RESTRICTION FRAGMENT LENGTH POLYMORPHISM MAPPING OF LOBLOLLY PINE: METHODS, APPLICATIONS, AND LIMITATIONS

D. B. Neale, C. G. Tauer, D. M. Gorzo, and K. D. Jermstad[1/]

Abstract.--A saturated restriction fragment length polymorphism (RFLP) linkage map for loblolly pine (Pinus taeda L.) is being constructed. RFLPs are mapped by classical segregation analysis in 3-generation pedigrees provided by the North Carolina State University- Industry Cooperative Tree Improvement Program. The map will provide a detailed description of the organization of the loblolly pine genome and serve as an important tool for tree breeding and gene transfer. It may be possible to use the map to identify quantitative trait loci (QTLs) and for marker-assisted early selection, although, several factors presently limit the utility of these applications in loblolly pine and other conifers.

Keywords: RFLPs, genetic mapping, Pinus taeda L.

INTRODUCTION

We are constructing a high-density genetic map for loblolly pine (Pinus taeda L.) based on restriction fragment length polymorphism (RFLP) markers. Loblolly pine was chosen among conifer species because of its commercial importance, immediate availability of 3-generation pedigrees, small genome size relative to many other conifers, and membership in the genus Pinus. Furthermore, loblolly pine is a likely candidate for improvement via gene transfer technology, which will be greatly facilitated by a genetic map.

Our goal is to understand more about the organization of conifer genomes. Aside from information obtained from karyotype analysis (Saylor 1972) and isozyme analysis (Adams and Joly 1980a, 1980b), little is known about the organization of the loblolly pine genome. By mapping a large number (200-300) of RFLP markers, we hope to obtain a detailed description of how the genome of this species is organized.

Aside from addressing basic questions on genome organization in loblolly pine, there will be direct applications of the genetic map. These include (1) the acquisition of a large number of new genetic markers for population genetic studies or for monitoring tree breeding activities, (2) identification of insertion sites following DNA transfer; and (3) use in

[1/] Geneticist, Sabbatical Scientist (Oklahoma State University), Postdoctoral, and Biologist, Institute of Forest Genetics, U.S.D.A. Forest Service, Placerville, CA 95667.

conjunction with other technologies to isolate individual genes and chromosomal segments encoding quantitative trait loci (QTL). There may also be an opportunity to use RFLPs for marker-assisted early selection.

RESTRICTION FRAGMENT LENGTH POLYMORPHISMS

A restriction fragment length polymorphism (RFLP) is most easily envisioned as a simply inherited Mendelian trait viewed at the molecular level. RFLPs are variations in DNA sequence between homologous chromosomes and can be detected using standard molecular biology protocols. First, DNA samples are isolated from individual trees; needle tissue is commonly used for conifers, although DNA can be isolated from most tissues. Then, DNAs are cut into discrete fragments by adding a restriction endonuclease. These enzymes recognize specific DNA sequences (usually 4-8 base pairs) and cleave DNA at or near these sites. Over 500 restriction enzymes have been isolated, mostly from microorganisms. The DNA fragments are size fractionated by gel electrophoresis. Restriction fragments are then transferred to a solid support membrane by a technique called Southern blotting (Southern 1975).

The next step in detecting RFLPs is to hybridize Southern blots with DNA probes. Radioactively labelled DNA probes are hybridized to homologous DNA sequences on the blots. The unbound probe is washed from the blot before exposing it to X-ray film. The resulting autoradiogram reveals only DNA sequences that are homologous to the probe. Variations in the position of bands among DNAs on the autoradiogram are the result of DNA sequence variation and are thus called restriction fragment length polymorphisms. RFLPs are usually the result of a gain or loss of restriction sites due to mutations within recognition sites, or are due to small insertions or deletions.

RFLPs have numerous advantages over other types of markers for genetic mapping. Contrasted with isozymes and morphological markers, RFLPs provide a virtually unlimited number of genetic markers. RFLP markers are developmentally stable and display normal Mendelian inheritance. RFLPs generally exhibit multiple alleles, codominance, minimal pleiotrophic effects, and little if any environmental effect (Beckman and Soller 1983, 1986). At present, RFLPs are the most effective, if not the only, method available for constructing a saturated genetic linkage map for loblolly pine.

Several problems associated with RFLP mapping will, at first, limit the widespread application of this technology to forest trees. One is the need for experience in molecular biology methodology which is just being developed in forestry research organizations. These methods must then be adapted to forest trees. Finally, the cost of this technology can be prohibitively high for some research organizations.

METHODS FOR CONSTRUCTING AN RFLP MAP FOR LOBLOLLY PINE

Isolation of Conifer DNA

Total loblolly pine genomic DNA is prepared from fresh needle tissue following modifications in the procedure of Murray and Thompson (1980). Fresh needles are frozen under liquid nitrogen and ground to a fine-textured powder which is then dispersed in a sorbitol-based extraction medium. Organelles are collected by sedimentation centrifugation, resuspended, and lysed with sarkosyl to release the DNA. The DNA is purified by chloroform/octanol extraction and banding on CsCl-ethidium bromide isopycnic gradients before ethanol precipitation. Yield is typically 0.5 to 1.0 mg per gram fresh weight of tissue.

Southern Blots

Genomic DNA is digested with each of several restriction endonucleases that were selected on the basis of cost, their capacity to digest genomic DNA completely, and their ability to reveal RFLPs in an initial screening study. The enzymes chosen included EcoRI, EcoRV, HindIII, and BamHI. Total DNA is digested to completion according to the supplier's recommendations. Restriction fragments are separated by horizontal agarose gel electrophoresis and transferred to Hybond-N (Amersham) nylon membranes for hybridization analyses.

DNA Probes

A complementary DNA (cDNA) library was constructed from messenger RNA (mRNA) isolated from two-week-old loblolly pine seedlings. A cDNA library was chosen as the source of probes because of its enrichment for single-copy sequences, as opposed to a genomic library, which is ordinarily high in repetitive sequences. Random cDNA clones have also been shown to reveal as much or more variability than genomic clones (Landry et al. 1987). Double-stranded cDNAs were inserted into a plasmid vector for cloning. Amplified clones are prepared for hybridization by radiolabeling to high specific activity with ^{32}P-dCTP (3000 Ci/mmol) using the random primer technique (Feinberg and Vogelstein 1983).

Southern Blot Hybridization Analysis

Southern blots are hybridized according to established protocols (Maniatis et al. 1982). Blots are prehybridized at 42°C for 3-5 hrs. with a solution containing 50% formamide. The prehybridization solution is replaced with fresh solution containing the radiolabeled probe at a concentration of approximately 5 x 10 dpm/ml. Hybridization is continued at 42°C for 16-24 hrs. Aqueous hybridizations at 65°C can also be used. Blots are washed at high stringency and autoradiographed at -80°C for 1-6 days.

Loblolly Pine Pedigrees

The RFLP map will be based on a single 3-generation pedigree from the North Carolina State University Industry Cooperative Tree Improvement Program. Measurement of linkage requires the joint segregation of two or more heterozygous loci in at least one parent of a cross (Bailey 1961; Mather 1951). Therefore, the more heterozygous loci an individual parent tree has, the more information on linkage it contains. A rapid estimate of individual tree heterozygosity can be obtained from isozyme analysis. Thus, to maximize heterozygosity in the cross selected for RFLP mapping, 155 candidate parent-pairs (second generation selections) were assayed for isozyme heterozygosity at 39 loci. Parent-pairs had from 0 to 14 heterozygous loci; the five most heterozygous pairs were selected for additional screening for RFLPs. In this preliminary screening for selection of a pedigree, it was assumed that isozyme and RFLP variability are correlated. This assumption is supported by comparative estimates obtained for maize and tomato (Helentjaris et al. 1985).

The five selected parent-pairs were screened for RFLPs by using combinations of four enzymes (EcoRI, EcoRV, BamHI, and HindIII) and 40 probes. Fifteen combinations revealed a polymorphism between parent-pairs for at least one cross. The most variable parent-pair possessed 10 of the 15 RFLPs found and has been selected as the parent-pair for the mapping project. This cross also had the most heterozygous loci based on the isozyme data.

We have begun screening the selected cross for RFLPs with the collection of random cDNA probes. The probe x enzyme combinations that identify RFLPs will be assayed on blots containing DNA for 100 or more progeny to test for Mendelian segregation of RFLPs. The four grandparents (first-generation selections) of this pedigree will also be screened to provide information on phase of alleles in the parent trees. Parental screening and progeny analysis will require approximately three years to complete.

Data Analysis

RFLP segregation data will be subjected to genetic linkage analysis similar to that used for human pedigrees. Human and conifer pedigrees are both outbred, unlike many crops, and the phase of alleles is often unknown. A general approach to linkage computation, the method of maximum likelihood, was advanced by Haldane and Smith (1947). Elston and Stewart (1971) formulated a general algorithm for calculating likelihood of a given map. Several computer program packages have been developed to implement this algorithm, including LIPED (Ott 1976) and LINKAGE (Lathrop and Lalouel 1984). The Elston-Stewart algorithm, however, is not well suited to a multilocus analysis of a large number of loci because the computation time needed to calculate likelihoods increases exponentially with the number of loci. Lander and Green (1987) described an alternative algorithm and accompanying computer package, MAPMAKER (Lander et al. 1987; Lincoln and Lander 1987), for multipoint analysis which overcomes the limitations of the Elston-Stewart technique. Hulbert et al. (1988) used MAPMAKER to analyze RFLP loci in

lettuce downy mildew (Bremia lactucae), an outcrossing fungus. Data from the present study will be analyzed with the MAPMAKER package using a microcomputer.

Number of Markers Required for Genome Saturation

The number of RFLPs mapped depends upon the degree of saturation desired and the size of the genome. It has been suggested that if the map is used to identify QTLs, then the maximum distance between markers should be 0.2 to 0.4 Morgans (c = 0.1 to 0.2)(Soller et al. 1976; Soller and Plotkin-Hazen 1977). Lange and Boehnke (1982) developed the expression n = log(1-p)/log(1-2c/k) to estimate number of markers where: n = number of markers required, p = proportion of a circular genome to be mapped, c = desired distance in Morgans between marker and QTL, and k = total length of genome in Morgans. Lange and Boehnke cautioned, however, that chromosome end effects increase this estimate by 20 to 30%.

The number of map units in the genome of loblolly pine (N = 12) has not been accurately determined. From isozyme linkage analyses, however, we estimate that the genome is approximately 2000 cM. Using the expression of Lange and Boehnke and assuming p = 0.9, c = 0.1, and k = 20, the corrected estimate is 275 markers. Since this estimate of genome size is based on limited data, it will be re-estimated once 50 to 100 RFLP markers have been mapped.

APPLICATIONS AND LIMITATIONS OF RFLPS IN TREE IMPROVEMENT

Genetic Markers

An important early application of the genetic map will be the acquisition of a large number of new genetic markers. To date, only a few morphological (Franklin 1970) and isozyme (Adams and Joly 1980a, 1980b) markers have been identified in loblolly pine. RFLP markers could improve the efficiency of clonal identification, parental verification in crosses, seed lot identification, and seed orchard studies (Adams 1981, 1983). More genetic markers will also improve estimation of other genetic parameters, such as recombination rates, genetic diversity, mutation rates, and segregation distortion. RFLP technology is currently expensive and labor intensive; thus isozymes will be the preferred marker where these constraints are present.

Construction of the initial RFLP map for loblolly pine will be very time consuming. However, once a full set (200-300) of RFLP probes has been identified and mapped, they will be potentially very useful as markers in other loblolly pine crosses and possibly in related pines. Mapping other crosses and related species should proceed much more rapidly than the initial map.

Identification of Quantitative Trait Loci

One of the most important, but longer term, applications of this map will be the opportunity to evaluate the potential for detecting linkages between RFLPs and quantitative trait loci (QTLs). Several QTL have recently been mapped in tomato by using either a regression approach (Nienhuis et al. 1987; Osborn et al. 1987; Tanksley and Hewitt 1988; Martin et al. 1989) or a new maximum likelihood method (Paterson et al. 1988; Lander and Botstein 1989). These analyses were conducted in either backcrosses or F_2's.

A theory has not been developed to detect linkages between RFLPs and QTLs in outbred populations in linkage equilibrium such as loblolly pine and other conifers. It may be possible to detect linkages between RFLPs and QTLs in specific crosses in which matings are made between trees of phenotypic extremes for the quantitative trait (Lander and Botstein 1989) and if the heritability of the trait is high and under the control of a small number of genes (Lander and Botstein 1986). An empirical test of the power of RFLP-QTL analysis in conifers should be made once the genetic map is completed. Specific traits and crosses should be chosen to maximize the opportunity of successfully detecting linkages. If this test proves successful, additional experiments could be designed to test for linkages between RFLPs and commercially important traits.

Marker-Assisted Selection

Forest tree breeders have continually searched for ways to select the best genotypes in breeding populations prior to rotation age. In loblolly pine, the approach has been to determine if correlations exist between the value of a trait expressed in a juvenile plant and a mature plant. If a strong juvenile-mature correlation can be established, then it may be possible to select for the trait in very young plants. Juvenile-mature correlations have been shown for traits measured in young plants (4-6 years) (Lambeth 1983; Lambeth et al. 1983; Foster 1986) and also for very young seedlings (<2 years) (Waxier and van Buijtenen 1981; Williams 1987,1988). First-year selection is currently practiced for fusiform rust resistance, but as yet, reliable first-year selection criteria have not been developed for yield or other commercial traits.

A second approach to early selection is marker-assisted indirect selection. This method has not been tried in loblolly pine because of the absence of genetic markers. This could change rapidly with the development of a saturated RFLP linkage map. The first step is to establish linkages between RFLPs and QTLs. Linkages between RFLPs and QTLs for commercially important traits have recently been demonstrated in tomato (Paterson et al. 1988). As noted earlier, however, linkages have been shown in F and backcross populations derived from highly inbred lines. In this situation, the "high" QTL allele is expected to be in cis with one RFLP allele and the "low" QTL allele with the alternative RFLP allele. In conifer populations, RFLP and QTL alleles will essentially be in linkage equilibrium. Several designs have been suggested for outbreeding populations to overcome this problem (Soller and Genizi 1978; Soller and Beckman 1983; Beckman and Soller

1988); however, each has evere limitations that make them impractical to apply to conifers. Currently, we do not consider marker-assisted selection applicable to existing loblolly pine breeding programs. Only if breeding programs were limited to a very small number of elite pedigrees would it become feasible to develop maps for each pedigree and practice indirect selection on RFLPs linked to QTLs for traits of interest.

ACKNOWLEDGMENTS. We thank If fat Ali and Paul Hodgskiss for technical assistance; Claire Kinlaw for help in constructing the cDNA library; Bob Weir, Steve McKeand, and members of the NCSU Coop for providing pedigrees; and Claire Williams and Barbara McCutchan for helpful comments on the manuscript. This work was supported by U.S.D.A. Forest Biology Grant (No. 88-33520-4074) to D.B.N. and Claire Kinlaw.

LITERATURE CITED

Adams, W. T. 1981. Applying isozyme analyses in tree-breeding programs. In. Proceedings of the Symposium on Isozymes of North American Forest Trees and Forest Insects. M. T. Conkle (ed.). Pacific Southwest Forest and Range Experiment Station, Berkeley, CA.

Adams, W. T. 1983. Applications of isozymes in tree breeding. In. Isozymes in Plant Genetics and Breeding, Part A. S. D. Tanksley and T. J. Orton (eds.), Elsevier, Amsterdam. pp. 381-400.

Adams, W. T. and R. J. Joly. 1980a. Genetics of allozyme variants in loblolly pine. J. Hered. 71:33-40.

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

Bailey, N. T. J. 1961. Introduction to the mathematical theory of genetic linkage. Oxford University Press, London. 298p.

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

Beckman, J. S. and M. Soller. 1986. Restriction fragment length polymorphisms and genetic improvement of agricultural species. Euphytica 35:111-124.

Beckman, J. S. and M. Soller. 1988. Detection of linkage between marker loci affecting quantitative traits in crosses between segregating populations. Theor. Appl. Genet. 76:228-236.

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

Feinberg, A. P. and B. Vogelstein. 1983. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. 132:6-13.

Foster, G. S. 1986. Trends in genetic parameters with stand development and their influence on early selection for volume growth in loblolly pine. Forest Sci. 32:944-959.

Franklin, E. C. 1970. Survey of mutant forms and inbreeding depression in species of the family Pinaceae. U.S.D.A. Forest Serv. Res. Pap. SE-61, Southeastern Forest Experiment Station, Ashville, NC 21 p.

Haldane, J. B. S. and C. A. B. Smith. 1947. A new estimate for the linkage between the genes for colour-blindness and haemophilia in man. Ann. Eugen. 14:10-31.

Helentjaris, T., G. King, M. Slocum, C. Siedenstrang, and S. Wegman. 1985. Restriction fragment polymorphism as probes for plant diversity and their development as tools for applied breeding. Plant Molec. Biol. 5:109-118.

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

Lambeth, C. C. 1983. Early testing--an overview with emphasis on loblolly pine. Proc. 17th South. For. Tree Imp. Conf. University of Georgia, Athens, GA p. 297-311.

Lambeth, C. C., J. P. van Buijtenen, and S. D. Duke. 1983. Early selection is effective in 20-year-old genetic tests of loblolly pine. Silvae Genet. 32:210-215.

Lander, E. S. and D. Botstein. 1986. Strategies for studying heterogeneous genetic traits in humans by using a linkage map of restriction fragment length polymorphisms. Proc. Natl. Acad. Sci. USA 83:7353-7357.

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. USA 84:2363-2367.

Lander, E. S., P. Green, J. Abrahamson, A. Barlow, M. Daly, S. Lincoln, 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., R. Kesseli, H. Leung, and R. W. Michelmore. 1987. Comparison of restriction endonucleases and sources of probes for their efficiency in detecting restriction fragment length polymorphisms in lettuce (Lactuca sativa L.). Theor. Appl. Genet. 74:646-653.

Lange, K. and M. Boehnke. 1982. How many polymorphic marker genes will it take to span the human genome? Am. J. Hum. Genet. 34:842-845.

Lathrop, G. M. and J. M. Lalouel. 1984. Easy calculation of lod scores and

genetic risks on small computers. Am. J. Hum. Genet. 36:460-465.

Lincoln, S. and E. S. Lander. 1987. Constructing genetic linkage maps with MAPMAKER: A tutorial and reference manual. Whitehead Institute for Biomedical Research.

Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory. Cold Spring Harbor, New York.

Martin, B., J. Nienhuis, G. King, and A. Schaefer. 1989. Restriction fragment length polymorphisms associated with water use efficiency in tomato. Science 243:1725-1728.

Mather, K. 1951. The Measurement of Linkage in Heredity. 2nd ed. John Wiley and Sons, Inc. New York 346p.

Murray, **M. G.** and W. F. Thompson. 1980. Rapid isolation of high molecular weight plant DNA. Nuc. Acids Res. 8:4321-4325.

Nienhuis, J., T. Helentjaris, M. Slocum, B. Ruggero, and A. Schaefer. 1987. Restriction fragment length polymorphism analysis of loci associated with insect resistance in tomato. Crop. Sci. 27:797-803.

Osborn, T. C., D. C. Alexander, and J. F. Forbes. 1987. Identification of restriction fragment length ploymorphisms linked to genes controlling soluable solids content in tomato fruit. Theor. Appl. Genet. 73:350-356.

Ott, J. 1976. A computer program for linkage analysis of general human pedigrees. Am. J. Hum. Genet. 28:528-529.

Paterson, A. H., E. S. Lander, J. D. Hewitt, S. Peterson, S. E. Lincoln 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.

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

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

Soller, M. and A. Genizi. 1978. The efficiency of experimental designs for the detection of linkage between a marker locus and a locus affecting a quantitative trait in segregating populations. Biometrics 34:47-55.

Soller, M., A. Genizi, and T. Brody. 1976. On the power of experimental designs for the detection of linkage between marker loci and quantitative loci in crosses between inbred lines. Theor. Appl. Genet. 46:35-39.

Soller, M. and J. Plotkin-Hazen. 1977. The use of marker alleles for the

introgression of linked quantitative alleles. Theor. Appl. Genet. 51:133-137.

Southern, E. M. 1975. Detection of specific sequences among DNA fragments separated be electrophoresis. J. Mol. Biol. 98:503-517.

Tanksley, S. D. and J. Hewitt. 1988. Use of molecular markers in breeding for soluable solids content in tomato-- a re-examination. Theor. Appl. Genet. 75:811-823.

Waxier, M. S. and J. P. van Buijtenen. 1981. Early genetic evaluation of loblolly pine. Can. J. For. Res. 11:351-355.

Williams, C. G. 1987. The influence of shoot ontogeny on juvenile-mature correlations in loblolly pine. Forest Sci. 33:411-422.

Willams, C. G. 1988. Accelerated short-term genetic testing for loblolly pine families. Can. J. For. Res. 18:1085-1089.