated group T) was formed (Figure 5.1). Three of the remaining four linked pairs identified in Nelson et al. (in press) mapped to terminal locations of three different linkage groups (Figure 5.2). The 71 primers selected for this study, identified a total of 148 heterozygous loci in longleaf pine 3-356 based on amplification of megagametophytic DNAs. Of these 148 loci, 133 loci (89.9%) were grouped into 14 groups (3 or more loci) and two pairs, four loci (2.7%) were grouped but could not be confidently ordered (even using relaxed LOD scores of < 2.0), and 11 loci (7.4%) remained unlinked. Of the 137 loci that could be grouped based on two-point analyses, 62 loci (45.3%) could not be scored when DNAs from F_1 progeny were amplified. Twenty-four of these 62 loci (38.7%) did not amplify from F_1 DNA as they appeared to be outside a window of consistent amplification for diploid DNAs, eight loci (12.9%) were contained within this window but did not amplify, and 30 loci (48.3%) could not be assessed due to the presence of slash pine bands of similar molecular weight that precluded accurate scoring. Of the 75 loci that were scorable, 26 (34.7%) were fixed or "non-segregating" (assuming slash pine H-28 to be homozygous band-present at the same locus), and 49 loci (65.3%) were segregating (42 1:1, and seven 3:1). Comparisons were made using the 49 loci common to both the haploid and diploid maps.

The 49 loci allowed us to determine homologous counterparts between maps. Seven homologous linkage groups (A-PpA, B-PpB, C-PpC, D-PpD1 and PpD2, E-PpE, N-PpN, O-PpO) were characterized by three or more loci (Table 5.1). Group D could not be correctly oriented (with respect to order) with either PpD1 or PpD2.

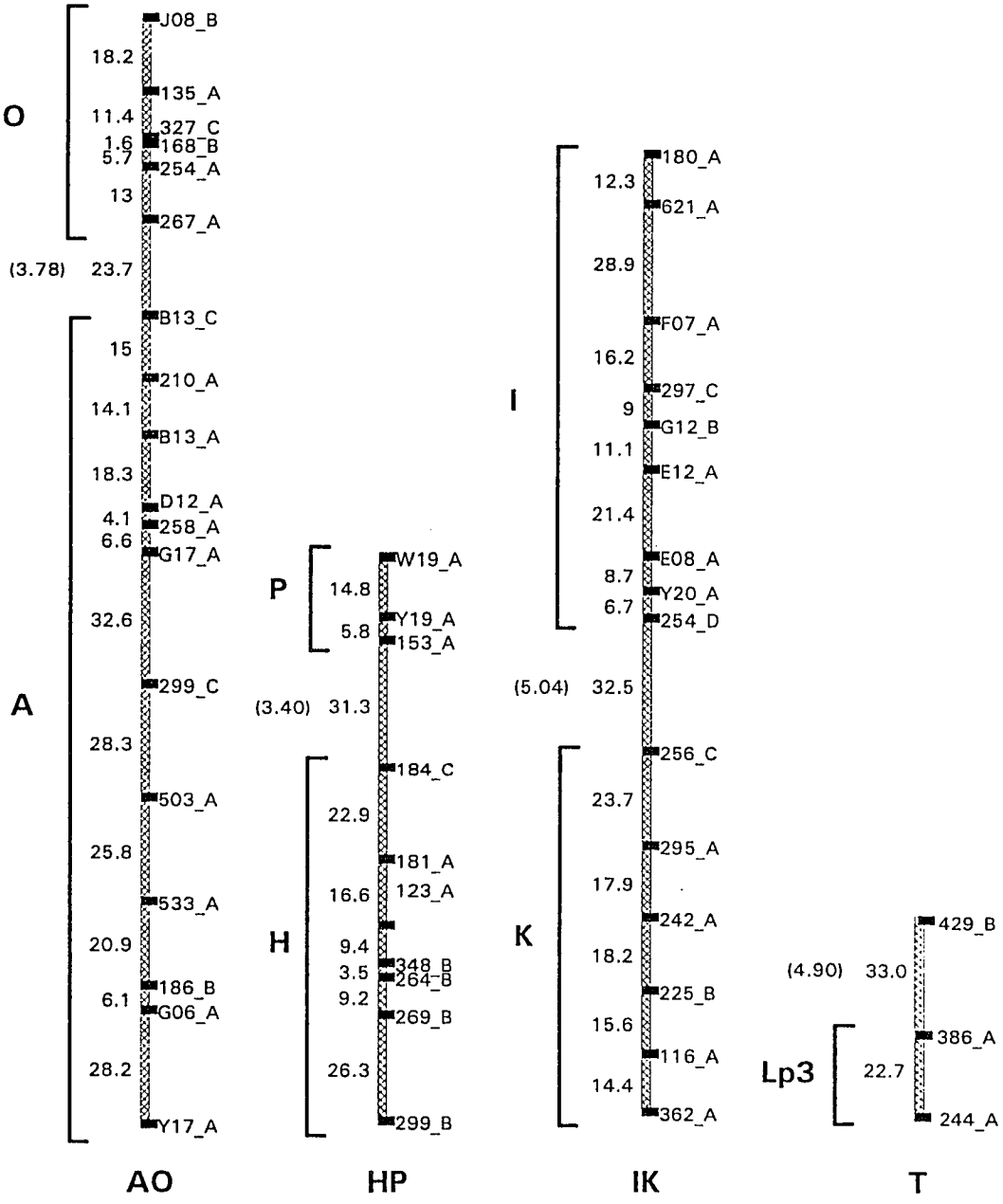

Figure 5.1. Convergence of linkage groups A and O, H and P, I with K, and an additional linkage group (previously a linked pair, now designated group T) identified in Nelson et al. (in press), as a result of additional megagametophyte data. Primer names and Haldane centiMorgan (cM) distances are provided. Log of the odds (LOD) ratios between groups are displayed in parentheses.

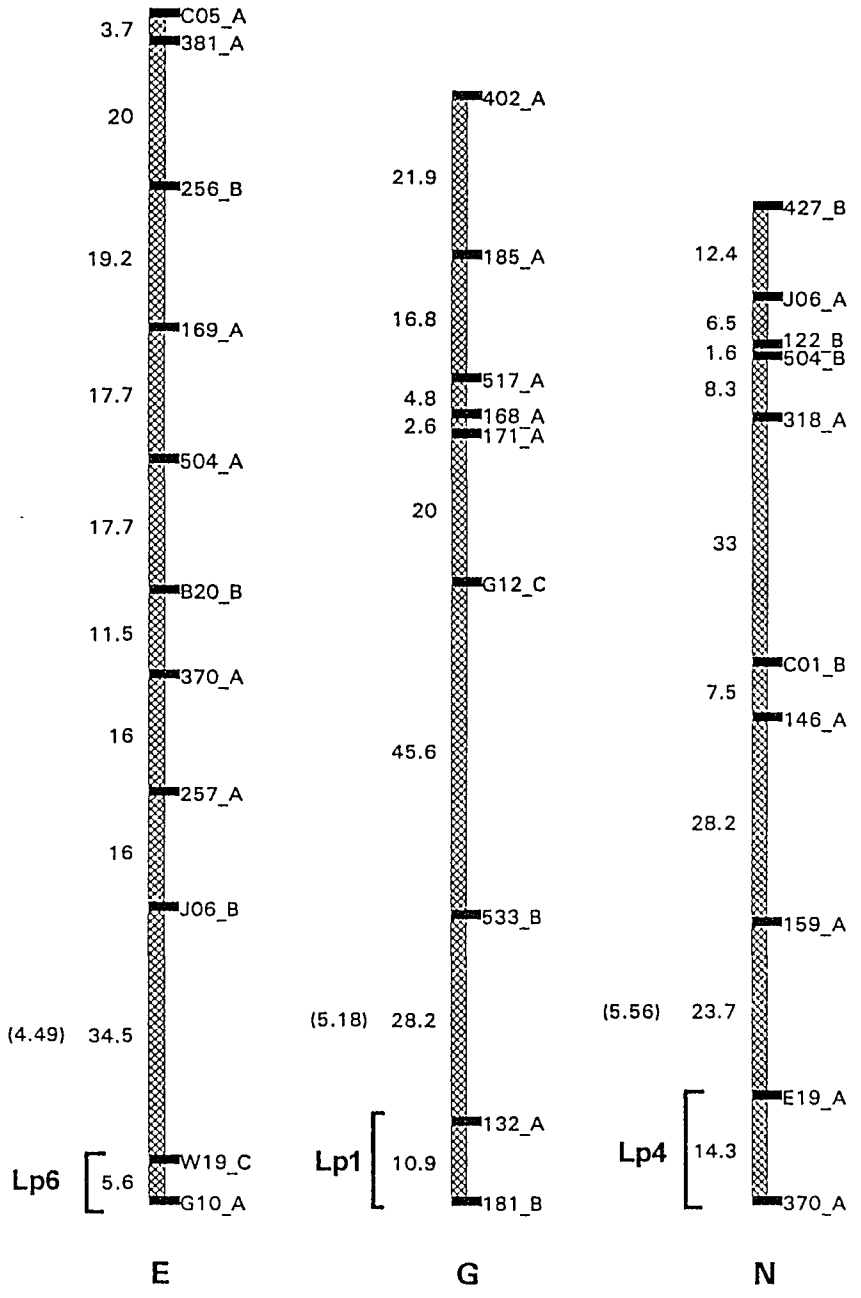

Figure 5.2. Convergence of linkage groups and linked pairs E and Lp6, G and Lp1, and N with Lp4 identified in Nelson et al. (in press), as a result of additional megagametophyte data. Primer names and cM distances are provided. LOD scores are displayed in parentheses.

PpD1 had only one marker in common, and one of the two markers in common between D and PpD2 (although grouped using two-point analysis) could not be confidently ordered using three- and multi-point analyses (Table 5.1). Five groups and one linked pair (F-PpF, G-PpG, H-PpH, I-PpI, J-PpJ, Lp3-PpT) were found to have two markers in common (Table 5.1). Single loci in the linked pairs Lp2 and Lp5 suggested homology with groups PpQ and PpS (Table 5.1). Seven of the comparable loci were found to segregate in a 3:1 Mendelian ratio in the slash pine H-28 (♂) × longleaf pine 3-356 (♀) F_1 mapping population, therefore genetic distances were not calculated. However, their association with markers on homologous groups was confirmed by χ^2 analyses (Kubisiak et al. in press) [see Table 5.1 for specific loci]. Three loci identified in the F_1 mapping population were grouped using two-point analyses, but could not be confidently ordered (Table 5.1). In total, 31 loci segregating (1:1) were placed on each map and used for comparative purposes. These 31 loci allowed us to align 10 homologous linkage groups (designated superscript H and D for haploid- and diploid-based linkage groups; Figure 5.3).

Although the order of loci in common between the two maps was found to be conserved, there appeared to be some fairly large discrepancies in terms of genetic distance estimates between linked loci (Table 5.1 and Figure 5.3). A paired comparison test (paired t-test) was utilized to determine whether the difference in genetic distance estimates between linked loci was significantly different from zero. In other words, this test was used to determine if one gametic subset was consistently over-estimating or under-estimating genetic distances between loci. The paired

comparison test resulted in an insignificant t-value ($t_{0.025,19}$ = -0.35803). Therefore there was no evidence to indicate that the two gametic subsets produced consistently (significantly) different genetic distance estimates.

Discussion

Overall, a fairly low percentage of loci that mapped in the haploid megagametophyte population [49 of 137 comparable loci (35.7%)] were also segregating either 1:1 or 3:1 in the diploid population. There were two reasons that might explain such a low percentage of comparable loci. First, 62 of the 137 loci (45.2%) mapped in the haploid population could not be scored when DNAs from F_1 progeny were amplified. The window of amplification (based on band size in nucleotide base pairs) appeared more restricted when F_1 DNAs were amplified, even though extension times were equivalent (2 min). In general, the higher molecular weight loci identified in Nelson et al. (in press) were not successfully amplified in the diploid F_1. This artifact could potentially be due to the fact that additional RAPD loci are simultaneously being amplified in DNAs derived from diploid versus haploid tissues. The presence of additional priming sites would only enhance primer competition during the initial cycles of amplification (Newbury and Ford-Lloyd 1993; Caetano-Anolles et al. 1994). Secondly, additional bands amplified from slash pine (the other genomic complement) precluded accurate scoring of some loci of interest, as they occurred at similar molecular weights.

Table 5.1. RAPD loci in common between homologous linkage groups constructed from separate gametic subsets (haploid and diploid mapping populations) of longleaf pine 3-356. Groups designated with a single letter (i.e. A) were mapped using haploid megagametophytic DNAs from wind-pollinated seed (Nelson et al. 1994) and those with three letters (i.e. PpA) were identified in slash pine H-28 (♂) × longleaf pine 3-356 (♀) F_1 DNAs (Kubisiak et al. in press). Distances between comparable loci are indicated in Haldane centiMorgans (cM). Those primer names designated with a three digit number were obtained from J.E. Carlson (Univ. of British Columbia, Vancouver, B.C., Canada), and those designated with a letter followed by a two digit number were obtained from Operon Technologies Inc. (Alameda, CA).

HAPLOID MAP		DIPLOID MAP				
Group & Primer	Distance in cM	Group & Primer	Distance in cM	(3:1) Loci	Pair	Unordered Loci
GROUP A		GROUP PpA				
210_A	32.4	210_A	18.4			
D12_A	4.1	D12_A	1.6			
258_A		258A				
GROUP B		GROUP PpB				
327_B	37.9				327_B	
362_B	4.1			362_B		
B04_B	12.9	B04_B	34.2			
A07_A	22.0			A07_A		
F07_C	12.9	F07_C	13.5			
306_A	1.2	306_A	5.3			
213_A	22.0	213_A	59.0			
269_C	35.2					269_C
256_A	6.5	256_A				
B13_A				B13_A		
GROUP C		GROUP PpC				
119_A	45.7	119_A	11.8			
248_A	60.5	248_A	73.1			
203_A	42.8	203_A	59.5			
271_A		271_A				
GROUP D		GROUPS PpD1 and PpD2				
159_B	30.5	159_B (PpD1)				
348_A	65.2					348_A
J08_A		J08_A (PpD2)				
GROUP E		GROUP PpE				
169_A	17.7	169_A	21.8			
504_A	29.2	504_A	49.3			
370_A	16.0			370_A		
257_A	50.5	257_A	89.0			
W19_A		W19_A				

(table con'd.)

HAPLOID MAP		DIPLOID MAP				
Group & Primer	Distance in cM	Group & Primer	Distance in cM	(3:1) Loci	Pair	Unordered Loci
GROUP F		GROUP PpF				
258_B	15.8	258_B	25.1			
299_A		299_A				
GROUP G		GROUP PpG				
181_B	39.1	181_B	47.7			
533_B		533_B				
GROUP H		GROUP PpH				
181_A	26.0	181_A	39.0			
348_B		348_B				
GROUP I		GROUP PpI				
297_C	41.5	297_C	25.1			
E08_A		E08_A				
GROUP J		GROUP PpJ				
184_A	25.2	184_A	46.7			
360_B		360_B				
GROUP N		GROUPS PpN1 and PpN2				
370_B	38.0	370_B	13.8			
159_A	28.2	159_A				
146_A	56.9			146_A		
J06_A	12.6	J06_A				
427_B				427_B		
GROUP O		GROUP PpO				
J08_B	31.2	J08_B	18.1			
168_B	5.7	168_B	4.5			
254_A	13.0	254_A				
267_A				267_A		
GROUP Lp2		GROUP PpQ				
216_A		216_A				
GROUP Lp3		GROUP PpT				
386_A	33.0			386_A		
429_B		429_B HAPLOID MAP				
GROUP Lp5		GROUP PpS				
D12_C		D12_C				

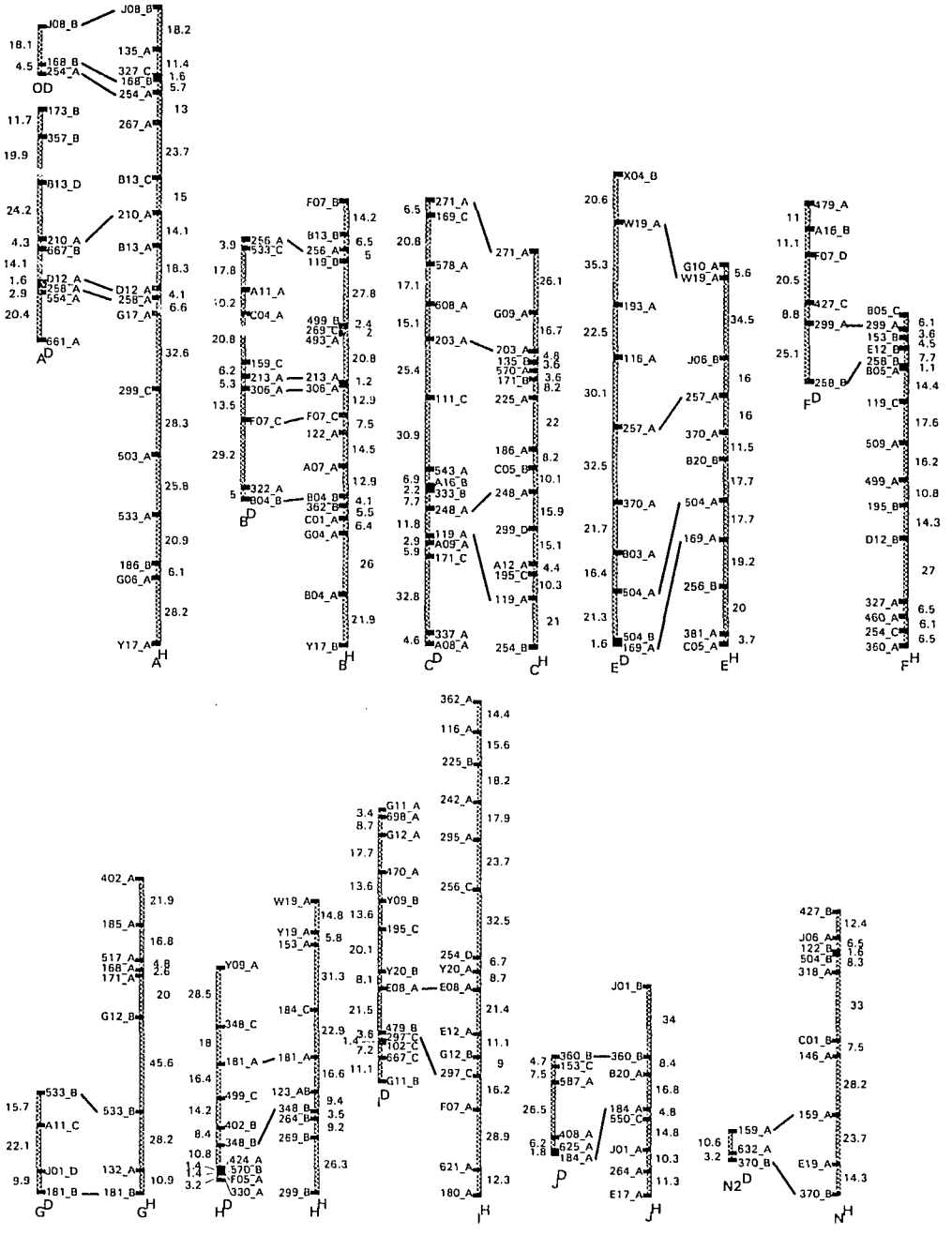

Figure 5.3. Longleaf pine 3-356 haploid-diploid comparison map. Linkage groups designated with a superscript D (i.e. A^D) indicate diploid-derived groups, those indicated with a superscript H indicate haploid-derived groups. Primer names and cM distances are provided. Markers in common between maps are indicated with heavy lines. Loci segregating 3:1 in the diploid mapping population are not displayed.

Although there was no evidence to suggest that the two gametic subsets derived from longleaf pine 3-356 produced consistently different genetic distance estimates, some of the differences between loci (mapped to the same linkage group) were rather large (Table 5.1 and Figure 5.3). The two linkage maps for longleaf pine 3-356 were not constructed from the megagametophytes and embryos (plants) contained within the same seeds. Such an approach would provide the same genetic distance estimates (providing no missing or misclassified data), as the megagametophytic tissue and embryo contained in a single seed receive the same genetic complement from the maternal parent (USDA 1974). As a result, variation associated with random gametic sampling might partially explain such differences in genetic distance estimates (Sall and Bengtsson 1989; Sall and Nilsson 1994). Such intra-individual heterogeneity in recombination has been observed in several species (Jones 1987).

Factors that distort segregation can also influence genetic distance estimates. It is obvious that misclassifications may disturb the results of linkage studies. Misclassifications systematically lead to a positive bias in map distances (Shields et al. 1991; Lincoln and Lander 1992). In mapping experiments using DNA markers, the risk of misclassification is especially pronounced because large data matrices are produced, subsequently interpreted, scored, and logged into a computer in several steps. Furthermore, for RAPD data where reproducible results require extreme experimental care, misclassifications must seriously be considered. Weeden et al. (1992) reported error rates as high as 10% in segregating pea and apple populations,

but with extra care such as using only high quality template DNA and scoring only clear polymorphisms they were able to keep their error rate below 4%. However, an error rate of 4% can result in substantial differences in genetic distance estimates, especially with limited sample sizes (Kubisiak et al. 1993; Sall and Nilsson 1994).

Segregation distortion at some loci may result due to a reduced viability of some of the resulting phenotypes. Significant deviations from normal Mendelian ratios are detected with the χ^2 test. It has been demonstrated that when only one locus shows distorted segregation the expectation of the recombination estimate is not biased, only the variance of the estimate is influenced (Bailey 1961; Ott 1991; Ritter et al. 1990). However, if more than one locus is experiencing distorted segregation as a result of reduced viability then recombination estimates are influenced. Two linkage groups identified in the diploid map (PpC and PpD1) contained regions to which multiple distorted markers mapped (Kubisiak et al. submitted). This might explain the large discrepancy in distance estimates between comparable loci in the groups C and PpC .

Conservation of marker order between the two maps strengthens the case for using RAPDs in genomic mapping studies. However, the intra-individual discrepancies in genetic distance estimates observed between comparable loci suggests that RAPDs may not be as efficient as other markers. This may prove to have some rather important implications for applications such as QTL searching or map-based cloning. Further refinements will no doubt be required to reduce the variation associated with genetic distance estimates. Conversion of RAPD loci to

sequence characterized amplified regions (SCARs) may provide such a means (Paran and Michelmore 1993). SCAR primers generally amplify only a single genetic locus, thus greatly simplifying the kinetics of the reaction. This may help to greatly reduce the possibility of misclassification and the associated inflation of map distances.

CHAPTER 6

CODOMINANT ALLOZYME MARKERS

Introduction

A new technique referred to as the random amplified polymorphic DNA (RAPD) assay has shown great promise for quickly identifying polymorphic markers for use in genetic linkage mapping (Williams et al. 1990). The RAPD technique amplifies DNA sequences by employing the polymerase chain reaction (PCR), and is based on the presence of short inverted DNA repeats. The RAPD technique identifies genetic polymorphisms that are inherited as dominant Mendelian markers. In other words, it is not possible to distinguish a heterozygous band-present genotype from a homozygous band-present genotype.

Application of the RAPD assay on DNAs obtained from haploid tissues has proven to be an extremely successful technique for constructing single-parent genetic linkage maps. Using the haploid tissue contained in conifer seeds (the megagametophyte), low- to medium-density RAPD maps have recently been constructed for several species such as white spruce (*Picea glauca* Voss) (Tulsieram et al. 1992), loblolly pine (*Pinus taeda* L.) (Grattapaglia et al. 1992), slash pine (*Pinus elliottii* Engelm.) (Nelson et al. 1993), and longleaf pine (*Pinus palustris* Mill.) (Nelson et al. in press).

A similar situation to the haploid megagametophyte system can be achieved in diploid tissues when a heterozygous F_1 individual is crossed with a homozygous recessive individual. In tree species, the construction of homozygous recessive testers

is not a practical strategy due to significant genetic loads (Zobel and Talbert 1984). However, as outcrossed tree species are highly heterozygous many loci which fit the "testcross-type" allelic configuration can be readily identified. If one parent in a controlled full-sib cross is heterozygous for a given marker and the other parent is homozygous null, the progeny will segregate 1:1. The opportunity to screen a large number of oligonucleotide primers on both parents and several progeny of controlled crosses for such informative markers provides a major advantage to the RAPD technique. Using this approach two classes of markers are identified; one set heterozygous in one of the parents, and a second set heterozygous in the other parent. Use of these markers results in the creation of two genetic maps, one for each parent (Grattapaglia and Sederoff 1994). As no RAPD loci are common between the parents, it is impossible to determine homologies to integrate the two maps. One way to integrate the parent-specific maps generated by the RAPD technique would be to employ multiallelic codominant markers that are simultaneously segregating in each parent (Ritter et al. 1990).

One multiallelic marker system that has been extensively used to study the genetics of forest tree species has been allozymes. Allozymes are variants of the same enzyme which display differential mobility when subjected to an electric field (Conkle et al. 1982). These alternative forms (alleles) can be identified by staining for enzyme activity in crude tissue extracts subsequent to starch gel electrophoresis. The identification, inheritance, and linkage of allozymes has been described for numerous pine species (Guries et al. 1978; Rudin and Ekberg 1978; Adams and Joly

1980; Conkle 1981; King and Dancik 1983; Cheliak and Pitel 1985; Furnier et al. 1986; Harry 1986; Strauss and Conkle 1986; Muona et al. 1987; Neibling et al. 1987; Perry and Knowles 1989; Adams et al. 1990; Xie et al. 1991). Their well-studied inheritance, ease of detection, and reliability make them an attractive choice for quickly ascertaining homologous linkage groups between the parent-specific RAPD linkage maps. In this study, 13 enzyme systems were employed to identify potentially homologous linkage groups between slash pine H-28 and longleaf pine 3-356 (CHAPTER 4).

Materials and Methods

Plant Material

Forty F_1 progeny from an interspecific cross between slash pine H-28 (♂) × longleaf pine 3-356 (♀) were used as the mapping population. The parents of this cross were previously mapped using RAPD markers (CHAPTER 4). Mortality, or a lack of lateral branches precluded the use of all 86 progeny employed in CHAPTER 4.

Allozymes

Following the instructions set forth in Conkle et al. (1982), enzymatic products from 9 of 13 enzyme systems were resolved in crude extracts obtained from dormant buds (see Table 6.1 for a list of the enzyme systems). Lateral branches were harvested in early February 1994 just prior to shoot elongation. The tissues were stored at 4°C until the day of the electrophoresis run. On the day of the run, bud scales were removed and the inner meristematic dome was excised and ground in

a 1.5 ml microcentrifuge tube using a plastic pestle mounted on a high-speed hand-held electric drill. Four μl of ice-cold extraction buffer was added per mg of meristematic tissue. Due to high levels of endogenous resins and phenolics an extraction buffer consisting of 0.2 M phosphate (pH 7.0), 5% bovine serum albumin, 5% polyvinylpyrrolidone, and a small quantity of 2-mercaptoethanol (1 drop/ 100 ml extraction buffer) were utilized to improve poorly resolved sets of enzyme bands. Both parents and ten progeny were initially screened for all 13 enzymes. Those systems that identified putative polymorphisms were later run against the remaining 30 progeny (three gels; each containing both parents and ten progeny).

Table 6.1 Thirteen allozyme systems scored on F_1 progeny between slash pine H-28 (♂) × longleaf pine 3-356 (♀).

Enzyme	Abbreviation
Aconitase	ACON
Alcohol dehydrogenase	ADH
Alanine aminopeptidase	AAP
Catalase	CAT
Alpha esterase	EST
Fluorescent esterase	FLEST
Glutamate dehydrogenase	GDH
Isocitric dehydrogenase	IDH
Leucine aminopeptidase	LAP
Malic dehydrogenase	MDH
Peroxidase	PER
6-phosphogluconate dehydrogenase	6PGD
Phosphoglucose isomerase	PGI

Electrophoresis

Wicks consisting of thin strips of Whatmann 3MM paper (1 mm x 20 mm) were used to absorb homogenates of each sample. One wick was used for each gel

system to be run. The wicks were loaded in 14% starch gels (Sigma Chemical Company, St. Louis, MO). A slice was made in each gel using a scalpel and the wicks placed at approximately one cm intervals along the entire length. Two wicks with dye marker (0.1% bromophenyl blue) were placed at each end of the gels to monitor migration. Gels were run at approximately 300 volts for 30 min after which the power source was turned off and the paper wicks removed. The run was subsequently continued for an additional 5 to 6 hours. After electrophoresis the gels were sectioned into the appropriate number of slices using monofilament fishing line, placed in plastic stain trays, and temporarily stored in a refrigerator. The staining procedures followed the instructions set forth in Conkle et al. (1982). All potentially polymorphic zymographs were recorded on paper, and subsequently entered as alpha-numeric data into a computer.

Segregation Analysis

Each allozyme locus was tested for goodness of fit to its expected Mendelian segregation ratio by chi-square (χ^2) analysis ($\alpha = 0.05$). The data were entered into the computer package MAPMAKER/EXP (version 3.0) (Lincoln et al. 1992) and analyzed using a modified backcross format (Nelson et al. 1993). The mapping strategy used was similar to those outlined in CHAPTER 4.

Results

Products from four of the 13 enzyme systems studied (AAP, ACON, CAT, and IDH) were either not present in meristematic tissue extracts or were too faint to be accurately scored.

Alcohol dehydrogenase

Two zones of activity were evident on gel slices stained for ADH. The faster migrating zone (ADH_1) consisted of two invariant bands present in both slash pine and longleaf pine. The slower migrating zone (ADH_2) was characterized by single bands of different mobility in each parent. Slash pine was apparently homozygous for a slightly slower migrating band (allele 1), and longleaf pine a slightly faster migrating band (allele 2). The banding phenotype for all the F_1 progeny consisted of three bands, two with similar migration as the parents and a third with intermediate mobility (see Figure 6.1). This banding phenotype is consistent with that expected for a dimeric protein. ADH_2 was also found to be dimeric in knobcone pine (Conkle 1971).

Alpha esterase

Gels stained for EST displayed an array of activity zones. A complex banding pattern consisting of as many as eight bands was visible. An intermediate zone consisting of a series of rather dark staining bands appeared to code for a monomer. Slash pine appeared to be homozygous for a slower migrating allele that was characterized by two bands (allele 1). The double-banded phenotype was similar to that reported for megagametophytes of ponderosa pine (Mitton et al. 1979) and virginia pine (Witter and Feret 1980). Longleaf pine was apparently heterozygous for alleles that were different from the slash pine allele (alleles 2 and 3). Both the longleaf pine alleles were represented by two bands, one band of each allele had similar mobilities (see Figure 6.1).

Fluorescent esterase

A single area of intense activity was evident on gels stained for FLEST. Both slash pine and longleaf pine appeared to be homozygous for alternate alleles as all the F_1 progeny contained a wide area of activity, which was most intense at the center (suggesting that FLEST probably had a dimeric structure). Mitton et al. (1979) found FLEST to be dimeric in ponderosa pine. A second slower-migrating area of activity was noted, however, this region was often entirely absent on some gels.

Glutamate dehydrogenase

Gels stained for GDH displayed only a single zone of activity, with both slash pine and longleaf pine apparently homozygous for the same allele, as no variation was noted in the F_1 progeny.

Leucine aminopeptidase

Two zones of activity were observed on gels stained for LAP. A faster migrating zone (LAP_1) consisting of a wide, darkly-staining band was present in both slash pine and longleaf pine. A slower migrating zone (LAP_2) appeared to code for a segregating monomeric protein, consistent with that reported for knobcone pine (Conkle 1971), scots pine (Rudin 1977), and Norway spruce (Lundkvist and Rudin 1977). Slash pine possessed both a fast and slow migrating allele (alleles 1 and 2), longleaf pine was apparently homozygous for the fast allele (allele 2) [see Figure 6.1].

Malate dehydrogenase

Gels stained for MDH appeared to display two zones of activity. A faster migrating zone (MDH_1) consisted of a broadly-staining region. It appeared as if this region might possibly consist of two or more bands, however, the relative migration of these bands was so similar as to preclude its usefulness. The slower migrating zone (MDH_2) appeared to code for a dimeric protein, consistent with that reported for pitch pine (Guries and Ledig 1978). Slash pine was homozygous for a slow migrating allele (allele 1), and longleaf pine was heterozygous for both fast and slow migrating alleles (alleles 1 and 2) [see Figure 6.1].

Peroxidase

Two zones of activity were visualized on gels stained for PER, one in the anodal slice and one in the cathodal slice. In the anodal region slash pine stained for four bands and longleaf three, however, no segregation was observed in the F_1 progeny. The banding pattern observed in the cathodal region consisted of up to eight bands, several of which were extremely faint. The pattern was extremely complex and precluded an accurate assessment of parental and F_1 genotypes.

Phosphoglucose isomerase

Gels stained for PGI displayed two regions of activity. The faster migrating region (PGI_1) appeared as a single faint band in both slash pine and longleaf pine. The mobility of these bands appeared to be slightly different (<1mm), however, all the progeny contained a similar apparently single-banded phenotype. The slower migrating region appeared as a series of dark-staining bands. Both slash pine and

longleaf pine appeared to be heterozygous for different alleles (slash pine alleles 1 and 3; longleaf pine alleles 2 and 4). The banding pattern observed for PGI_2 suggests that it is dimeric (see Figure 6.1), as has been observed for ponderosa pine (Mitton et al. 1979) and pitch pine (Guries and Ledig 1978).

6-phosphogluconate dehydrogenase

Two zones of activity were evident on gels stained for 6PGD. A faster, darkly-staining zone (6PGD_1) appeared to code for a dimeric protein [as was noted

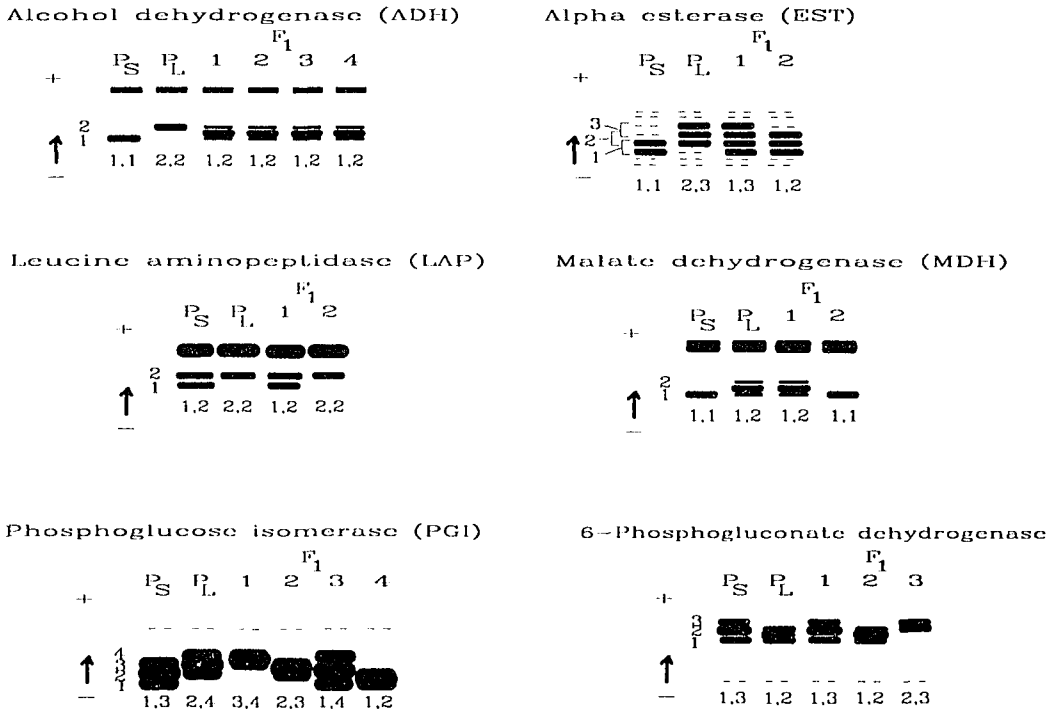

Figure 6.1. Banding patterns for six polymorphic allozyme loci identified in a slash pine H-28 (♂) × longleaf pine 3-356 (♀) cross.

for ponderosa pine (Mitton et al. 1979)]. Both slash pine and longleaf pine were apparently heterozygous, each harboring one common allele and one uncommon allele (slash pine alleles 1 and 3; longleaf pine alleles 1 and 2) [see Figure 6.1]. The slower, faintly-staining zone was not consistent from gel to gel (i.e. the parents sometimes stained on one gel and did not stain when run on another gel).

Linkage Analysis

Each allozyme locus was tested for goodness of fit to its expected 1:1 Mendelian segregation ratio by chi-square (χ^2) analysis (see Table 6.2). None of the loci was found to deviate significantly from their expected 1:1 ratio. Initially, linkage between/among all allozyme loci common to a specific parent was examined. None of the polymorphic allozyme loci identified in slash pine H-28 (LAP, PGI_2, or 6PGD_1) was found to be linked based on two-point analyses. Likewise, no linkage was suggested between/among any of the polymorphic allozyme loci identified in longleaf pine 3-356 (EST, MDH, PGI_2, 6PGD_2).

Table 6.2. Chi-square (χ^2) test for information for allozyme loci identified in both slash pine and longleaf pine, based on their inheritance in 40 F_1 progeny.

		Allele[1]				
Parent	Locus	F	S	Expected	χ^2 value	Prob
Slash	LAP_2	18	22	20	0.40	0.51
	PGI_2	17	22	19.5	0.64	0.42
	6PGD_1	18	21	19.5	0.23	0.63
Longleaf	EST	18	22	20	0.40	0.51
	MDH_2	17	23	20	0.90	0.34
	PGI_2	20	20	20	0.0	1.0
	6PGD_1	23	16	19.5	1.26	0.26

[1]F = fast migrating band; S = slow migrating band

In total, six of the seven allozyme loci mapped. In slash pine, PGI_2 and LAP mapped to internal regions of linkage groups PeE and PeI, respectively (see Figure 6.2). The locus encoding for 6PGD_1 in slash pine was found to be unlinked. All four of the polymorphic allozyme loci identified in longleaf pine mapped. The locus encoding for PGI_2 mapped to an internal region of linkage group PpH, suggesting homology between linkage group PeE in slash pine and PpH in longleaf pine. The remaining three loci EST, 6PGD_1, and MDH mapped to the linkage groups PpD1, PpG, and PpS, respectively. None of the allozyme loci allowed any further convergence towards the haploid chromosome number (n=12) in either map.

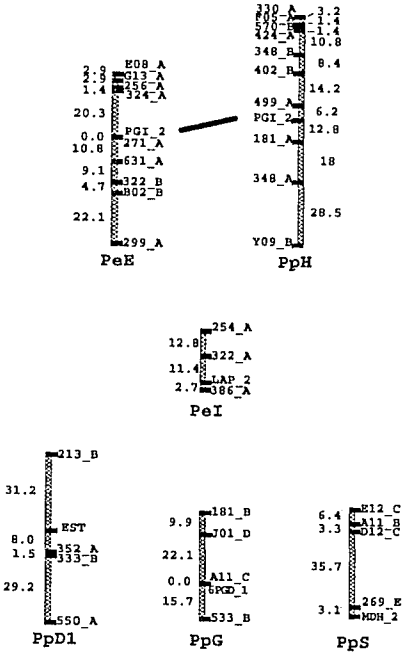

Figure 6.2. Linkage relationship of five different allozyme loci (PGI_2, LAP_2, EST, 6PGD_1 and MDH_2) to RAPD loci identified in the parents of a slash pine H-28 (♂) × longleaf pine 3-356 (♀) cross. Linkage groups are designated Pe_ for slash pine, and Pp_ for longleaf pine. Haldane centiMorgan (cM) distances are provided.

Discussion

The goal of this research was to identify allozyme loci that were polymorphic in both longleaf pine 3-356 and slash pine H-28, and to use these loci to identify homologous linkage groups between the parent-specific maps produced by the RAPD technique. Unfortunately, only five heterozygous loci were identified, two of which were heterozygous in both parents (PGI_2 and 6PGD_1). Locus 6PGD_1 could not be mapped in slash pine which precluded its usefulness for identifying homology between the maps. The locus PGI_2, however, suggested homology between the linkage groups PpH in longleaf pine 3-356 and PeE in slash pine H-28 (see Figure 6.2).

Although the allozyme systems employed in this study were not very informative in terms of identifying homologous linkage groups between longleaf pine 3-356 and slash pine H-28, they can be compared to linkage groups identified in various other pine species by Conkle (1981). In our analysis an esterase locus mapped to longleaf pine linkage group PpD1. Conkle (1981) mapped two esterase loci, one to a linkage group designated A (EST-1), and a second to linkage group B (Est-4). As only one locus could be unambiguously scored on our gels, we could not definitively identify which linkage group (A or B) PpD1 might be homologous with. However, Conkle (1981) found locus EST-4 to map to the same linkage group as locus PGI-2 (linked at approximately 44.9 cM). We mapped EST and PGI_2 to different linkage groups (PpD1 and PpH, respectively). This might suggest that the esterase locus we identified is EST-1 (Conkle 1981), or that the longleaf pine linkage

groups PpD1 and PpH are really part of the same chromosome and that current levels of genomic coverage have precluded their convergence. The allozyme locus LAP_2 which mapped to linkage group PeI in slash pine H-28 suggests that this linkage group is homologous to linkage group A identified in Conkle (1981). PGI_2 and LAP_2 were both found to map to linkage group A (approximately 57 cM apart) in Conkle (1981), suggesting that PeE and PeI may be part of the same chromosome. The loci 6PGD_1 and MDH_2 could not be mapped in any of the pine species studies in Conkle (1981), therefore, these loci were not useful for comparative purposes.

It is clear that allozymes will not provide the necessary number of loci required to efficiently integrate the parent-specific RAPD maps constructed for longleaf pine 3-356 and slash pine H-28. Other sources of data will need to be developed in order to serve as bridges. Codominant markers such as RFLPs or minisatellites will likely be the most logical choice for such an endeavor as these markers are often highly polymorphic [i.e. multiallelic] (Devey et al. 1991; Nybom and Rogstad 1990; Condit and Hubbell 1991; Broun et al. 1992; Thomas and Scott 1993).

CHAPTER 7

EARLY HEIGHT GROWTH-QTL MAPPING

Introduction

Many traits of agronomic and economic importance in trees result from the combined action of multiple genes and the environment. These include such traits as rotation age height and volume, or various wood quality characteristics such as specific gravity and fiber length (Zobel and Talbert 1984). Polygenic traits exhibit phenotypes with continuous distributions that are difficult to analyze.

Sax (1923) introduced the concept of using qualitative genes to locate genes of lesser effect controlling quantitative traits. Thoday (1961) used single gene morphological markers to conduct detailed studies of quantitative traits in *Drosophila melanogaster*. More systematic attempts to resolve quantitative traits into their individual genetic components were initially limited by a lack of polymorphic qualitative markers that could cover large parts of the genome. These limitations have been partially overcome by the use molecular markers such as allozymes (Tanksley et al. 1982), restriction fragment length polymorphisms (RFLPs) (Paterson 1988), and more recently by random amplified polymorphic DNAs (RAPDs) (Williams et al. 1990).

The traditional approach to QTL mapping consists of simply comparing the phenotypic means of the two to three genotypic classes for each marker (Soller et al. 1976; Tanksley et al. 1982). This technique has proven to be useful for mapping traits (Keim et al. 1990; Diers et al. 1992; Stuber et al. 1992; Young et al. 1993),

however, there are many problems with this method of mapping QTLs. The method cannot tell whether the markers are closely linked to a small-effect QTL or loosely linked to a major effect QTL, nor can it tell whether the markers are associated with multiple QTLs. The effects of QTLs are likely to be underestimated because they are confounded by recombination frequencies, and because of this confounding effect the technique is not very powerful and many individuals may be required.

In an attempt to take advantage of the information provided by genetic maps, Lander and Botstein (1989) introduced interval mapping. This method localizes the effect of a single quantitative trait locus (QTL) to a genomic segment located between pairs of qualitative markers rather than by associating the effect with an individual marker . Because a double cross-over is necessary to disassociate an interval from a QTL, interval mapping more efficiently exploits information obtained from genetic maps. Compared with the traditional method, the interval mapping method has several advantages. The probable position of the QTL can be inferred by the support interval, and the estimated locations and effects of QTLs tend to be asymptotically unbiased if there is only one QTL on a chromosome. The method also requires fewer individuals than the traditional approach.

Despite these advantages, the interval mapping method still has some problems. If there is more than one QTL on a chromosome the test statistic at the position being tested will be affected by all those QTL, and as a result the estimated position and effects of QTL will likely be biased (Haley and Knott 1992; Martinez and Curnow 1992; Zeng 1994). For this reason, it would appear that employing only

two markers at a time to do the test is not very efficient as the information from other markers is not utilized.

In an attempt to increase the reliability and accuracy of QTL mapping by taking into account the effects of possible linked QTLs, multiple regression analysis has been employed (Cowen 1989). Conditioning on linked markers in the multiple regression analysis reduces the chance of interference of possible linked QTLs on hypothesis testing and parameter estimation. This conditioning also reduces the sampling variance of the test statistic by controlling some residual genetic variation and thus increasing the power of the test (Weisberg 1985).

In this investigation we examined a population derived from an interspecific cross between slash pine H-28 (♂) × longleaf pine 3-356 (♀). RAPD markers and maps for each parent were employed to localize putative QTL influencing various early height growth (EHG) measurements. Three QTL mapping strategies single-marker or nonsimultaneous (NS) regression models, multiple-marker or simultaneous (S) regression models, and interval mapping were utilized and their results compared.

Materials and Methods

Mapping Population

Seventy-two progeny from an interspecific cross between slash pine H-28 (♂) × longleaf pine 3-356 (♀) were employed for genomic mapping and QTL searching. Seeds were germinated and grown in containers in a greenhouse at the Harrison Experimental Forest in Gulfport, Mississippi. At four months of age, the seedlings were transplanted to a nursery bed and grown for another five months. During this

period all seedlings were protected from diseases (primarily damping-off and brown spot needle blight) and given uniform growing conditions. At nine months, the seedlings were out-planted to a field site on the forest and allowed to grow under natural conditions (i.e. no irrigation or brown spot control). All the heterozygous RAPD loci identified in each parent [132 markers for longleaf pine, 101 for slash pine], and their corresponding maps (Kubisiak et al. in press) were used for mapping loci influencing various EHG mesurements.

Growth Variables

The growth variables measured included: hypocotyl length at two months (HYP2) and three months of age (HYP3); total height at three months (TH3), five months (TH5), nine months (TH9), and 21 months of age (TH21); height to green needles at 21 months of age (HTG21); and root collar diameter at 11 months (RCD11) and 21 months of age (RCD21). As brown spot disease was not controlled once the seedlings were out-planted to the field site, total height to green needles was used as an indicator of brown spot infection. The degree of brown spot infection (BS) was calculated as the difference between TH21 and HTG21. As TH21 was most likely confounded by the presence of brown spot, the following equation was used to transform TH21; ADJHT21 = TH21 + TH21/BS.

Nonsimultaneous and Simultaneous Models

The degree of association between individual RAPD markers and QTL influencing the early growth measurements (HYP2, HYP3, TH3, TH5, TH9, TH21, ADJHT21, BS, RCD11, and RCD21) were investigated by employing single-locus or

nonsimultaneous (NS) analysis of variance (ANOVA) models in which the individual marker-genotypes (+/- and -/-, where + = band present and - = band absent) were used as class variables (Keim et al. 1990). The proportion of the phenotypic variance explained by segregation of the marker was determined by the R^2 value. In an attempt to reduce the residual inflation common to single-locus models and to increase the probability of detecting secondary QTLs, multiple-loci or simultaneous (S) models were constructed. A separate model was constructed for each linkage group or pair, allowing for the removal of variation associated with any other QTL which might be present on the same linkage group. The best multi-variable regression model was determined by using the maximum R^2 improvement method available in the statistical software SAS® (SAS Institute Inc. 1989). A set of genome-wide simultaneous models was then constructed which included all the significant markers ($\alpha \leq 0.05$) located on each linkage group. Finally, simultaneous models taking into account any interaction between/among significant marker loci was employed to investigate the possibility of epistasis.

Interval Mapping

The program MAPMAKER/QTL version 1.1 (Lincoln et al. 1992) was used to obtain QTL maximum likelihood scans for the various EHG variables (HYP2, HYP3, TH3, TH5, TH9, TH21, ADJHT21, BS, RCD11, RCD21). Positions within intervals which had log of the odds (LOD) scores above 2.0 were used to define intervals containing putative QTL (Luo and Kearsey 1989).

Results

Phenotypic Means, Ranges and Variances

The ten variables selected for mapping were found to be normally distributed, therefore no further transformations of the data were required. The mean, range and phenotypic variance of each growth measurement are displayed in Table 8.1. Limited variation was observed in the F_1 progeny for both hypocotyl length at two months and three months, as well as for root collar diameter at 11 months and 21 months. However, rather substantial variation was observed for the total height growth and brown spot measurements.

Table 7.1. Mean and total phenotypic variation in the 72 F_1 progeny for each growth measurement (greenhouse measurements of hypocotyl length at two months and three months; nursery bed measurements of total height at three months, five months, and nine months, field measurements of total height at 21 months and adjusted height at 21 months; field measurement of brown spot resistance at 21 months; and field measurements of root collar diameter at 11 months and 21 months).

Growth Measurement	F_1 Progeny Mean (cm)	Range (cm)	F_1 Progeny Variance (cm^2)
Hypocotyl Length			
2 Months	1.52	0.7 - 2.8	0.19
3 Months	1.77	0.5 - 2.8	0.26
Total Height			
3 Months	7.36	1.5 - 10.7	2.37
5 Months	7.60	3.0 - 12.5	3.19
9 Months	24.36	8.0 - 45.0	53.10
21 Months	36.93	11.0 - 68.0	145.77
Adjusted Height			
21 Months	42.08	13.8 - 69.9	124.23
Brown-spot Resistance			
21 Months	10.80	1.0 - 32.0	11.15
Root Collar Diameter			
11 Months	1.32	0.5 - 1.8	0.08
21 Months	1.50	0.7 - 2.2	0.14

Mapping EHG QTL Using Nonsimultaneous and Simultaneous Models

The single-locus or nonsimultaneous (NS) models constructed for each of the 132 longleaf pine RAPD loci, suggested association between 82 markers and QTL influencing the various EHG measures at a significance level of $\alpha \leq 0.05$ (see Table 7.2). The NS models constructed for the 101 slash pine loci suggested association between 30 marker loci and QTL for the various EHG variables.

Table 7.2. Associations suggested between marker loci and QTL for 10 different early height growth variables using single-marker or nonsimultaneous ANOVA models ($\alpha \leq 0.05$).

LONGLEAF PINE

Hypocotyl length at two months (HYP2)

Marker	Linkage Group	Prob>F	R^2
F07_C	PpB	0.0162	0.081
213_A	PpB	0.0446	0.056
427_C	PpF	0.0188	0.076
A16_B	PpF	0.0489	0.063
299_A	PpF	0.0337	0.065
297_C	PpI	0.0150	0.083
479_B	PpI	0.0073	0.099
667_C*	PpI	0.0411	0.151
698_A	PpI	0.0305	0.067
G11_B	PpI	0.0176	0.081
G12_A	PpI	0.0462	0.056
102_C	PpI	0.0186	0.077
F14_A	PpN1	0.0250	0.072
270_A	PpR	0.0241	0.071
628_B	PpR	0.0330	0.066
527_A	PpT	0.0433	0.058
X04_A	unlinked	0.0068	0.103

Hypocotyl length at three months (HYP3)

Marker	Linkage Group	Prob>F	R^2
F07_C	PpB	0.0105	0.091
213_A	PpB	0.0446	0.056
427_C	PpF	0.0070	0.100
A16_B	PpF	0.0050	0.121
299_A	PpF	0.0388	0.061

(table con'd.)

Hypocotyl length at three months (HYP3) con't.

Marker	Linkage Group	Prob>F	R^2
348_C	PpH	0.0351	0.062
499_C	PpH	0.0122	0.086
297_C	PpI	0.0368	0.062
479_B	PpI	0.0193	0.076
G11_B	PpI	0.0470	0.055
102_C	PpI	0.0388	0.061
408_A	PpJ	0.0380	0.061
625_A	PpJ	0.0312	0.065
F14_A	PpN1	0.0441	0.058

Total height at three months (TH3)

Marker	Linkage Group	Prob>F	R^2
111_C	PpC	0.0200	0.077
111_B	PpC	0.0200	0.077
119_A	PpC	0.0466	0.055
543_A	PpC	0.0029	0.122
A09_A	PpC	0.0405	0.059
A16_A	PpC	0.0069	0.112
257_A	PpE	0.0191	0.077
499_C	PpH	0.0252	0.070

Total height at five months (TH5)

Marker	Linkage Group	Prob>F	R^2
119_A	PpC	0.0165	0.079
169_C	PpC	0.0281	0.067
543_A	PpC	0.0294	0.067
A09_A	PpC	0.0110	0.089
A16_A	PpC	0.0127	0.096
271_A	PpC	0.0446	0.057
181_A	PpH	0.0175	0.078
348_A	unlinked	0.0335	0.063

Total height at nine months (TH9)

Marker	Linkage Group	Prob>F	R^2
543_A	PpC	0.0182	0.078
A09_A	PpC	0.0494	0.054
A16_A	PpC	0.0247	0.079
A11_C	PpG	0.0207	0.075
X04_A	unlinked	0.0099	0.094

Total height at 21 months (TH21)

Marker	Linkage Group	Prob>F	R^2
119_A	PpC	0.0444	0.058
171_C	PpC	0.0120	0.089
A08_A	PpC	0.0223	0.075
A09_A	PpC	0.0191	0.078
123_B	PpN1	0.0348	0.064

(table con'd.)

Adjusted height at 21 months (ADJHT21)

Marker	Linkage Group	Prob>F	R^2
171_C	PpC	0.0462	0.057
543_A	PpC	0.0413	0.061
A08_A	PpC	0.0155	0.083
A11_C	PpG	0.0278	0.070
123_B	PpN1	0.0065	0.106

Brown spot resistance at 21 months (BS)

Marker	Linkage Group	Prob>F	R^2
258_A	PpA	0.0457	0.059
119_A	PpC	0.0120	0.089
171_C	PpC	0.0047	0.112
337_A	PpC	0.0327	0.101
A08_A	PpC	0.0276	0.069
A09_A	PpC	0.0001	0.116
257_A	PpE	0.0480	0.057
A16_B	PpF	0.0232	0.086
299_A	PpF	0.0305	0.069
297_C	PpI	0.0327	0.066
102_C	PpI	0.0437	0.059
J08_B	PpO	0.0491	0.059

Root collar diameter at 11 months (RCD11)

Marker	Linkage Group	Prob>F	R^2
337_A	PpC	0.0482	0.079
186_C	PpD2	0.0242	0.072
181_B	PpG	0.0019	0.130
J01_D	PpG	0.0043	0.115
348_A	unlinked	0.0415	0.058

Root collar diameter at 21 months (RCD21)

Marker	Linkage Group	Prob>F	R^2
337_A	PpC	0.0141	0.122
A08_A	PpC	0.0063	0.105
258_B	PpF	0.0168	0.082
153_C	PpJ	0.0475	0.057

SLASH PINE

Hypocotyl length at two months (HYP2)

Marker	Linkage Group	Prob>F	R^2
299_B	PcH	0.0390	0.060
590_A	PcH	0.0208	0.074
190_B	pair	0.0474	0.056
B13_B	pair	0.0217	0.075
B08_C	unlinked	0.0204	0.078

(table con'd.)

Hypocotyl length at three months (HYP3)

Marker	Linkage Group	Prob>F	R^2
677_B*	PeC	0.0244	0.180
B08_C	unlinked	0.0022	0.131

Total height at three months (TH3)

Marker	Linkage Group	Prob>F	R^2
698_A	PeC	0.0255	0.071
268_A	PeG	0.0121	0.088
E02_A	PeG	0.0487	0.058
G09_C	PeJ	0.0098	0.096
B08_F	unlinked	0.0233	0.076

Total height at five months (TH5)

Marker	Linkage Group	Prob>F	R^2
268_A	PeG	0.0020	0.130
E02_A	PeG	0.0042	0.117
631_A	PeE	0.0329	0.063

Total height at nine months (TH9)

Marker	Linkage Group	Prob>F	R^2
X04_A	unlinked	0.0144	0.084

Total height at 21 months (TH21)

Marker	Linkage Group	Prob>F	R^2
452_A	pair	0.0197	0.087
X04_A	unlinked	0.0030	0.124

Adjusted height at 21 months (ADJHT21)

Marker	Linkage Group	Prob>F	R^2
452_A	pair	0.0265	0.079
X04_A	unlinked	0.0006	0.164

Brown spot resistance at 21 months (BS)

Marker	Linkage Group	Prob>F	R^2
660_A	PeB	0.0087	0.097
608_A	PpG	0.0437	0.059
452_A	pair	0.0195	0.088
X04_A	unlinked	0.0021	0.132

Root collar diameter at 11 months (RCD11)

Marker	Linkage Group	Prob>F	R^2
590_A	PeH	0.0473	0.055
111_A	unlinked	0.0202	0.077

(table con'd.)

Root collar diameter at 21 months (RCD21)

Marker	Linkage Group	Prob>F	R^2
660_A	PeB	0.0392	0.061
299_B	PeH	0.0280	0.069
590_A	PeH	0.0091	0.096
452_A	pair	0.0102	0.105

*These markers contained high levels of missing data.

Based on the NS models nine marker loci located on the various longleaf pine linkage groups PpB, PpF, PpI, and PpN1 were found to be associated with putative QTL influencing hypocotyl length at both two and three months of age (see Table 7.2). Marker loci located on linkage group PpC were consistently (i.e. throughout time) found to be associated with QTL affecting total height growth as well as root collar diameter. The locus A09_A on PpC was found to be significantly associated with QTL influencing total height at three, five, nine, and 21 months of age. Similarly, a single locus (337_A on PpC) was found to be associated with QTL influencing root collar diameter at both 11 and 21 months of age.

The NS models suggested that the unlinked slash pine marker B08_C was associated with QTL affecting hypocotyl length at both two and three months of age. Markers located on the slash pine linkage group PeG (268_A and E02_A) were found to be associated with QTL affecting total height growth at both three and five months of age. The unlinked slash pine marker X04_A was found to be associated with QTL influencing total height growth at nine months, 21 months, adjusted height growth at 21 months of age, as well as with brown spot infection at 21 months. The locus 590_A (located on linkage group PeH) was found to be associated QTL affecting root collar diameter at 11 and 21 months of age.

The multiple-marker or simultaneous (S) models constructed for each of the 22 longleaf pine linkage groups suggested association between 56 marker loci and QTL influencing the various EHG measures at a significance level of $\alpha \leq 0.05$ (see Table 7.3). The S models constructed for the 19 slash pine linkage groups suggested association between 41 slash pine markers and QTL for various EHG measures.

Table 7.3. Associations suggested between marker loci and QTL for 10 different early height growth measures using linkage group-specific simultaneous regression models ($\alpha \leq 0.05$).

LONGLEAF PINE

Hypocotyl length at two months (HYP2)

Marker	Linkage Group	Prob>F	R^2 *
119_A	PpC	0.0243	0.352
171_C	PpC	0.0189	
A09_A	PpC	0.0081	
352_A	PpD1	0.0046	
333_B	PpD1	0.0039	
F14_A	PpN1	0.0164	
628_C	pair	0.0491	

Hypocotyl length at three months (HYP3)

Marker	Linkage Group	Prob>F	R^2
119_A	PpC	0.0199	0.214
171_C	PpC	0.0116	
A08_A	PpC	0.0157	
A09_A	PpC	0.0032	
352_A	PpD1	0.0390	
J01_D	PpG	0.0402	
J06_A	PpN1	0.0187	
F14_A	PpN1	0.0068	

Total height at three months (TH3)

Marker	Linkage Group	Prob>F	R^2
111_B	PpC	0.0246	0.443
A08_A	PpC	0.0349	
499_C	PpH	0.0406	
G11_A	PpI	0.0450	
J06_A	PpN1	0.0385	
632_A	PpN2	0.0061	

(table con'd.)

Total height at three months (TH3) con't.

Marker	Linkage Group	Prob>F	R^2
159_A	PpN2	0.0003	
337_A	pair	0.0007	

Total height at five months (TH5)

Marker	Linkage Group	Prob>F	R^2
116_B	PpE	0.0039	0.334
257_A	PpE	0.0096	
E08_A	PpI	0.0186	
Y20_B	PpI	0.0259	
408_A	PpJ	0.0491	
625_A	PpJ	0.0336	
J06_A	PpN1	0.0478	
632_A	PpN2	0.0086	

Total height at nine months (TH9)

Marker	Linkage Group	Prob>F	R^2
F07_C	PpB	0.0199	0.071
X18_A	PpR	0.0320	

Total height at 21 months (TH21)

Marker	Linkage Group	Prob>F	R^2
119_A	PpC	0.0316	0.093
479_A	PpF	0.0465	
A16_B	PpF	0.0057	
427_C	PpF	0.0247	

Adjusted height at 21 months (ADJHT21)

Marker	Linkage Group	Prob>F	R^2
F07_C	PpB	0.0279	0.140
625_A	PpJ	0.0325	
123_B	PpN1	0.0304	

Brown spot resistance at 21 months (BS)

Marker	Linkage Group	Prob>F	R^2
479_A	PpF	0.0155	0.284
A16_B	PpF	0.0008	
427_C	PpF	0.0052	
499_C	PpH	0.0440	
348_B	PpH	0.0189	
Y20_B	PpI	0.0049	
195_C	PpI	0.0112	

Root collar diameter at 11 months (RCD11)

Marker	Linkage Group	Prob>F	R^2
203_A	PpC	0.0082	0.071
W19_A	PpE	0.0416	

(table con'd.)

Root collar diameter at 21 months (RCD21)

Marker	Linkage Group	Prob>F	R^2
479_A	PpF	0.0298	0.099
A16_B	PpF	0.0034	
427_C	PpF	0.0377	
625_A	PpJ	0.0420	

SLASH PINE

Hypocotyl length at two months (HYP2)

Marker	Linkage Group	Prob>F	R^2
661_B	PeA	0.0084	0.127
116_A	PeA	0.0158	
G09_A	PeA	0.0208	

Hypocotyl length at three months (HYP3)

Marker	Linkage Group	Prob>F	R^2
116_A	PeA	0.0369	0.194
A12_A	PeA	0.0397	

Total height at three months (TH3)

Marker	Linkage Group	Prob>F	R^2
567_A	PeA	0.0471	0.265
327_A	PeB	0.0173	
493_A	PeB	0.0241	
322_B	PeE	0.0433	
B02_B	PeE	0.0157	
336_A	PeK	0.0337	

Total height at five months (TH5)

Marker	Linkage Group	Prob>F	R^2
493_A	PeB	0.0264	0.118
A16_A	PeB	0.0164	
X16_A	PeB	0.0043	
631_A	PeE	0.0142	

Total height at nine months (TH9)

Marker	Linkage Group	Prob>F	R^2
631_A	PeE	0.0087	0.104
460_A	pair	0.0448	

Total height at nine months (TH9)

Marker	Linkage Group	Prob>F	R^2
631_A	PeE	0.0087	0.245
460_A	pair	0.0448	
B08_E	PeA	0.0485	
493_A	PeB	0.0010	
A16_A	PeB	0.0131	
631_A	PeE	0.0109	
514_A	pair	0.0464	

(table con'd.)

Adjusted height at 21 months (ADJHT21)

Marker	Linkage Group	Prob>F	R^2
327_A	PeB	0.0151	0.246
493_A	PeB	0.0010	
631_A	PeE	0.0122	
299_A	PeE	0.0202	
452_A	pair	0.0006	

Brown spot resistance at 21 months (BS)

Marker	Linkage Group	Prob>F	R^2
B08_E	PeA	0.0138	0.543
660_A	PeB	0.0384	
562_A	PeB	0.0080	
661_A	PeB	0.0024	
327_A	PeB	0.0026	
493_A	PeB	0.0001	
A16_A	PeB	0.0004	
631_A	PeE	0.0227	
337_B	PeL	0.0416	

Root collar diameter at 11 months (RCD11)

Marker	Linkage Group	Prob>F	R^2
631_A	PeE	0.0324	0.084

Root collar diameter at 21 months (RCD21)

Marker	Linkage Group	Prob>F	R^2
660_A	PeB	0.0317	0.142
A16_A	PeB	0.0408	
631_A	PeE	0.0060	

*Amount of variation explained by the "best-fit" model containing all loci.

Based on the simultaneous (S) models, five marker loci located on three different longleaf pine linkage groups (PpC, PpD1, and PpN1) were found to be associated with putative QTL influencing hypocotyl length at both two and three months of age (see Table 7.3). The only longleaf pine marker locus found to be associated with QTL influencing hypocotyl length in both the NS and S models was locus F14_A located on linkage group PpN1. Three marker loci located on linkage groups PpN1 (J06_A) and PpN2 (159_A and 632_A) were found to be associated with total height growth at both three and five months of age. The only longleaf

pine locus found to be associated with QTL influencing total height in both NS and S models was 111_B on linkage group PpC. This locus was associated with QTL influencing total height growth at three months of age. Using NS models, markers on linkage group PpC were found to be consistently (i.e. throughout time) associated with QTL influencing total height growth and root collar diameter. However, with the S models loci on linkage group PpC were only found to be associated with QTL affecting total height growth at three and 21 months, as well as root collar diameter at 11 months.

Using NS models, the only slash pine locus found to be associated with QTL affecting hypocotyl length at both two and three months of age was the unlinked locus B08_C. However, using S models this locus was not significant, instead locus 116_A on linkage group PeA was found to be associated with QTL affecting hypocotyl length at both two and three months of age (see Figure 7.3). Using the S models, linkage groups PeB and PeE were consistently found to harbor loci associated with QTL influencing total height growth and root collar diameter.

A second set of S models consisting of those loci found to be significant ($\alpha \leq 0.05$) in the linkage group-specific S models, as well as any significant ungrouped or unlinked loci, were constructed. When all the significant marker loci located throughout the genome were used in a simultaneous model, a number of loci were no longer found to be significantly associated with QTL. Only 15 of the 56 longleaf pine loci identified in the linkage group-specific S models were suggested to be associated with QTL (see Table 7.4). Likewise, only 14 of the 41 slash pine loci

were associated with QTL. No time-associated linkages between the same locus and a particular EHG measure were observed. In other words, a locus found to be significantly associated with a particular trait such as hypocotyl length at two months of age was observed to be significant at three months.

Table 7.4. Associations suggested between marker loci and QTL for different early height growth measures using genome-wide simultaneous regression models ($\alpha \leq 0.05$).

LONGLEAF PINE

Hypocotyl length at two months (HYP2)

Marker	Linkage Group	Prob>F	R^2
119_A	PpC	0.0095	0.265
A09_A	PpC	0.0116	
352_A	PpD1	0.0057	
333_B	PpD1	0.0060	

Hypocotyl length at three months (HYP3)

Marker	Linkage Group	Prob>F	R^2
119_A	PpC	0.0447	0.093
F14_A	PpN1	0.0118	

Total height at three months (TH3)

Marker	Linkage Group	Prob>F	R^2
A08_A	PpC	0.0484	0.233
632_A	PpN2	0.0277	
337_A	pair	0.0195	

Total height at five months (TH5)

Marker	Linkage Group	Prob>F	R^2
116_B	PpE	0.0056	0.158
257_A	PpE	0.0342	

Adjusted height at 21 months (ADJHT21)

Marker	Linkage Group	Prob>F	R^2
123_B	PpN1	0.0072	0.040

Brown spot resistance at 21 months (BS)

Marker	Linkage Group	Prob>F	R^2
A16_B	PpF	0.0102	0.164
499_C	PpH	0.0150	
348_B	PpH	0.0056	

(table con'd.)

SLASH PINE

Hypocotyl length at two months (HYP2)

Marker	Linkage Group	Prob>F	R^2
661_B	PeA	0.0276	0.035

Hypocotyl length at three months (HYP3)

Marker	Linkage Group	Prob>F	R^2
B08_C	unlinked	0.0021	0.131

Total height at three months (TH3)

Marker	Linkage Group	Prob>F	R^2
322_B	PeE	0.0115	0.196
B02_B	PeE	0.0467	
B08_F	unlinked	0.0023	

Adjusted height at 21 months (ADJHT21)

Marker	Linkage Group	Prob>F	R^2
X04_A	unlinked	0.0157	0.164

Brown spot resistance at 21 months (BS)

Marker	Linkage Group	Prob>F	R^2
660_A	PeB	0.0093	0.458
562_A	PeB	0.0009	
661_A	PeB	0.0045	
327_A	PeB	0.0130	
493_A	PeB	0.0027	
A16_A	PeB	0.0041	

Root collar diameter at 11 months (RCD11)

Marker	Linkage Group	Prob>F	R^2
111_A	unlinked	0.0229	0.077

Root collar diameter at 21 months (RCD21)

Marker	Linkage Group	Prob>F	R^2
660_A	PeB	0.0071	0.061

*Amount of variation explained by the "best-fit" model containing only significant loci.

A final set of S models was constructed which included all the significant marker loci identified with the genome-wide S models and their possible interactions among one another. These models were constructed in order to obtain information about any possible epistatic effects between QTL. Based on these S models, it appears as if there are epistatic interactions occuring between QTL influencing

hypocotyl length at 3 months, total height at three months, brown spot resistance in longleaf pine, as well as for brown spot resistance in slash pine (see Table 7.5).

Table 7.5. Epistatic interactions between QTL influencing various early height growth measures ($\alpha \leq 0.05$).

LONGLEAF PINE			
Hypocotyl length at three months (HYP3)			
Marker	Linkage Group	Prob>F	R^{2*}
119_A	PpC	0.0825	0.178
F14_A	PpN1	0.0413	
119_A*F14_A		0.0109	
Total height at three months (TH3)			
Marker	Linkage Group	Prob>F	R^2
A08_A	PpC	0.8543	0.348
632_A	PpN2	0.1620	
337_A	pair	0.0007	
A08_A*337_A		0.0092	
Brown spot resistance at 21 months (BS)			
Marker	Linkage Group	Prob>F	R^2
A16_B	PpF	0.0152	0.293
499_C	PpH	0.7784	
348_B	PpH	0.0209	
A16_B*348_B		0.0055	
SLASH PINE			
Brown spot resistance at 21 months (BS)			
Marker	Linkage Group	Prob>F	R^2
660_A	PeB	0.0001	0.486
562_A	PeB	0.0092	
661_A	PeB	0.0341	
327_A	PeB	0.1481	
493_A	PeB	0.0045	
A16_A	PeB	0.0004	
562_A*A16_A		0.0036	
562_A*327_A		0.0345	

*Amount of variation explained by the "best-fit" model containing all significant loci and interactions.

Mapping EHG QTL Using Interval Mapping

QTL influencing five of the eight early height growth measurements were localized using the interval mapping method (Figure 7.1). One interval on longleaf pine 3-356 linkage group PpI (G11_B - 667_C) was found to explain 17.3% (LOD = 2.35) of the total phenotypic variance for hypocotyl length at two months. This QTL had a negative effect on hypocotyl length with an additive estimated effect of -0.357 cm. At a LOD threshold of 2.0, no QTL influencing hypocotyl length at two months was detected in the slash pine genome, nor were any QTL influencing hypocotyl length at three months detected in either longleaf pine 3-356 or slash pine H-28.

Two QTL influencing total height growth were detected. One in longleaf pine influencing total height growth at three months (LOD = 2.21), and the other in slash pine influencing total height growth at 5 months (LOD = 2.02). The longleaf pine-associated QTL was located on linkage group PpC between the marker loci 111_C - 543_A. This QTL explained 16.5% of the phenotypic variation and had an estimated additive effect of -1.238 cm. The slash pine-associated QTL was located on linkage group PeG between the marker loci E02_A - 268_A, and explained 12.7% of the phenotypic variation with an estimated additive effect of -1.258 cm.

One QTL influencing brown spot resistance was detected (LOD = 2.68) on longleaf pine linkage group PpC between the markers A09_A and 171_C. This locus was found to explain 12.9% of the phenotypic variation for brown spot resistance at 21 months, and had an estimated additive effect of +5.6 cm (measured in terms of the length of stem containing live needles).

One QTL influencing root collar diameter was detected (LOD = 2.17). This QTL was detected in longleaf pine and was located on linkage group PpG between the marker loci J01_D - A11_C. It was found to explain 13.4% of the phenotypic variation for root collar diameter at 11 months, and had an estimated additive effect of +0.206 cm.

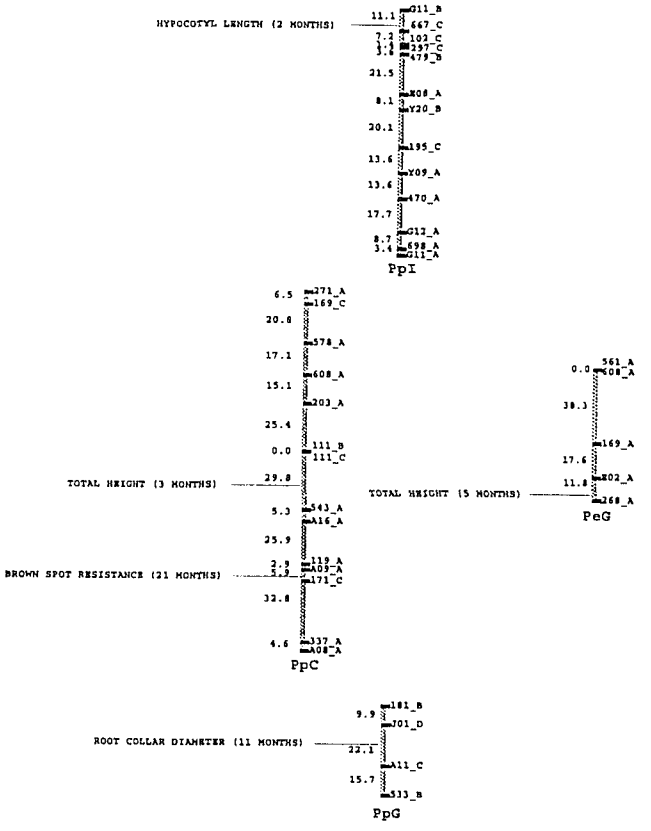

Figure 7.1. Intervals from the longleaf pine and slash pine RAPD maps depicting the approximate location of quantitative trait loci for various early height growth measurements (for details see Materials and Methods).

Discussion

Comparison of Methodologies

The present study provided the opportunity to compare different methodologies for QTL mapping; traditional single-marker analysis of variance (Soller et al. 1976), interval mapping (Lander and Botstein 1989), and multiple regression (Cowen 1989). The traditional single-marker analysis of variance and interval mapping appear to be directly comparable. In every case where a marker was declared significant with the single-marker or nonsimultaneous (NS) models ($\alpha \leq 0.05$), the log of the odds (LOD) scan surpassed 1.0. Although the two methods appeared to yield identical results in terms of identifying loci significantly linked to QTL, the interval mapping method offers some advantages. Interval mapping uses information on flanking markers, which may provide more power than the single-marker method for detecting QTLs (Lander and Botstein 1989). In addition, it also provides information on the likely location and effect of the QTL.

The multiple regression technique did not appear to be comparable to the results obtained from either the single-marker analysis or interval mapping. Only a single comparison could be made and that was for a QTL influencing total height at three months of age in longleaf pine. The interval mapping technique localized a QTL to the interval between markers 111_C and 543_A in linkage group PpC, whereas the multiple regression technique suggested linkage of a QTL to marker A08_A at the end of the same linkage group. No other QTL identified by interval mapping was identified by the multiple regression technique.

These particular findings might lead one to question whether the multiple regression method is a valid technique for mapping QTL. In an attempt to answer this question, a number of simulated data sets were contructed and analyzed using multiple regression. The data sets were contructed with the genetic simulation software GREGOR (Nick Tinker, McGill University). Each simulated genome consisted of two chromosomes with 21 loci per chromosome (5 cM between loci). QTL (for a single trait) with varying additive effects were placed along the chromosomes. The corresponding data sets were analyzed as described for the simultaneous models in the Materials and Methods section. Based on these simulations, the multiple regression method proved to be highly efficient at identifying markers linked to QTL. Occassionally the multiple regression method failed to declare marker loci as being significantly linked to QTL, however, this only occurred when other QTL with large additive effects were located on the same chromosome.

Although some very different conclusions (in terms of the number and location of QTL) were drawn from the multiple regression method when compared to either the single-marker or interval mapping methods, it must be kept in mind that neither of the latter two methods accounts for multiple linked or unlinked QTL. As a result the test statistic at the position being tested will be affected by these QTL and are likely biased (Zeng 1994). It would, however, at this time be ill-advised to totally discount associations identified by single-marker or interval mapping methods. Not until the theoretical properties of the multiple regression method are analyzed in

detail, and direct results comparing the actual gain obtained by employing each of these methods become available, should any one of these methods be disregarded.

CHAPTER 8

CONCLUSIONS AND FUTURE RECOMMENDATIONS

As a result of work conducted in this dissertation, several factors have come to light which will need to be addressed in future research. First, the linkage maps constructed are not complete, and efforts to converge the maps towards 12 chromosomes and expand the coverage towards 100% are needed. Such efforts would greatly increase the efficiency of future QTL searching and marker-aided selection efforts in the southern pine. Secondly, development of multiallelic (i.e. codominant) markers for conifers is essential. Such markers would provide a means with which to integrate the parent-specific linkage maps produced with dominant RAPD markers. Thirdly, the conversion of RAPD markers to sequence characterized amplified region (SCAR) markers would increase the reliability of PCR-based markers by simplifying the kinetics of the RAPD reaction. The specificity of SCAR primers might also increase the transferability of markers across different genotypes, possibly even leading to the construction of a unified map for the southern pine. Fourthly, problems associated with highly polygenic inheritance, genotype by environment interaction, and linkage equilibrium will need to be considered for each species and trait which the tree breeder may desire to improve using molecular markers. Finally, careful consideration of how the markers will be implemented will be critical to the overall success of a marker-aided selection program.

Searching for genes influencing agronomic and economic traits of interest requires a complete linkage map (not necessarily a highly-saturated map). While

karyological studies have revealed that pine genomes consist of 12 similar-sized pairs of homologous chromosomes (Saylor 1972; Kormutak 1975), all the genetic maps constructed from work carried out in this dissertation contained greater than 12 linkage groups. For example, the most complete map consisted of 162 markers in 14 groups and 2 pairs. The fact that more than 12 linkage groups were suggested and that some markers still remain unlinked, implies that the maps are incomplete. Based on estimates of genome size and levels of coverage, screening more primers will be inefficient as most additional polymorphisms will likely map to regions already adequately covered by markers. However, more systematic means for converging the maps towards 12 linkage groups and expanding the coverage towards 100% might be taken.

One means of increasing map coverage would be to simply increase the sample size (i.e. number of progeny). An increased sample size could potentially increase the LOD ratio for some associations (e.g. see Figure 5.1 and 5.2). A more efficient approach would be to use a form of bulked segregant analysis (Michelmore et al. 1991) where bulks are constructed for terminal and unlinked marker classes (Reiter et al. 1992). It would be most advantageous to concentrate on the unlinked markers, the linked pairs, and the terminal loci of the smallest linkage groups first, as the larger linkage groups are more likely to be nearly complete. Additional primers could then be screened against these bulks and new polymorphisms identified. These polymorphisms would then need to be scored on the entire mapping population and added to the existing map, hopefully allowing for further convergence.

The RAPD marker linkage maps constructed in these studies were parent-specific. The dominant nature of RAPDs precludes the use of any loci at which one parent is homozygous band-present, as all the progeny will contain the band (i.e. it does not segregate). Therefore, the basic strategy when scoring RAPDs in diploid populations is to select those loci which are band-present in one parent, band-absent in the other, and segregating in the progeny (termed the pseudo-testcross strategy; Grattapaglia and Sederoff 1994). Using this strategy two sets of loci are obtained, one set heterozygous in one parent and a second set heterozygous in the other. As none of the heterozygous loci identified in one parent are common to the other parent, it is impossible to determine homologies to integrate the two maps. However, those loci which are heterozygous in both parents (i.e. segregating 3:1 in the progeny) can be used in linkage analyses to assign homologies (e.g. see Figure 4.1). In crosses made between distantly-related individuals (e.g. interspecific crosses) extremely low numbers of loci appear to be simultaneously heterozygous in both parents. For example, Grattapaglia and Sederoff (1994) found only 11 of 558 loci (2.0%) to be heterozygous in both parents of an interspecific eucalyptus cross. Similarly, only 14 of 247 loci (5.7%) were identified to be heterozygous in both parents in our slash pine × longleaf pine cross. The number of loci found to be heterozygous in both parents might be expected to increase as more closely-related individuals are crossed. As an example, Roy et al. (1992) found 5 of 14 loci (35.7%) identified in three different intraspecific crosses of yellow birch (*Betula alleghaniensis*) to be heterozygous in both parents. Integration of RAPD maps

constructed from crosses between distantly-related individuals will most likely require other sources of data to serve as bridges (Ritter et al. 1990). Codominant markers such as RFLPs or minisatellites would be the most logical choice for such an endeavor as these markers are often highly polymorphic [i.e. multiallelic] (Devey et al. 1991; Nybom and Rogstad 1990; Condit and Hubbell 1991; Broun et al. 1992; Thomas and Scott 1993).

Regardless of its conceptual simplicity, the kinetics of the RAPD reaction are quite complex. Annealing temperature, degree of sequence similarity at priming sites, and primer competition all can affect the amplification of RAPD markers. As a result, amplification inconsistencies can produce spurious data in the form of misclassified individuals. Misclassification of individuals directly results in inflated map distances (see Discussion in CHAPTER 6). Therefore, further technical refinements will be required to reduce the error associated with genetic distance estimates. One possible means of reducing this error would be to convert RAPD loci to sequence characterized amplified regions [SCARs] (Paran and Michelmore 1993).

The theory behind the SCAR technique is to increase the length of the primers employed so that only a single genetically-defined locus is amplified. SCAR primers can be constructed by cloning RAPD loci of interest directly from low melting point agarose gels. Once the fragments (loci) have been cloned, they are then sequenced 5' to 3' on both sense and antisense strands. Actual primer lengths are adjusted in order to best match primer pairs based on thermal melting points (Tm), delta G's, and complementarity. Equal Tm's, positive delta G's (indicating a

non-preference for hair-pin loop formation), and no complementarity are the main goals. In general, 25 base pair primers have been found to be efficient for single band amplification (Williams et al. 1991; Paran and Michelmore 1993). As only a single locus is amplified, the kinetics of the reaction are greatly simplified. This increased specificity would eliminate the competition for primers common to the multilocus RAPD technique hence reducing the possibility of misclassification and the associated inflation of map distances.

One major reservation regarding the use of RAPD markers for genetic mapping and QTL searching is that preliminary comparisons among species-specific RAPD maps suggest that bands of similar molecular weight amplified by the same primer often map to discrete, unlinked locations in different genotypes (Johns 1992; Nance et al. 1992a; Kesseli et al. 1994). This is not surprising as RAPD amplification products can often contain (or are contained within) dispersed repetitive DNA sequences (Williams et al. 1991; Kazan et al. 1992; Paran and Michelmore 1993; Kesseli et al. 1994). As a result, a RAPD marker found to be associated with a QTL in one genotype may not be associated with the QTL in another. One possible way to circumvent this potential problem might, again, be to convert RAPDs to SCARs. The increased specificity of the SCAR primers would decrease the chance of amplifying unlinked loci in other genotypes.

The conversion of RAPDs to SCARs may allow for the construction of a unified map for the southern pine. SCARs, however, would not circumvent the need to localize favorable QTL alleles within each individual in a breeding population, as

it has long been recognized that one of the major problems facing marker-aided selection in allogamous species is the linkage equilibrium encountered between marker loci and QTL of interest (Soller et al. 1976; Strauss et al. 1992). Since SCARs are PCR-based they are amenable to automation, and should thus allow for the rapid collection of data needed to make such an immense task feasible. In addition, multiple SCAR loci could be simultaneously amplified in the same reaction-well greatly reducing the number of assays required.

In addition to the aforementioned concerns, much discussion has been voiced regarding the potential of molecular markers to aid in the improvement of agronomically and economically important traits in operational tree breeding programs (Neale and Williams 1991; Bernatzky and Mulcahy 1992; Bradshaw and Foster 1992; Nance et al. 1992b; Tauer et al. 1992; Strauss et al. 1992; Williams and Neale 1992). At least three factors may diminish the potential for marker-aided selection in tree breeding programs as they are currently practiced: (1) if the assumption that the traits of interest are controlled by a large number of genes each with small additive effects were true, it would be nearly impossible to detect associations between markers and QTL unless extremely large progeny sizes are employed; (2) large genotype by environment interactions would necessitate developing a different set of markers for each environment; (3) linkage equilibrium will make it impossible to predict whether a marker allele is in coupling or repulsion phase with the desired QTL allele.

Unfortunately, the assumption of highly polygenic inheritance can only be adequately tested by conducting marker-QTL analyses. Numerous studies conducted to date, indicate that most quantitative traits of agronomic or economic importance are controlled by genes having a variable effect on the phenotype. In other words, the detection of genes with both large and small effects on the phenotype has generally been the rule (Paterson et al. 1988; Keim et al. 1990; Stuber 1992; Goldman et al. 1993; Komatsuda et al. 1993; Mansur et al. 1993; Wang et al. 1994; Leonards-Schippers et al. 1994; and many others).

Marker genotype by environment interactions should parallel quantitative trait by environment interactions. Therefore, if quantitative trait by environment interaction has been observed for a particular trait in a particular species [such as has been observed for fusiform rust resistance in loblolly pine (Powers and Mathews 1980), and wood quality in Japanese larch (Loo et al. 1982)], then the added cost associated with mapping these QTL in different environments would greatly impede the adoption and success of applying molecular markers for their improvement.

The most frequently debated problem facing marker-aided selection in tree breeding programs is linkage equilibrium [see Strauss et al. (1992) for an in-depth discussion]. One solution to the equilibrium dilemma would be to map all the individuals in an elite population. This might be an economically reasonable alternative if the population consisted of a limited number of individuals, but would not be a reasonable alternative with larger population sizes. Even if it is not possible to construct a map for every individual, it is possible that a correlated response to

selection on markers might achieve a greater response than direct selection on the trait itself. Another solution to the equilibrium problem would be to saturate the region around major effect QTL. Theoretically, tighter linkages experience higher disequilibrium due to the reduced chance of recombination between the loci. Therefore, selection on tightly linked, flanking markers around major effect QTL should greatly reduce the problem of linkage equilibrium in operational tree breeding programs. In the immediate future, however, it appears as if the greatest benefits will be obtained by employing molecular markers within specifically designed pedigrees constructed to address unique problems (e.g. Bernatzky and Mulcahy 1992).

The thrust for work conducted in this dissertation was to develop molecular markers for use in a backcross breeding program to speed the introgression of genes influencing rapid early height growth (EHG) from slash pine (*Pinus elliottii* Engelm. var. *elliottii*) into longleaf pine (*Pinus palustris* Mill.). Along these lines, the efficacy of RAPD markers for genetic mapping in the *Pinaceae* was determined using both half-sib and full-sib families. In terms of achieving the overall goal (a longleaf pine capable of rapid early height growth), the next step will be to produce BC_1 families and map QTL influencing EHG. We currently are anticipating the formation of strobili on several selected F_1 progeny so that we can proceed with the research. Our plans are to produce and test two divergent backcross families. In one family, a longleaf pine will be used as the recurrent parent, and in the other family, a slash pine will be used as the recurrent parent. Use of the same F_1

individual as the male parent in each of the backcross families should allow for the identification of EHG loci in the most comparable genetic background (with positive effect EHG alleles coming from the hybrids' slash pine parent and negative effect (grass stage) EHG alleles coming from the hybrids' longleaf pine parent). By analyzing both backcross families, we expect to find "real" (not false positive) EHG loci as they should map to the same marker loci in both families.

Successful completion of the next stage of research should result in the development of longleaf pine hybrid genotypes which exhibit vigorous early height growth as a result of their harboring high numbers of positive-effect EHG alleles. These genotypes will be selected for further backcrossing to increase the proportion of the recurrent longleaf pine genome, clonally propagated and used directly in production plantings, and crossed to fix loci influencing EHG. Selected progeny, fixed at EHG loci, could then be used as parents in seed orchards as well as in further backcrosses.

REFERENCES

Adams, W.T. 1992. Gene dispersal within forest tree populations. *In* Population genetics of forest trees. *Edited by* Adams, W.T., Strauss, S.H., Copes, D.L., and A.R. Griffin. Kluwer Academic Publishers. p 217-240.

Adams, W.T., and R.J. Joly. 1980. Linkage relationships among twelve allozyme loci in loblolly pine. J. Heredity 71:199-202.

Adams, W.T., Neale, D.B., Doerksen, A.H., and D.B. Smith. 1990. Inheritance and linkage of isozyme variants from seed and vegetative bud tissues in coastal Douglas-fir (*Pseudotsuga menziesii* var. *menziesii* (Mirb) Franco). Silvae Genet. 39:3-4.

Anderson, N.G., Tollaksen, S.L., Pascoe, F.H., and L. Anderson. 1985. Two-dimensional electrophoresis analysis of wheat seed proteins. Crop Sci. 25:667-674.

Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A., and K. Struhl. 1987. Current protocols in molecular biology. John Wiley & Sons, New York.

Bahrman, N. and C. Damerval. 1989. Linkage relationships of loci controlling protein amounts in maritime pine (*Pinus pinaster* Ait.) Heredity 63:267-274.

Bailey, N.T.J. 1961. Introduction to the mathematical theory of genetic linkage. Clarendon Press, Oxford.

Beckmann, J.S., and M. Soller. 1983. Restriction fragment length polymorphisms in genetic improvement: methodologies, mapping, and costs. Theor. Appl. Genet. 67:35-43.

Beckmann, J.S., and M. Soller. 1986. Restriction fragment length polymorphisms in plant genetic improvement. Oxford Surv. Plant Mol. Biol. 3:197-246.

Bernatzky, R., and D.L. Mulcahy. 1992. Marker-aided selection in a backcross breeding program for resistance to chestnut blight in the American chestnut. Can J. For. Res. 22:1031-1035.

Bernatzky, R. and S. Tanksley. 1986. Toward a saturated linkage map in tomato based on isozymes and random cDNA sequences. Genetics 112:887-898.

Bradshaw, H.D., and G.S. Foster. 1992. Marker-aided selection and propagation systems in trees: advantages of cloning for studying quantitative inheritance. Can. J. For. Res. 22:1044-1049.

Broun, P., Ganal, M.W., and S.D. Tanksley. 1992. Telomeric assays display high levels of heritable polymorphism among closely related plant varieties. Proc. Natl. Acad. Sci. USA 89:1354-1357.

Brown, C. L. 1964. The seedling habit of longleaf pine. A Monograph. Georgia Forest Research Council and School of Forestry, University of Georgia. 68 p.

Bush, R.M., and P.E. Smouse. 1992. Evidence for the adaptive significance of allozymes in tree species. In Population genetics of forest trees. *Edited by* Adams, W.T., Strauss, S.H., Copes, D.L., and A.R. Griffin. Kluwer Academic Publishers. p 179-196.

Byram T. D., and W. J. Lowe. 1985. Longleaf pine tree improvement in the western gulf region. Proc. 18th South. For. Tree Imp. Conf. p.78-87.

Caetano-Anolles, G., Bassam, B.J., and P.M. Gresshoff. 1991. DNA amplification fingerprinting using very short arbitrary oligonucleotide primers. Bio/Tech. 9:553-557.

Caetano-Anolles, G., Bassam, B.J., and P.M. Gresshoff. 1994. Multiple arbitrary amplicon profiling using short oligonucleotide primers. *In* Current Topics in Plant Molecular Biology: Plant Genome Analysis. CRC Press. p 29-45.

Carlson, J.E., Tulsieram, L.K., Glaubitz, J.C., Luk, V.W.K., Kauffeldt, C., and R. Rutledge. 1991. Segregation of random amplified DNA markers in F_1 progeny of conifers. Theor. Appl. Genet. 83:194-200.

Castle. 1921. An improved method of estimating the number of genetic factors concerned in cases of blending inheritance. Science 54:223.

Chaparro, J.X., Werner, D.J., O'Malley, D.O., and R.R. Sederoff. 1994. Targeted mapping and linkage analysis of morphological, isozyme, and RAPD markers in peach. Theor. Appl. Genet. 87:805-815.

Cheliak, W.M., and J.A. Pitel. 1985. Inheritance and linkage of allozymes in *Larix laricina*. Silvae Genet. 34:142-148.

Cockerham, C.C. 1986. Modifications in estimating the number of genes for a quantitative character. Genetics 114:659-654.

Colby, S., Groover, A., Kinlaw, C., Harry, D., and D. Neale. 1993. Advancing toward a transcriptional map of the loblolly pine genome. *In* Proceedings of the 22nd South. For. Tree Impr. Conf. June 14-17. Atlanta, GA. p 232-240.

Condit, R., and S.P. Hubbell. 1991. Abundance and DNA sequence of two-base repeat regions in tropical trees. Genome 34:66-71.

Conkle, M.T. 1971. Inheritance of alcohol dehydrogenase and leucine aminopeptidase isozymes in knobcone pine. For. Sci. 17:190-194.

Conkle, M.T. 1981. Isozyme variation and linkage in six conifer species. In: Conkle MT (technical coordinator) Proceedings of the Symposium on Isozymes of North American Forest Trees and Insects, Berkeley, California, July 27, 1979. General Technical Report PSW-48, Berkeley, California: USDA Forest Service, Pacific Southwest Forest Experiment Station; 11-17.

Conkle, M.T. 1992. Genetic diversity - seeing the forest through the trees. *In* Population genetics of forest trees. *Edited by* Adams, W.T., Strauss, S.H., Copes, D.L., and A.R. Griffin. Kluwer Academic Publishers. p 5-22.

Conkle, M.T., Hodgskisss, P.D., Nunnally, L.B., and S.C. Hunter. 1982. Starch gel electrophoresis of conifer seeds: a laboratory manual. USDA For. Ser. Gen. Tech. Report PSW-64. 18 pp.

Cowen, N.M. 1989. Multiple linear regression analysis of RFLP data sets used in mapping QTLs. *In* Development and Applications of Molecular Markers to Problems in Plant Genetics. *Edited by* T. Helentjaris and B. Burr. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y. p113-116.

Darvasi, A., and J.I. Weller. 1992. On the use of the moments method of estimation to obtain approximate maximum likelihood estimates of linkage between a genetic marker and a quantitative trait locus. Heredity 68:43-46.

Derr, H.J. 1966. Longleaf x slash hybrids at age 7: Survival, growth, and disease susceptibility. J. For. 64:236-239.

Derr, H.J. 1969. Intraspecific and interspecific crosses for testing brown-spot resistance of longleaf pine. Final Report of Study FS-SO-1102-5.2. New Orleans, Louisiana: USDA Forest Service, Southern Forest Experiment Station.

Devey, M. E., Jermstad, K. D., Tauer, C. G., and D. B. Neale. 1991. Inheritance of RFLP loci in a loblolly pine three-generation pedigree. Theor. Appl. Genet. 83:238-242.

Dhillon, S.S., Miksche, J.P., and R.A. Cecich. 1984. DNA changes in senescing leaves of *Populus deltoides*. Plant Physiol. (Suppl.) 75:120.

Diers, B.W., Keim, P., Fehr, W.R., and R.C. Shoemaker. 1992. RFLP analysis of soybean seed protein and oil content. Theor. Appl. Genet. 83:608-612.

Donner, B.L., and F. Hines. 1987. Forest statistics for Mississippi counties - 1987. Resource Bull. SO-129. New Orleans, LA: USDA, Forest Service, Southern Forest Experiment Station. pp. 79.

Durham, R.E., Liou, P.C., Gmitter Jr., E.G., and G.A. Moore. 1992. Linkage of restriction fragment length polymorphisms and isozymes in *Citrus*. Theor. Appl. Genet. 84:39-48.

El-Kassaby, Y.A., and O. Sziklai. 1982. Genetic variation of allozyme and quantitative traits in selected Douglas-fir (*Pseudotsuga menziesii* var. *menziesii* (Merb.) Franco) population. For. Ecol. Manag. 4:115-126.

Ellstrand, N.C. 1992. Gene flow among seed plant populations. *In* Population genetics of forest trees. *Edited by* Adams, W.T., Strauss, S.H., Copes, D.L., and A.R. Griffin. Kluwer Academic Publishers. p 241-256.

Elston, R.C., and J. Stewart. 1971. A general model for the genetic analysis of pedigree data. Hum. Hered. 21:523-542.

Epperson, B.K. 1992. Spacial structure of genetic variation within populations of forest trees. *In* Population genetics of forest trees. *Edited by* Adams, W.T., Strauss, S.H., Copes, D.L., and A.R. Griffin. Kluwer Academic Publishers. p 257-278.

Faure, S., Noyer, J.L., Horry, J.P., Lanaud, C., and D. Gonzalez de Leon. 1993. A molecular marker-based linkage map of diploid bananas (*Mus acuminata*). Theor. Appl. Genet. 87:517-526.

Fowler, D.P., and R.W. Morris. 1977. Genetic diversity in red pine: evidence for low genic heterozygosity. Can. J. For. Res. 7:343-347.

Furnier, G.R., Knowles, P., Aleksiuk, M.A., and B.P. Dancik. 1986. Inheritance and linkage of allozymes in seed tissue of whiteback pine. Can. J. Genet. Cytol. 28:601-604.

Furnier, G.R., Stine, M., Mohn, C.A., and M.A. Clyde. 1991. Geographic patterns of variation in allozymes and height growth in white spruce. Can. J. For. Res. 21:707-712.

Gebhardt, C., Ritter, E., Debener, T., Schachtschabel, U., Walkemeiern, B., Uhrig, H., and F. Salamini. 1989. RFLP-analysis and linkage mapping in *Solanum tuberosum*. Theor. Appl. Genet. 78:65-75.

Gerber, S., Rodolphe, F., Bahrman, N., and Ph. Baradat. 1993. Seed-protein variation in maritime pine (*Pinus pinaster* Ait.) revealed by two-dimensional electrophoresis: genetic determinism and construction of a linkage map. Theor. Appl. Genet. 85:521-528.

Goldman, I.L., Rocheford, T.R., and J.W. Dudley. 1993. Quantitative trait loci influencing protein and starch concentration in the Illinois long term selection maize strains. Theor. Appl. Genet. 87:217-224.

Grattapaglia, D., Chaparro, J., Wilcox, P., McCord, S., Werner, D., Amerson, H., McKeand, S., Bridgewater, F., Whetten, R., O'Malley, D., and R. Sederoff. 1992. Mapping in woody plants with RAPD markers: application to breeding in forestry and horticulture. Proc. of the Symp. Appl. of RAPD Tech. to Plant Breed. November 1, 1992. Minneapolis, Minnesota p 37-40.

Grattapaglia, D. and R. Sederoff. 1994. Genetic linkage maps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudo-testcross: mapping strategy and RAPD markers. Genetics 137:1121-1137.

Groover, A., Devey, M., Fiddler, T., Lee, J., Megraw, R., Mitchell-Olds, T., Sherman, B., Vujcic, S., Williams, C., and D. Neale. 1993. Genetic mapping of quantitative trait loci influencing wood specific gravity in loblolly pine (*Pinus taeda*). *In* Proceedings of the 22nd South. For. Tree Impr. Conf. June 14-17. Atlanta, GA. p 257-269.

Guerra, M.S. 1984. Cytogenetics of *Rutaceae*. II. Nuclear DNA content. Caryologia 37:219-226.

Guries, R. P., Friedman, S. T., and F. T. Ledig. 1978. A megagametophyte analysis of genetic linkage in pitch pine (*Pinus rigida* Mill.). Heredity 40:309-31.

Guries, R.P., and F.T. Ledig. 1978. Inheritance of some polymorphic isoenzymes in pitch pine (*Pinus rigida*). Heredity 40:27-32.

Haldane, J.B.S. 1919. The combination of linkage values, and the calculation of distances between the loci of linked factors. J. Genet. 8:299-309.

Haldane, J.B.S., and C.A.B. Smith. 1947. A new estimate of the linkage between the genes for color blindness and hemophilia in man. Ann. Eugen. (Lond) 14:10-31.

Haley, C.S., and S.A. Knott. 1992. A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. Heredity 69:315-324.

Hamrick, J.L., Godt, M.J.W., and S.L. Sherman-Broyles. 1992. Factors influencing levels of genetic diversity in woody plant species. *In* Population genetics of forest trees. *Edited by* Adams, W.T., Strauss, S.H., Copes, D.L., and A.R. Griffin. Kluwer Academic Publishers. p 95-124.

Hare, R.C. 1984. Stimulation of early height growth in longleaf pine with growth regulators. Can. J. For. Res. 14:459-462.

Harlow, W. H., Harrar, E. S., and F. M. White. 1978. Textbook of dendrology (Sixth Edition). McGraw-Hill Book Company. 510 pp.

Harry, D.E. 1986. Inheritance and linkage of isozyme variants in incense-cedar. J. Hered. 77:261-266.

Havey, M.J., and F.J. Muehlbauer. 1989. Linkages between restriction fragment length, isozyme, and morphological markers in lentil. Theor. Appl. Genet. 77:395-401.

Helentjaris, T., Slocum, M., Wright, S., Schaefer A., and J. Nienhuis. 1986. Construction of genetic linkage maps in maize and tomato using restriction fragment length polymorphisms. Theor. Appl. Genet. 72:761-769.

Hemmat, M., Weeden, N.F., Manganaris, A.G., and D.M. Lawson. 1994. Molecular marker linkage map for apple. J. Heredity 85:4-11.

Hulbert, T.W., Legg, E.J., Lincoln, S.E., Lander, E.S., and R.W. Michelmore. 1988. Genetic analysis of the fungus, *Bremia lactucae*, using restriction fragment length polymorphisms. Genetics 120:947-958.

Jansen, R.C., and P. Stam. 1994. High resolution of quantitative traits into multiple loci via interval mapping. Genetics 136:1447-1455.

Jarrell, D.C., Roose, M.L., Traugh, S.N., and R.S. Kupper. 1992. A genetic map of citrus based on segregation of isozymes and RFLPs in an intergeneric cross. Theor. Appl. Genet. 84:49-56.

Jensen, J. 1989. Estimation of recombination parameters between a quantitative trait locus (QTL) and two marker gene loci. Theor. Appl. Genet. 78:613-618.

Johns, M.A., 1992. Comparison of RAPD maps for two sets of maize recombinant-inbred lines. Plant Genome I. International Conference of the Plant Genome. Nov. 9-11. San Diego, CA. p 32.

Jones, G.H. 1987. Chiasmata. *In* Meiosis. *Edited by* P.B. Moens. Academic Press, London. p 213-244.

Kais, A.G. 1975. Fungicidal control of *Scirrhia acicola* on longleaf pine seedlings. Plant. Dis. Rep. 59:686-688.

Kais A.G. and M.M. Griggs. 1986. Control of brown spot needle blight on longleaf pine through benomyl treatment and breeding. In: Proceedings of the Conference on Recent Research on Conifer Needle Blight Diseases, Gulfport, Mississippi, October 14-18, 1986. General Technical Report WO-50, Washington DC: USDA Forest Service; 15-19.

Kais, A.G., Snow, G.A., and D.H. Marx. 1981. The effects of benomyl and *Pisolithus tinctorius* ectomycorrhizae on survival and growth of longleaf pine seedlings. South. J. Appl. For. 5:189-195.

Kaufman, C.M. 1977. Growth of young slash pine. For. Sci. 23:217-226.

Kazan, K., Manners, J.M., and D.F. Cameron. 1992. Inheritance of random amplified polymorphic DNA markers in an interspecific cross in the genus *Stylosanthes.* Genome 36:50-56.

Keim, P., Diers, B.W., Olson, T.C., and R.C. Shoemaker. 1990. RFLP mapping in soybean: Association between marker loci and variation in quantitative traits. Genetics 126:735-742.

Kelly, J. F. and W. A. Bechtold. 1989. The longleaf pine resource. Proc. Symp. on the Management of Longleaf pine. April 4-6, 1989. Long Beach, MS. USDA For. Ser. Gen. Tech. Rep. SO-75. South. For. Exp. Sta., New Orleans, La pp. 11-21.

Kesseli, R.V., Paran, I., and R.W. Michelmore. 1994. Analysis of a detailed genetic linkage map of *Lactuca sativa* (Lettuce) constructed from RFLP and RAPD markers. Genetics 136:1435-1446.

King, J.N., and B.P. Dancik. 1983. Inheritance and linkage of isozymes in white spruce (*Picea glauca*). Can. J. Genet. Cytol. 25:430-436.

Knapp, S.J. 1991. Using molecular markers to map multiple quantitative trait loci: models for backcross, recombinant inbred, and doubled haploid progeny. Theor. Appl. Genet. 81:333-338.

Knapp, S.J., Bridges, W.C., and D. Birkes. 1990. Mapping quantitative trait loci using molecular marker linkage maps. Theor. Appl. Genet. 79:583-592.

Komatsuda, T., Annaka, T., and S. Oka. 1993. Genetic mapping of a quantitative trait locus (QTL) that enhances the shoot differentiation rate in *Hordeum vulgare* L. Theor. Appl. Genet. 86(6):713-720.

Kormutak, A. 1975. Karyological structure of some *Pinus* species. Biologia 30:545-550.

Kosambi, D.D. 1944. The estimation of map distances from recombination values. Ann. Eugen. (Lond) 12:172-175.

Kubisiak, T.L., Nelson, C.D., Nance, W.L., and M. Stine. In press. RAPD linkage mapping in a longleaf pine × slash pine F_1 family. Theor. Appl. Genet.

Kubisiak T.L., Nelson C.D., and M. Stine 1993. Simulation study of linkage map construction with missing and mis-scored RAPD data. In: Proceedings of the Twenty-second Southern Forest Tree Improvement Conference, Atlanta, Georgia, June 14-17, 1993. Publication 44 of the Southern Forest Tree Improvement Committee, Springfield, Virginia: The National Technical Information Service; 241-248.

Kubisiak, T.L., Stine, M., Nelson, C.D., and W.L. Nance. 1992. Single tree RAPD linkage mapping of longleaf pine. Proceedings of Plant Genome I, The International Conference on the Plant Genome, Nov. 9-11, San Diego, CA. p 49.

Lalouel, J.M. 1992. Linkage analysis in human genetics. *In* Plant genomes: methods for genetic and physical mapping. *Edited by* Beckmann, J.S., and T.C. Osborn. Kluwer Academic Publishers. p 167-180.

Lande, R. 1981. The minimum number of genes contributing to quantitative variation between and within populations. Genetics 99:541-553.

Lande, R., and R. Thompson. 1990. Efficiency of marker-assisted selection in the improvement of quantitative traits. Genetics 124:743-756.

Lander, E.S., and D. Botstein. 1989. Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121:185-199.

Lander, E.S., and P. Green. 1987. Construction of multilocus linkage maps in humans. Proc. Natl. Acad. Sci. U.S.A. 84:2363-2367.

Lander E.S., Green P., Abrahamson J., Barlow A., Daly M.J., Lincoln S.E., and L. Newburg. 1987. MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1:174-181.

Landry, B.S., Kesseli, R., Farrara, B., and R.W. Michelmore. 1987. A genetic map of lettuce (*Lactuca sativa* L.) with restriction fragment length polymorphism, isozyme, disease resistance and morphological markers. Genetics 116:331-337.

Layton, P. A., and R. E. Goddard. 1982. Environmental and genetic effects on duration of the grass stage of longleaf pine. Proc. 7th N. Am. For. Bio. Workshop. pp.131-136.

Ledig, F.T., and M.T. Conkle. 1983. Gene diversity and genetic structure in a narrow endemic, Torrey pine (*Pinus torreyana* Parry ex Carr). Evolution 37:79-85.

Leonards-Schippers, C., Gieffers, W., Schafer-Pregel, R., Ritter, E., Knapp, S.J., Salamini, F., and C. Gebhardt. 1994. Quantitative resistance of *Phytophthora infestans* in potato: a case study for QTL mapping in allogamous plant species. Genetics 137:67-77.

Loo, J.A., Fowler, D.P., and M.H. Schneider. 1982. Geographic variation in specific gravity among Japanese larch from different provenances. Wood Fiber 14:281-286.

Loveless, M.D. 1992. Isozyme variation in tropical trees: patterns of genetic organization. *In* Population genetics of forest trees. *Edited by* Adams, W.T., Strauss, S.H., Copes, D.L., and A.R. Griffin. Kluwer Academic Publishers. p 67-94.

Lin, S. Y., and H. Ikehashi. 1993. A gamete abortion locus detected by segregation distortion of isozyme locus *Est-9* in wide crosses of rice (*Oryza sativa* L.). Euphytica 67:35-40.

Lincoln, S., Daly, M. J. and E. S. Lander. 1992. Constructing genetic linkage maps with Mapmaker 3.0: A tutorial and reference manual. Whitehead Institute for Biomedical Research.

Lincoln, S.E., and E.S. Lander. 1992. Systematic detection of errors in genetic linkage data. Genomics 14:604-610.

Liu, Z., and G.R. Furnier. 1993. Inheritance and linkage of allozymes and restriction fragment length polymorphisms in trembling aspen. J. Heredity 84:419-424.

Liu, B. H., Sederoff, R., and D. O'Malley. 1993. Linkage mapping using open-pollinated populations. Proc. of the 22nd South. For. Tree Impr. Conf. June 14-17, 1993. Atlanta, GA. p. 489.

Lundkvist, K., and D. Rudin. 1977. Genetic variation in eleven populations of *Picea abies* as determined by isozyme analysis. Hereditas 85:67-74.

Luo, Z.W., and M.J. Kearsey. 1989. Maximum likelihood estimation of linkage between a marker gene and a quantitative locus. Heredity 63:401-408.

Mansur, L.M., Lark, K.G., Kross, H., and A. Oliveira. 1993. Interval mapping of quantitative trait loci for reproductive, morphological, and seed traits of soybean (*Glycine max* L). Theor. Appl. Genet. 86: 907-913.

Martinez, O., and R.N. Curnow. 1992. Estimation of the locations and the sizes of the effects of quantitative trait loci using flanking markers. Theor. Appl. Genet. 85:480-488.

McCouch, S.R., Kochert, G., Yu, Z.H., Wang, Z.Y., Kush, G.S., Coffman, W.R., and S.D. Tanksley. 1988. Molecular mapping of rice chromosomes. Theor. Appl. Genet. 76:815-829.

Michelmore, R. W., Paran, I., and R. V. Kesseli. 1991. Identification of markers linked to disease resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions using segregating populations. Proc. Natl. Acad. Sci. USA 88:9828-9832.

Mitton, J.B. 1992. The dynamic mating system of conifers. *In* Population genetics of forest trees. *Edited by* Adams, W.T., Strauss, S.H., Copes, D.L., and A.R. Griffin. Kluwer Academic Publishers. p 197-216.

Mitton, J.B., Linhart, Y.B., Sturgeon, K.B., and J.L. Hamrick. 1979. Allozyme polymorphisms detected in mature needle tissue of ponderosa pine. J. Heredity 70:86-89.

Montgomery, D.C. 1984. Design and analysis of experiments. Second Edition. John Wiley & Sons. pp. 538.

Moran, G.F. 1992. Patterns of genetic diversity in Australian tree species. *In* Population genetics of forest trees. *Edited by* Adams, W.T., Strauss, S.H., Copes, D.L., and A.R. Griffin. Kluwer Academic Publishers. p 49-66.

Morton, N.E. 1955. Sequential tests for the detection of linkage. Am. J. Hum. Genet. 7:277-318.

Muller-Starck, G., Baradat, P.H., and F. Bergmann. 1992. Genetic variation within European tree species. *In* Population genetics of forest trees. *Edited by* Adams, W.T., Strauss, S.H., Copes, D.L., and A.R. Griffin. Kluwer Academic Publishers. p 23-48.

Muona, O., Yazdani, R., and G. Lindqvist. 1987. Analysis of linkage in *Picea abies.* Hereditas 106:31-36.

Nance, W. L., and C. D. Nelson. 1989. Restriction fragment length polymorphisms and their potential use in marker assisted selection in southern pine improvement. Proc. 20th S. Tree Improv. Conf. pp 50-59.

Nance, W.L., Nelson, C.D., Stine, M., Kubisiak. T.L., van Buijtenen, J.P., Kong, X., and R.L. Doudrick. 1992a. A RAPD-based genetic linkage map of slash pine (Pinus elliottii Engelm. var. elliottii) with comparisons to other RAPD-based maps. Plant Genome I. International Conference of the Plant Genome. Nov. 9-11. San Diego, CA. p 60.

Nance, W.L., Tuskan, G.A., Nelson, C.D., Doudrick, R.L. 1992b. Potential applications of molecular markers for genetic analysis of host-pathogen systems in forest trees. Can. J. For. Res. 22:1036-1043.

Neale, D. B., and C. G. Williams. 1991. Restriction fragment length polymorphism mapping in conifers and applications to forest genetics and tree improvement. Can. J. For. Res. 21:545-554.

Nelson, C. D., Kubisiak, T. L., Stine, M. and W. L. Nance. In press. A genetic linkage map of longleaf pine based on random amplified polymorphic DNAs (J. Heredity).

Nelson, C.D., Nance, W.L., and R.L. Doudrick. 1992. A partial genetic linkage map of slash pine (*Pinus elliottii* var. *elliottii*) based on randomly amplified polymorphic DNAs. Proceedings of Plant Genome I, The International Conference on the Plant Genome, Nov. 9-11, San Diego, CA. p 39.

Nelson, C.D., Nance, W.L., and R.L. Doudrick. 1993. A partial genetic linkage map of slash pine (*Pinus elliottii* Englm. var *elliottii*) based on random amplified polymorphic DNAs. Theor. Appl. Genet. 87:145-151.

Newbury, H.J., and B.V. Ford-Lloyd. 1993. The use of RAPD for assessing variation in plants. Plant Growth Regulation 12:43-51.

Niebling, C.R., Johnson, K., and H.D. Gerhold. 1987. Electrophoretic analysis of genetic linkage in scots pine (*Pinus sylvestris* L.). Biochem. Genet. 25:803-814.

Nybon, H., and S.H. Rogstad. 1990. DNA "fingerprints" detect genetic variation in *Acer negundo* (*Aceraceae*). Plant Syst. Evol. 173:49-56.

O'Brien, S.J. 1993. Genetic maps: locus maps of complex genomes. Sixth Edition. Cold Spring Harbor Press.

Ohri, D., and T.N. Khoshoo. 1986. Genome size in gymnosperms. Plant Sys. Evol. 153:119-132.

Ott, J. 1985. Analysis of human genetic linkage. The John Hopkins Press Ltd., London. 223 p.

Ott, J. 1991. Analysis of human genetic linkage. The John Hopkins Press Ltd., London. 302 p.

Paran, I., Kesseli, R., and R. Michelmore. 1991. Identification of restriction fragment length polymorphism and random amplified polymorphic DNA markers linked to downy mildew resistance genes in lettuce, using near-isogenic lines. Genome 34:1021-1027.

Paran, I., and R.W. Michelmore. 1993. Development of reliable PCR-based markers linked to downy mildew resistance genes in lettuce. Theor. Appl. Genet. 85:985-993.

Paterson, A.H., Damon, S., Hewitt, J.D., Zamir, D., Rabinowitch, H.D., Lincoln, S.E., Lander, E.S., and S.D. Tanksley. 1991. Mendelian factors underlying quantitative traits in tomato: comparison across species, generations, and environments. Genetics 127:181-197.

Paterson, A.H., Lander, E.S., Hewitt, J.D., Peterson, S., Lincoln, S.E., and S.D. Tanksley. 1988. Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. Nature 335:721-726.

Perry, D.J., and P. Knowles. 1989. Inheritance and linkage relationships of allozymes of eastern white cedar (*Thuja occidentalis*) in northwestern Ontario. Genome 32:245-250.

Pham, J. L., and B. Bougerol. 1993. Abnormal segregations in crosses between 2 cultivated rice species. Heredity 70:466-471.

Powers Jr., H.R., and F.R., Matthews. 1980. Comparison of six geographic sources of loblolly pine for fusiform rust resistance. Phytopathology 70:1141-1143.

Reiter R.S., Williams J.G.K., Feldmann K.A., Rafalski J.A., and S.V. Tingey. 1992. Global and local genome mapping of *Arabidopsis thaliana* by using recombinant inbred lines and random amplified polymorphic DNAs. Proc Nat Acad Sci USA 89:1477-1481.

Richards, A. J. 1986. Plant breeding systems. George Allen & Unwin (Publishers) Ltd. 529 pp.

Ritter, E., Gebhardt, C., and F. Salamini. 1990. Estimation of recombination frequencies and construction of RFLP linkage maps in plants from crosses between heterozygous parents. Genetics 125:645-654.

Roy, A., Frascaria, N., MacKay, J., and J. Bousquet. 1992. Segregating random amplified polymorphic DNAs (RAPDs) in *Betula alleghaniensis*. Theor. Appl. Genet. 85:173-180.

Rudin, D. 1977. Leucine-amino-peptidases (LAP) from needles and macrogametophytes of Pinus sylvestris L. Inheritance of allozymes. Hereditas 85:219-226.

Rudin, D., and I. Ekberg. 1978. Linkage studies in *Pinus sylvestris* L. - using macrogametophyte allozymes. Silvae Genetica 27:1-12.

Sall, T., and B.O. Bengtsson. 1989. Apparent negative interference due to variation in recombination frequencies. Genetics 122:935-942.

Sall, T., and N.-O. Nilsson. 1994. The robustness of recombination frequency estimates in intercrosses with dominant markers. Genetics 137:589-596.

SAS Institute Inc., SAS/STAT® User's Guide, Version 6, Fourth Edition, Volume 1, Cary, NC: SAS Institute Inc., 1989. 943 pp.

Sax, K. 1923. The association of size differences with seed-coat pattern and pigmentation in *Phaseolus vulgaris*. Genetics 8:552-560.

Saylor, L.C. 1972. Karyotype analysis of the genus *Pinus* subgenus *Pinus*. Silvae Genet. 19:155-163.

Schmidtling, R. C., and T. L. White. 1989. Genetics and tree improvement of longleaf pine. Gen. Tech. Rep. SO-75. 13 pp.

Shattuck-Eidens, D.M., Bell, R.N., and T. Helentjaris. 1992 *In* Plant genomes: methods for genetic and physical mapping. *Edited by* J.S. Beckmann and T.C. Osborn. Kluwer Academic Publishers. p 55-70.

Shaw, C.R., and R. Prasad. 1970. Starch gel electrophoresis of enzymes - a compilation of recipes. Biochem. Genet. 4:297-320.

Shields, D.C., Collins, A., Buetow, K.H., and N.E. Morton. 1991. Error filtration, interference, and the human linkage map. Proc. Natl. Acad. Sci. USA 88:6501-6505.

Siggers, P.V. 1944. The brown spot needle blight of pine seedlings. U.S. Dept. Agr. Tech. Bul. 870, 36 p.

Simpson, S.P. 1989. Detection of linkage between quantitative trait loci and restriction fragment length polymorphisms using inbred lines. Theor. Appl. Genet. 77:815-819.

Snyder, E. B., and H. J. Derr. 1972. Breeding longleaf pines for resistance to brown spot needle blight. Phytopathology 62:325-329.

Soller, M., Genizi, A., and T. Brody. 1976. On the power of experimental designs for the detection of linkage between marker loci and quantitative trait loci in crosses between inbred lines. Theor. Appl. Genet. 47:35-39.

Soller, M., and J. S. Beckmann. 1983. Genetic polymorphism in varietal identification and genetic improvement. Theor. Appl. Genet. 67:25-33.

Strauss, S.H., and W.T. Conkle. 1986. Segregation, linkage and diversity of allozymes in knobcone pine. Theor. Appl. Genet. 72:483-493.

Strauss, S.H., Lande, R., and G. Namkoong. 1992. Limitations of molecular-aided-selection in forest tree breeding. Can. J. For. Res. 22:1050-1061.

Stuber, C.W., Lincoln, S.E., Wolff, D.W., Helentjaris, T., and E.S. Lander. 1992. Identification of genetic factors contributing to heterosis in a hybrid from two elite maize inbred lines using molecular markers. Genetics 132:823-839.

Sturtevant, A.H. 1913. The linear arrangement of six sex-linked factors in Drosophila, as shown by their mode of association. J. Exp. Zool. 14:43-59.

Tanksley S.D., Medina-Filho H., and C.M. Rick. 1982. Use of naturally-occurring enzyme variation to detect and map genes controlling quantitative traits in an interspecific backcross of tomato. Heredity 49:11-25.

Tauer, C.G., Hallgren, S.W., and B. Martin. 1992. Using marker-aided selection to improve tree growth response to abiotic stress. Can. J. For. Res. 22:1018-1030.

Thoday, J.M. 1961. Location of polygenes. Nature 191:368-370.

Thomas, M.R., and N.S. Scott. 1993. Microsatellite repeats in grapevine reveal DNA polymorphisms when analyzed as sequence-tagged sites (STSs). Theor. Appl. Genet. 86:985-990.

Tingey S.V., Rafalski J.A., and J.G.K. Williams. 1992. Genetic analysis with RAPD markers. In: Applications of RAPD technology to Plant Breeding, Minneapolis, Minnesota, Nov 1, 1992. Crop Science Society of America; 3-8.

Tulsieram, L. K., Glaubitz, J. C., Kiss, G. and J. E. Carlson. 1992. Single tree genetic linkage mapping in conifers using haploid DNA from megagametophytes. Bio/Technology 10:686-690.

USDA Forest Service. 1965. Silvics of forest trees of the United States. USDA-Forest Service Agriculture Handbook No. 271. pp. 762.

USDA Forest Service. 1974. Seeds of woody plants in the United States. USDA-Forest Service Agriculture Handbook No. 450. pp 883.

USDA Forest Service. 1982. 1981 directory of forest tree seed orchards in the United States. USDA-Forest Service FS-278, pp. 47.

van Buijtenen, J.P., Kong, X., Funkhouser, E., Nance, W.L., Nelson, C.D., Nelson, L.S., and G.N. Johnson. 1992. Linkage map of slash pine based on megagametophytic DNA. Proceedings of Plant Genome I, The International Conference on the Plant Genome, Nov. 9-11, San Diego, CA. p 51.

Wagner, D. B., Furnier, G. R., Saghai-Maroof, M. A., Williams, S. M., Dancik, B. P., and R. W. Allard. 1987. Chloroplast DNA polymorphisms in lodgepole and jack pines and their hybrids. Proc. Natl. Acad. Sci. USA 84:2097-2100.

Wang, .G-L., Mackall, D.J., Bonman, M.J., McCouch, S.R., Champoux, M.C., and R.J. Nelson. 1994. RFLP mapping of genes conferring complete and partial resistance to blast in a durably resistant rice cultivar. p 1421-1434.

Weeden, N.F. 1994. Approaches to mapping in horticultural crops. *In* Current topics in plant molecular biology: plant genome analysis. *Edited by* P.M. Gresshoff. CRC Press. p 57-68.

Weeden, N.F., and M. Hemmat. 1993. *Malus domestica*, apple. *In* Genetic maps: locus maps of complex genomes. Sixth Edition. *Edited by* S.J. O'Brien. Cold Spring Harbor Laboratory Press. p 6.53-6.58.

Weeden, N.F., Timmerman, G.M., Hemmat, M., Kneen, B.E., and M.A. Lodhi. 1992. Inheritance and reliability of RAPD markers. Proc. of the Symp. Appl. of RAPD Tech. to Plant Breed. November 1, 1992. Minneapolis, Minnesota p 12-17.

Weisberg, S. 1985. Applied linear regression. Second Edition. John Wiley & Sons pp. 324.

Weller, J.I. 1986. Maximum likelihood techniques for the mapping and analysis of quantitative trait loci with the aid of genetic markers. Biometrics 42:627-640.

Weller, J.I. 1992. Statistical methodologies for mapping and analysis of quantitative trait loci. *In* Plant genomes: methods for genetic and physical mapping. *Edited by* Beckmann, J.S., and T.C. Osborn. Kluwer Academic Publishers. p 181-207.

Wells, O. O., and E. B. Snyder. 1976. Longleaf pine half-sib progeny test. For. Sci. 22(3):404-406.

Welsh, J., and M. McClelland. 1990. Fingerprinting genomes using PCR with arbitrary primers. Nuc. Acids Res. 18:7213-7218.

Williams, J.G.K., Hanafey, M.K., Rafalski, J.A., and S.V. Tingey. 1993. Genetic analysis using random amplified polymorphic DNA markers. Methods in Enzymology 218:705-740.

Williams, J.G.K., Kubelik, A.R., Livak, K.J., Rafalski, J.A., and S.V. Tingey. 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Research 18:6531-6535.

Williams, C.G., and D.B. Neale. 1992. Conifer wood quality and marker-aided selection: a case study. Can. J. For. Res. 22:1009-1017.

Williams, M.N.V., Pande, N., Nair, S., Mohan, M., and J. Bennett. 1991. Restriction fragment length polymorphism analysis of polymerase chain reaction products amplified from mapped loci of rice (*Oryza sativa* L.) genomic DNA. Theor. Appl. Genet. 82:489-498.

Witter, M.S., and P.P. Feret. 1980. Inheritance of esterase and acid phosphatase isozymes in Virginia pine *Pinus virginiana* and the application of the isozyme technique to a seed orchard population. Silvae Genetica 28:213-220.

Xie, C.Y., Dancik, B.P., and F.C. Yeh. 1991. Inheritance and linkage of isozymes in *Thuja orientalis*. J. Heredity 82:329-334.

Young, N.D., Danesh, D., Menancio-Hautea, D., and L. Kumar. 1993. Mapping oligogenic resistance to powdery mildew in mungbean with RFLPs. Theor. Appl. Genet. 87:243-249.

Zeng, Z-B. 1992. Correcting the bias of Wright's estimate of the number of genes affecting a quantitative character: A further improved method. Genetics 131:987-1001.

Zeng, Z-B. 1994. Precision mapping of quantitative trait loci. Genetics 136:1457-1468.

Zeng, Z-B, Houle, D., and C.C. Cockerham. 1990. How informative is Wright's estimator of the number of genes affecting a quantitative character? Genetics 126:235-247.

Zobel, B., and J. Talbert. 1984. Applied forest tree improvement. John Wiley & Sons, Inc. 505 pp.

APPENDIX A

PERMISSION AND PROOF OF AUTHORSHIP

The University of Georgia
Cooperative Extension Service
College of Agricultural and Environmental Sciences

Extension Forest Resources
P. O. Box 1209
Tifton, GA 31793
July 27, 1994

T.L. Kubisiak
School of Forestry, Wildlife and Fisheries
227 Forestry, Wildlife & Fisheries Bldg.
Louisiana State University
Baton Rouge, LA 70803

Dear Tom:

There are no restrictions in the use in your dissertation of the paper by Kubisiak, Nelson, and Stine as it was published in the Proceedings of the 22nd Southern Forest Tree Improvement Conference. The Southern Forest Tree Improvement Committee holds no copyright claims to papers published in their proceedings.

Sincerely,

David J. Moorhead
Associate Professor

pb

THEORETICAL
AND
APPLIED GENETICS

International Journal
of Breeding Research and Cell Genetics

Springer
Springer International

Bundesanstalt für Züchtungsforschung an Kulturpflanzen
Institut für Resistenzgenetik · D-85461 Grünbach · Germany

Dr. T. L. Kubisiak
Louisiana Agricult. Exp. Stat.
Louisiana State University
Agricultural Center
School of Forestry, Wildlife and Fisheries
Baton Rouge, LO 70803
U.S.A.

Federal Centre for Breeding Research
on Cultivated Plants
Institute for Resistance Genetics
Graf-Seinsheim-Straße 23
D-85461 Grünbach
Germany
Telephone: (0) 8122/1651
Telefax: (0) 8122/40642

Your letter of

Date:

Oct. 24, 1994

Dear Dr. Kubisiak,

referring to your letter dated October 12, 1994, I am able to tell you that TAG grants you permission for reprinting the mentioned article in your dissertation.

Sincerely yours,

G. Wenzel

APPENDIX B

RAPD MARKER[3], ALLOZYME, AND QTL DATA SETS

Longleaf pine clone 27-168 preliminary RAPD marker data set (#1)

Primer[1]	~Band Size (bp)	Presence or Absence of RAPD Band in 12 Megagametophytes[2]
A04	0500	-HAA-AHHHHHH
A04	0600	-AAH-HHHHAAH
A05	0525	-HHA-AAHHAAA
A07	0325	-HAA-AHHHAHH
A07	0375	-HHA-AHAHHAH
A07	0800	-AHA-HAHAHAA
A07	1650	-HAH-HHAAAHH
A11	0425	-AHA-HHHAAHH
A12	0650	-HAH-HHHAHAA
A15	1625	AHHHHHAHHH--
A16	1050	HAHHHAAAHH--
A16	1150	HAAHAHAAHH--
A17	0400	AHHAAHAHAA--
A17	0700	HHAAAHHHAH--
A17	0775	HHHAAHHHAH--
A18	1050	AHAHAHAHHA--
B04	1600	HHHAHHAHHH--
B04	2000	AHAAAAHAHH--
B05	1500	HHHHAAAHHA--
B06	0700	AHHAHAHAHA--
B06	0800	AAAHHHHHHA--
B08	0385	AAAHHHHAAH--
B08	0700	AAHHHAAAAH--
B08	1000	HHHAAHHHHA--
B08	1400	AAAAHHHHHA--
B12	0750	HHAHHHHAAA--
B15	0425	HAHAAAHHAH--
B15	0550	HHAAAAHHHH--
B18	0375	HAHHAHAAHH--
B18	0700	AAHHHHAAAA--
B20	1150	AAAHHAHAAH--
E02	1300	AAHAAHHAHA-
E02	0750	HHHHHHAAHA-
E02	1500	AAHHHAAAAH-
E06	0900	HAHAHHAAAH-
E08	0600	H-HAHAHHAH-
E08	0850	H-HHHHAHAA-

[3]Both the haploid and diploid-based RAPD data sets are available on the Internet in a genome database for forest trees, called TreeGenes. Instructions for getting TreeGenes by file transfer protocol can be obtained by sending electronic mail to **dendrome@s27w007.pswfs.gov**.

E08	1300	A-AHAAHHAH-
E11	0950	A-AAAAHAHH-
E12	0600	H-AHHHAHAA-
E12	0900	H-AHHHAAHH-
E12	1600	A-HAHHAHAA-
E13	1000	--AHHAHHHH-
E17	0700	H-AHHAAHAH-
E19	1400	H-AAHAAHAH-
E19	2000	H-HHHAAAAA-
G04	0500	H--AAAAAHHH
G04	1500	A--AHAAHAHH
G04	1600	H--HAHAAHAA
G07	2000	---AHAHHHHH
G09	0750	H--AHHAAAAA
G09	1500	H--AHAAHAHH
G11	0450	A-AHHHHHAAA
G11	1200	H-HHHHAAHAH
G12	0375	A-HHHAAHAHH
G12	0800	H-AHHHAHAAA
G13	0950	A-AHAAHHHHH
G13	0900	A-AAAAHHHAA
G15	1200	H-HAHHAHAAH
G15	1600	A-HHAAHHHAA
G15	0775	A-AAAHHHHAA
G18	0650	H-AHHAAHHHA
W01	0600	--AHHAAHAHA
W11	0450	H-HHAHHHHAA
W11	1000	A-HAHAHAAHA
W14	0900	H-HHAHHHAAA
W15	0600	H-AHAAHHAH-
X04	1600	H-AHAHAHHAA
X09	1000	H-AAAHAHHAA
X16	1200	A-AAAAHHAHH
X16	1250	H-AAHAAAHHH
X18	0450	H-HAHHHAAHA
X18	0575	A-AHAHHHAAH
X18	0800	H-AAAHAHHAA
X20	1000	A--AHAAHHHH
Y02	1050	H-AAAHAAHHA
Y04	0525	A-HAAAHHAHH
Y04	0700	A-AAHHAAAHH
Y13	0600	A-AHAHHHAAA
Y14	0400	A-AHAHAHAHA
Y15	1250	A-AHHAAHAAA
Y16	2000	--HHHAHAAAA
Y17	1600	H-HHAHA-HHA
Y18	0675	H-AAAHAAHAA
Y20	0800	H-AAAHAAHHA

[1]Operon Technologies Inc. (Alameda, CA) primer designation
[2]H = band present; A = band absent; - = missing data

Allozyme data scored on 40 F_1 progeny from cross between slash pine H-28 (♂) × longleaf pine 3-356 (♀).

Slash Pine H-28

	Allozyme[1]		
Progeny ID	PGI_2	LAP_2	6PGD_1
1	3	1	3
3	3	1	1
7	1	2	1
11	3	2	3
14	1	2	1
16	1	2	3
19	3	1	1
20	1	1	3
21	1	2	3
23	1	2	1
24	3	1	3
27	3	2	1
29	1	1	1
30	1	1	1
31	1	1	3
32	3	1	3
35	3	1	1
36	3	2	3
37	1	1	3
42	3	1	1
43	1	1	3
44	1	1	1
46	3	2	3
47	1	2	1
48	3	1	1
50	1	1	1
51	3	2	3
53	3	1	1
54	3	2	1
55	1	1	3
56	3	1	1
57	1	-	-
58	1	2	3
61	1	2	3
63	1	1	1
64	1	1	3
65	3	2	1
78	1	2	1
82	3	1	1
84	1	1	3

[1]1,2,3 = allele designation (see Figure 6.1); - = missing data

Allozyme data scored on 40 F_1 progeny from cross between slash pine H-28 (♂) × longleaf pine 3-356 (♀).

Longleaf pine 3-356

	Allozyme[1]			
Progeny ID	PGI_2	EST	6PGD_1	MDH_2
1	4	3	2	1
3	2	2	2	1
7	4	2	1	1
11	2	2	1	2
14	2	2	2	1
16	4	3	1	2
19	4	2	2	2
20	4	2	1	1
21	4	2	1	1
23	2	3	2	2
24	4	3	2	1
27	4	3	1	1
29	2	2	1	1
30	2	3	1	2
31	4	2	2	1
32	4	3	2	1
35	2	2	2	2
36	2	2	2	2
37	4	3	2	2
42	4	3	2	1
43	4	2	1	2
44	2	3	2	2
46	4	2	2	2
47	2	3	1	2
48	2	2	2	2
50	4	2	1	1
51	2	2	2	1
53	2	2	2	2
54	2	3	2	1
55	4	3	2	1
56	4	3	1	1
57	2	2	-	1
58	4	2	2	2
61	4	3	1	1
63	2	2	1	2
64	4	2	1	1
65	2	3	2	2
78	4	2	1	1
82	2	3	2	1
84	2	2	2	1

[1]1,2,3,4 = allele designation (see Figure 6.1); - = missing data

Early height growth measurements on F_1 progeny from cross between slash pine H-28 (♂) × longleaf pine 3-356 (♀).

Progeny	Trait1	Trait2	Trait3	Trait4	Trait5	Trait6	Trait7	Trait8	Trait9	Trait10[1]
1	1.8	2.0	6.0	6.0	25.0	37.0	1.6	1.8	13	39.8
2	1.4	1.6	5.8	5.0	23.0	26.0	1.0	1.0	1	52
3	1.0	1.1	8.0	8.0	34.0	44.0	1.5	1.7	16	46.8
4	2.5	2.8	8.0	6.5	17.0	21.0	1.0	0.9	2	31.5
5	1.8	1.7	8.4	7.5	20.0	23.0	0.9	0.9	6	26.8
6	1.4	1.6	5.4	5.0	18.0	23.0	0.9	1.0	5	27.6
7	1.9	2.4	7.0	7.5	24.0	45.0	1.0	1.3	18	47.5
8	2.8	2.8	7.0	7.0	28.0	32.0	1.0	1.1	6	37.3
9	1.7	2.2	6.4	7.0	19.0	29.0	1.0	1.0	4	36.3
10	0.9	1.0	7.5	6.0	17.0	26.0	1.0	1.1	8	29.3
11	1.4	1.8	9.9	10.5	30.0	41.0	1.5	1.8	12	44.4
12	1.9	2.1	6.7	6.5	19.0	30.0	1.0	1.0	5	36
13	2.1	2.2	8.3	7.5	26.0	39.0	1.4	1.6	13	42
14	1.7	1.8	7.5	8.0	26.0	45.0	1.5	1.9	12	48.8
15	1.2	1.3	6.3	6.0	24.0	30.0	1.4	1.5	3	40
16	1.1	1.4	7.4	8.5	14.0	25.0	1.4	1.2	8	28.1
17	2.1	2.2	7.9	7.5	15.0	15.0	0.8	0.7	2	22.5
18	1.8	1.8	7.9	9.0	20.0	28.0	1.0	1.1	3	37.3
19	1.2	1.3	7.5	8.5	32.0	56.0	1.7	2.1	32	57.8
20	1.4	1.4	7.4	7.0	29.0	51.0	1.6	1.8	20	53.6
21	1.4	1.4	10.2	10.0	29.0	32.0	1.1	1.0	4	40
22	1.3	1.6	7.3	7.0	10.0	14.0	0.8	0.8	3	18.7
23	1.8	2.4	8.5	7.5	22.0	29.0	1.2	1.5	2	43.5
24	0.8	0.5	1.5	5.5	8.0	.	0.5	.	.	.
25	2.2	2.0	9.0	8.0	32.0	43.0	1.6	1.7	6	50.2
26	2.1	2.3	10.0	9.0	45.0	60.0	1.7	2.0	20	63
27	1.5	1.6	5.7	4.0	17.0	21.0	1.1	1.1	1	42
28	2.1	2.6	6.2	7.5	20.0	35.0	1.6	1.5	7	40
29	2.0	2.2	7.6	10.0	22.0	38.0	1.5	1.5	8	42.8
30	1.4	1.6	7.7	7.5	12.0	23.0	1.1	1.1	6	26.8
31	1.3	1.0	7.0	6.5	21.0	25.0	1.4	1.2	3	33.3
32	1.6	1.8	5.0	3.0	16.0	11.0	1.0	0.8	4	13.75
33	1.0	1.2	5.0	6.5	32.0	41.0	1.4	1.5	13	44.2
34	1.8	2.1	8.5	8.0	23.0	36.0	1.3	1.5	8	40.5
35	2.0	2.6	8.7	7.5	27.0	36.0	1.2	1.2	2	54
36	1.4	1.5	8.4	7.0	17.0	29.0	1.2	1.2	4	36.3
37	1.5	1.1	7.5	9.0	36.0	58.0	1.7	2.1	26	60.2
38	1.6	1.8	6.8	9.5	20.0	27.0	1.3	1.5	3	36
39	1.8	1.7	7.8	7.0	27.0	46.0	1.5	1.7	13	49.5
40	1.4	1.3	6.1	6.0	20.0	33.0	1.4	1.5	6	38.5
41	1.3	1.5	6.8	7.0	21.0	28.0	1.0	1.1	2	42
42	1.1	1.2	4.5	4.0	15.0	.	1.5	.	.	.
43	1.2	1.5	7.9	8.0	24.0	40.0	1.7	1.8	10	44
44	1.1	1.2	6.8	8.5	24.0	29.0	1.3	1.2	4	36.3
45	1.1	1.8	8.0	9.0	28.0	43.0	1.4	1.6	13	46.3
46	1.0	1.5	9.7	10.0	33.0	46.0	1.5	1.5	9	51.1
47	1.1	1.5	10.2	8.5	35.0	68.0	0.9	2.0	35	69.9
48	1.5	2.2	9.0	9.0	37.0	54.0	1.7	2.0	23	56.3

49	1.2	1.4	6.0	6.5	17.0	30.0	1.4	1.4	7	34.3
50	1.0	1.0	4.8	4.0	25.0	33.0	1.3	1.3	5	39.6
51	1.8	1.9	8.7	9.5	20.0	25.0	1.1	1.2	4	31.3
52	1.5	2.0	7.0	7.5	18.0	24.0	1.2	1.3	6	28
53	2.2	2.7	8.5	8.0	35.0	50.0	1.0	1.6	14	53.6
54	1.2	1.4	7.0	8.5	32.0	56.0	1.8	2.0	17	59.3
55	1.6	2.2	7.9	8.0	31.0	43.0	1.5	1.8	13	46.3
56	0.7	0.9	7.0	8.0	22.0	28.0	1.5	1.5	12	30.3
57	1.7	2.6	7.9	8.0	11.0	26.0	1.2	1.6	17	27.5
58	1.9	2.0	7.3	9.0	20.0	47.0	1.2	1.9	32	48.5
59	1.7	2.5	6.0	6.0	21.0	34.0	1.6	1.8	6	39.7
60	1.5	2.2	5.5	5.5	24.0	33.0	1.8	1.6	5	39.6
61	2.5	2.6	8.9	10.5	30.0	42.0	1.1	1.4	11	45.8
62	1.1	1.5	8.3	9.5	31.0	49.0	1.5	1.7	19	51.6
63	1.1	1.5	6.5	6.0	30.0	48.0	1.4	1.8	24	50
64	1.2	1.5	6.8	7.0	22.0	45.0	1.5	1.6	15	48
65	1.1	1.3	6.9	7.0	23.0	39.0	1.4	1.3	4	48.8
66	1.5	2.0	9.7	12.5	41.0	60.0	1.7	2.1	25	62.4
67	1.8	2.0	7.8	11.0	33.0	54.0	1.5	2.1	24	56.3
68	1.9	1.9	6.7	10.0	30.0	49.0	1.5	1.9	24	51.0
69	1.2	2.2	6.2	6.5	15.0	21.0	1.3	1.7	6	24.5
70	1.1	1.8	6.5	6.0	23.0	29.0	1.6	2.0	8	32.6
71	0.9	1.3	10.7	10.0	28.0	49.0	1.6	2.2	18	51.7
72	1.5	1.8	6.2	7.0	24.0	38.0	1.2	1.6	16	40.4

[1]Trait1 = hypocotyl length (2 months)
Trait2 = hypocotyl length (3 months)
Trait3 = total height (3months)
Trait4 = total height (5 months)
Trait5 = total height (9 months)
Trait6 = total height (21 months)
Trait7 = root collar diameter (11 months)
Trait8 = root collar diameter (21 months)
Trait9 = brown spot infection
Trait 10 = adjusted height at 21 months

VITA

Thomas L. Kubisiak was born the youngest of six children on August 17, 1966 in Midland, Michigan. His father passed away prior to his third birthday and he therefore was greatly influenced by the teachings of his mother, older siblings, and immediate relatives. Thomas infrequently stayed inside and loved to roam the parks and woodlots near his home. His preoccupation with nature and the outdoors was apparent to his mother at an early age, but not fully recognized by himself until his late teens. Thomas graduated from Midland High School in the summer of 1984, and subsequently enrolled at Michigan State University with the intent of obtaining a degree in chemical engineering. During his second year in college he realized that chemical engineering was not where his true interests lay, but was more the influence of his family and relatives. After much self reflection Thomas decided to pursue a career in Forestry. He graduated with a B.S. in Forestry in the summer of 1988 and subsequently enrolled in a Masters of Science program in Forestry at the University of Minnesota. He obtained his degree in December of 1990, after which he immediately enrolled in a Doctor of Philosophy program in Forestry at Louisiana State University. Upon completion of his Ph.D., Thomas hopes to apply his knowledge towards the genetic improvement of various economically important timber species. His feelings are that there will be an even greater need in the future for more productive forests as current public and private land-bases dwindle due to such factors as a rising human population, a push for multiple-use, and the continued protection of land for endangered species.

DOCTORAL EXAMINATION AND DISSERTATION REPORT

Candidate: Thomas Lester Kubisiak

Major Field: Forestry

Title of Dissertation: Molecular Marker Linkage Mapping in Southern Pine (Longleaf Pine and Slash Pine)

Approved:

Major Professor and Chairman

Dean of the Graduate School

EXAMINING COMMITTEE:

Date of Examination:

October 27, 1994