Theor Appl Genet (1994) 88:273–278

M. E. Devey · T. A. Fiddler · B.-H. Liu · S. J. Knapp
D. B. Neale

An RFLP linkage map for loblolly pine based on a three-generation outbred pedigree

Received: 1 August 1993 / Accepted: 9 September 1993

Abstract A genetic linkage map for loblolly pine (*Pinus taeda* L.) was constructed using segregation data from a three-generation outbred pedigree consisting of four grandparents, two parents, and 95 F_2 progeny. The map was based predominantly on restriction fragment length polymorphism (RFLP) loci detected by cDNA probes. Sixty-five cDNA and three genomic DNA probes revealed 90 RFLP loci. Six polymorphic isozyme loci were also scored. One-fourth (24%) of the cDNA probes detected more than 1 segregating locus, an indication that multigene families are common in pines. As many as six alleles were observed at a single segregating locus among grandparents and it was not unusual for the progeny to segregate for three or four alleles per locus. Multipoint linkage analysis placed 73 RFLP and 2 isozyme loci into 20 linkage groups; the remaining 17 RFLP and 4 isozyme loci were unlinked. The mapped RFLP probes provide a new set of codominant markers for genetic analyses in loblolly pine.

Key words RFLP · Genetic linkage mapping · *Pinus taeda* L.

Communicated by P. M. A. Tigerstedt

M. E. Devey
Division of Forestry/CSIRO, P.O. Box 4008, Queen Victoria Terrace, Canberra, ACT 2600, Australia

T. A. Fiddler
Oregon Health Sciences, School of Medicine, Portland, OR 97201, USA

B.-H. Liu
Department of Forestry, North Carolina State University, Raleigh, NC 27695, USA

S. J. Knapp
Department of Crop and Soil Sciences, Oregon State University, Corvallis, OR 97331, USA

D. B. Neale (✉)
Institute of Forest Genetics, USDA/Forest Service, Pacific Southwest Research Station, P.O. Box 245, Berkeley, CA 94701, USA

Introduction

Loblolly pine (*Pinus taeda* L.) is the most economically important forest tree species in the United States. A genetic linkage map for this species would be useful for an understanding of its genome organization and gene expression; as well, it would provide a source of genetic markers for tree improvement programs and gene resource management. Early attempts to construct such maps using isozymes (Adams and Joly 1980 a,b) were limited by the number of available markers. No classical or cytogenetic maps have been developed for loblolly pine.

Pines are diploid organisms with $2n = 24$ chromosomes. They are characterized by their longevity, outbred mating systems (Dorman 1976), and high genetic variability relative to other plant species (Hamrick et al. 1981). DNA content is also high, varying from 17 to 35 pg per nucleus (Dhillon 1987). When compared to other plant species, the loblolly pine genome was found to be approximately 3.7 times larger than maize and 110 times larger than *Arabidopsis* (Neale and Williams 1991). A large proportion of the pine genome is made up of repetitive sequences (Kriebel 1985; Miksche 1985). In terms of map units, however, the conifer genome is probably not much larger than that of other crop species. It has been estimated that 300 markers would cover the genome of loblolly pine (approximately 2500 cM) at an average distance of 20 cM between markers.

For natural populations in which inbred lines are not available (e.g., humans, trees) crosses between heterozygous individuals must be used to construct multilocus linkage maps. Pine pedigrees consist of two or three generations: four grandparents (optional), two parents, and a large number of progeny. In plants, the computational aspects of this type of analysis are difficult (Ritter et al. 1990). Computation time increases greatly using CEPH (Centre d'Etude du Polymorphisme Humain) pedigrees with large sibships, particularly when certain ambiguities in the data occur, such as when parental

phases are unknown. The usual programs for analysis of this type of data, e.g., MAPMAKER (Lander et al. 1987) or LINKAGE (Lathrop and Lalouel 1984), are very inefficient for the construction of maps beginning with a large number of markers and a large number of progeny. GMENDEL 2.0 (Liu and Knapp 1990) uses multiple two-point locus-ordering algorithms, which can easily handle a large number of markers and progeny. CRI-MAP (Green et al. 1988) can handle a large number of loci but is not as efficient for analysis of data from large sibships.

We have constructed a genetic map for loblolly pine using restriction fragment length polymorphism (RFLP) loci. Most of the probes used for the construction of this map are cDNA sequences, which will provide information on genes and genome organization in loblolly pine. In addition, these probes will cross-hybridize to other species. The set of markers published here is being used to construct maps in *P. radiata*, *P. sylvestris*, *Pseudotsuga menziesii*, and *Picea abies*, thus providing information on comparative genome organization in conifers. Application of these probes for mapping genes controlling wood specific gravity and other quantitative traits in loblolly pine has also been initiated.

We report here the linkage analysis and genetic map developed for loblolly pine based on RFLPs and a small number of isozyme loci. The map consists of 73 RFLP and 2 isozyme loci arranged into 20 linkage groups; an additional 17 RFLP and 4 isozyme loci were unlinked. These groups should consolidate as more markers are added; another 100 markers are probably needed to arrive at the expected number of 12 linkage groups.

Materials and methods

Plant materials

A three-generation outbred pedigree (Fig. 1) consisting of four grandparents, two parents, and 95 progeny was selected from a large number of candidate pedigrees on the basis of RFLP and isozyme heterozygosities in the parent trees (Devey et al. 1991). The plant materials were sampled from several seed orchard and progeny test sites in Arkansas and Oklahoma.

RFLP and isozyme procedures

Procedures relating to DNA extraction, Southern blotting, and probe hybridization have been described previously (Devey et al. 1991). The primary source of mapping probes was a loblolly pine cDNA library prepared from total RNA extracted from 12-day-old seedling roots and shoots. Genomic DNA clones were prepared from *Pst*I-*Bam*HI-digested and size-selected (0.5–2.0 kb) loblolly pine DNA.

A total of 516 cDNA and 100 genomic DNA probes were hybridized to blots of *Hin*dIII-digested grandparent and parent DNAs to identify polymorphic probes that would segregate in the F_2 population. Selected probes were then hybridized to progeny blots to obtain F_2 genotypic data.

Isozyme markers were tested for polymorphism using eight megagametophytes from each parent (Conkle et al. 1982). Sixteen isozymes were evaluated including Glycerate dehydrogenase (GLYDH), 6-phosphogluconate dehydrogenase (6PGD), glucose dehydrogenase (GDH), phosphoglucose isomerase (PGI), Phosphoglucomutase (PGM), Shikimate dehydrogenase (SKD), fluorescent esterase (FEST), aconitase (ACO), fructose-1,6-diphosphatase (FDP), uridine-5-diphosphoglucose pyrophosphorylase (UGPP), diaphorase (DIA), malic enzyme (ME), menadione reductase (MNR), fumarase (FUM), glutamate oxaloacetate transaminase (GOT), and catalase (CAT). The progeny were genotyped for polymorphic loci using protein extracts from needle tissue following methods described by Neale et al. (1984).

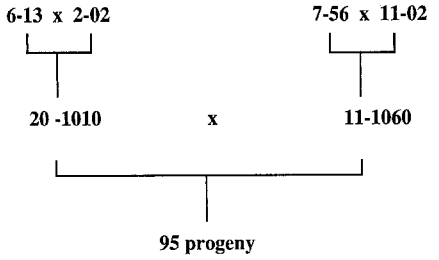

Fig. 1 Three-generation loblolly pine pedigree used for genetic mapping

Linkage analysis

Multipoint linkage analysis was performed on a Sun Microsystems SPARC station 2 using the GMENDEL 2.0 linkage analysis program (Liu and Knapp 1990). This program provides options for standard Mendelian crosses, e.g., backcross and F_2, and for crosses between heterozygous individuals. GMENDEL 2.0 will accept no more than two alleles per locus; thus loci that showed segregations of three or four alleles were coded to two alleles/locus. Some of the progeny classes were combined because up to 16 classes are possible with four alleles per locus. Although we had data from the grandparents, GMENDEL 2.0 does not use this information; the program determines phase from progeny segregation data. Loci were placed into linkage groups on the basis of a recombination value less than 30% and $P < 0.001$ (LOD score > 3.0), and ordering was done using multiple two-point algorithms. The technical aspects of linkage mapping using heterozygous outbred matings are to be discussed in a subsequent paper by S. J. Knapp (personal communication).

A second analysis using CRI-MAP (Green et al. 1988) was also conducted. This program uses all of the information available from the grandparents and does not code data from loci segregating for three or four alleles. Linkage groups were determined from a two-point analysis of segregation data, and order was established on a per linkage group basis using markers that were linked to each other. With this data set, it was not possible to order more than 6 to 8 loci at a time.

Database and nomenclature

The genetic map and associated laboratory information from this study has been submitted for electronic publishing in the Dendrome database. Dendrome is a database of genome mapping information for forest trees. The genetic map and information pertaining to RFLP probes, pedigrees, segregation data, and computerized images of autoradiograms can be found within a database called TreeGenes. Inquiries to Dendrome can be sent by electronic mail to dendrome@s 27w007.pswfs.gov.

The nomenclature we proposed in Devey et al. (1991) for probes and loci has been slightly modified. The *c* or *g* notation for the cDNA or genomic DNA library following the IFG signature has been dropped, and locus identifications have been changed to a, b, c, etc.; e.g., PtIFG602a.

Results

RFLP and isozyme variation

We screened a large number of cDNA and genomic DNA probes against grandparent/parent DNAs and selected the best probes for mapping. A large amount of genetic variation was revealed by probes that detected single- and low-copy number sequences. Many polymorphic probes were not included in the linkage analysis either because the complexity of the pattern was genetically uninterpretable or because the signal was inadequate for efficient scoring. Segregation data were collected on 65 cDNA and 3 genomic DNA probes (Table 1). A much higher percentage of the genomic DNA probes (> 50%) detected dispersed or tandemly repeated sequences.

Of the 65 mapped cDNA probes 16 (24%) revealed more than one segregating and scorable locus. An average of 1.33 interpretable loci were detected for each cDNA probe. Lowercase letters designate different loci detected by the same probe, "a" being the highest molecular weight. The loci detected by 4 of the probes, 669, 1950, 2540, and 2738, were tightly linked, e.g., 669a, b, and c mapped within 9.6 cM of each other. Loci detected by other multilocus probes were unlinked to each other on the basis of the present map.

We often observed what appeared to be a family of several segregating loci detected by the same probe; however, it was not possible to determine which fragments were allelic, and consequently these loci were not mapped. Although of little value for mapping purposes, this class of probes would be useful for DNA fingerprinting applications.

As many as six RFLP alleles per locus were observed among the four grandparents of the cross. In the progeny, 4 loci were observed to segregate for four alleles per locus, 24 for three alleles, and 9 for two alleles (both parents heterozygous). In 53 of the loci, only one of the parents was segregating for the marker. Most of the genomic DNA probes revealed high-copy or monomorphic restriction fragment patterns, and only a small percentage of these (3%) were actually mapped. The use of two restriction enzymes for cloning of genomic DNA evidently increased the number of high-copy sequences in this library. A higher percentage (approximately 75%) of single- or low-copy sequences were obtained using *Pst*I alone in *P. radiata* (Smith and Bell, unpublished data).

Among the 16 isozyme loci evaluated, 6 were polymorphic between the two parents: *Gdh*, *To*, *Pgm1*, *Skd*, *6Pgd*, and *Ugpp*. Data were collected from segregating progeny, and these markers were included in the linkage analysis.

Significant lack of fit ($P < 0.05$) to expected Mendelian ratios was observed for 6 of the loci: 1918e, 2323b, 2575a, 2718a, *To*, and *6Pgd*. These deviations may have resulted from ambiguities in scoring certain alleles or

Table 1 Linkage group, map position (cM) as shown in Treegenes database, and G-statistic for 90 RFL loci and 6 isozyme loci in loblolly pine cross 20-1010 × 11-1060

Locus	Linkage map		G-statistic
	Group	Position	
602a	1	0.0	2.797
605a	1	91.9	1.374
605b	2	119.2	2.187
616b	3	0.0	0.065
624a	4	0.0	0.912
653a	5	48.9	1.378
653b	11	0.0	1.562
653d	6	20.9	0.263
655a	Unlinked		0.011
658b	7	0.0	1.304
658c	6	0.0	4.740
669a	4	28.2	0.516
669b	4	25.4	3.639
669c	4	24.1	0.891
670a	Unlinked		0.000
674a	Unlinked		0.410
739a	11	2.3	1.823
846a	Unlinked		0.200
1154b	1	92.4	0.693
1165a	19	0.0	0.011
1454a	2	86.8	0.043
1457b	7	18.9	0.178
1584a	Unlinked		0.383
1588a	Unlinked		0.099
1599a	5	29.6	2.237
1623a	8	14.5	0.728
1626a	8	0.0	0.974
1626b	18	0.0	0.263
1633a	8	38.0	0.191
1633b	12	0.0	0.851
1633c	4	0.0	1.165
1643a	1	33.4	0.044
1750a	Unlinked		0.011
1849a	2	8.9	0.097
1869c	Unlinked		0.099
1917a	5	39.1	0.782
1918a	20	0.0	2.344
1918d	13	0.0	2.947
1918e	Unlinked		3.996*
1918f	9	0.0	3.531
1934a	1	67.2	5.228
1950a	1	18.0	2.768
1950b	1	19.3	0.044
1956b	12	22.0	0.400
2002a	Unlinked		2.860
2009a	14	0.0	0.409
2022a	2	0.0	0.829
2025a	1	19.3	1.443
2068a	1	50.2	2.925
2150a	10	19.7	1.246
2220a	2	79.6	1.650
2256a	13	15.1	1.449
2256c	3	25.2	0.582
2290a	Unlinked		1.304
2291a	20	17.3	3.545
2295a	Unlinked		5.787
2295b	2	97.3	2.943
2317a	Unlinked		0.095
2323b	4	41.8	6.847*
2345a	Unlinked		0.564
2346a	7	12.2	0.011
2357a	2	4.4	0.633
2369a	Unlinked		0.267
2530a	9	30.3	0.049

Table 1 (*Continued*)

Locus	Linkage map		G-statistic
	Group	Position	
2538b	10	0.0	4.452
2538d	10	27.4	0.011
2540a	2	29.2	0.032
2540c	2	29.2	0.099
2540d	2	25.8	0.244
2540e	2	42.0	0.097
2553a	5	58.4	2.191
2559a	15	0.0	0.043
2564b	10	42.4	1.191
2564c	9	8.6	0.363
2568a	5	0.0	0.506
2568b	4	18.0	2.594
2574c	2	105.3	2.925
2575a	1	66.6	6.066*
2610a	16	0.0	0.022
2615b	17	0.0	0.606
2621a	15	4.7	0.047
2697a	6	31.2	0.080
2703a	18	0.0	0.633
2718a	Unlinked		17.400*
2723a	16	4.4	0.170
2738a	3	7.6	2.857
2738b	3	7.6	1.605
1A7c	17	27.6	0.011
4D4a	14	24.7	1.570
1D11a	Unlinked		2.093
6Pgd	19	16.1	6.272*
Gdh	9	26.4	1.435
Pgm1	Unlinked		0.933
Skd	Unlinked		0.013
To	Unlinked		8.258*
Ugpp	Unlinked		4.918

* Significant lack of fit ($P < 0.05$)

loci rather than from some biological cause or chance alone. Three of the markers with significant distortion, 1918e, 2718a, and *To*, were not linked to any of the markers tested; *P* values of the other three loci were just under 0.05, and these loci were included in the map. Segregation distortion has been previously observed at the *6Pgd* locus in loblolly pine (Adams and Joly 1980a).

The genetic linkage map

Genetic linkage relationships among markers were determined using GMENDEL 2.0. Seventy-three RFLP and 2 isozyme loci were arranged into 20 linkage groups, each with two or more loci per group (Fig. 2). Ten linkage groups consisted of only two loci (Fig. 2). Seventeen of the RFLP loci and 4 isozyme loci were not linked to any other locus (maximum recombination value 0.3, $P < 0.001$). A relatively uniform distribution of markers was observed with linked loci covering 632 cM.

With cDNA probes it was common to find a number of different probes that detected identical hybridization patterns on *Hin*dIII-digested DNAs. These were not included on the map; however, in some instances 2 probes (each with different hybridization patterns) were mapped to the same locus, e.g., loci 1950b and 2025a map to the same position on group 1. Sequence analysis of these probes may determine if they are in fact from the same gene.

Results from the CRI-MAP analysis were generally consistent with that of GMENDEL 2.0. Both programs use the same algorithm to estimate recombination frequencies, and there was essentially no difference in these values for a given locus order. The orders were the same on 7 of the 10 groups that had three or more loci. In 3 of the groups the difference in order involved a switch in position of only one locus. The CRI-MAP analysis for group 4 had essentially the same order as that obtained by GMENDEL except for a switch in the tightly linked 669a, b, and c loci. Locus 2568b was not included in this group, and 1633b was substituted for 1633c, although in a different position. The order and number of loci included in group 2 (the largest linkage group) were substantially different between the two analyses.

Discussion

Segregation of RFLP markers in the progeny of a three-generation pedigree allowed the construction of a genetic linkage map for loblolly pine. Twenty linkage groups were obtained. As more markers are added, the expected number of 12 linkage groups should be obtained. The limited number of loci that were mapped did not always allow a unique order to be established. GMENDEL 2.0 does not provide any information on alternative gene orders; however, another analysis using the program CRI-MAP (data not shown) produced identical linkage groups except for group 2. With any multipoint linkage analysis, order and distance between markers will vary and acquire increased precision as additional markers are added. The same set of markers, along with newly identified RFLP markers, is being used to construct a map in a second three-generation pedigree. Information from this study should help to provide a more precise order.

The ability of GMENDEL 2.0 to handle large sibships derived from matings of heterozygous parents makes it suitable for linkage analysis of pine outbred pedigrees. For many forest tree species, three-generation pedigrees are unavailable; the use of two-generation pedigrees for mapping may then be necessary. Although grandparental data contribute additional information on the phase of marker alleles in the parents, these data are not necessary for this analysis. A limitation of GMENDEL 2.0, however, is that loci segregating for three or four alleles are coded as two alleles/locus; thus, some amount of information is lost. It is not known how serious this problem is. The comparison with CRI-MAP, which uses all available allelic information, indicates that locus order will probably not be affected in most cases.

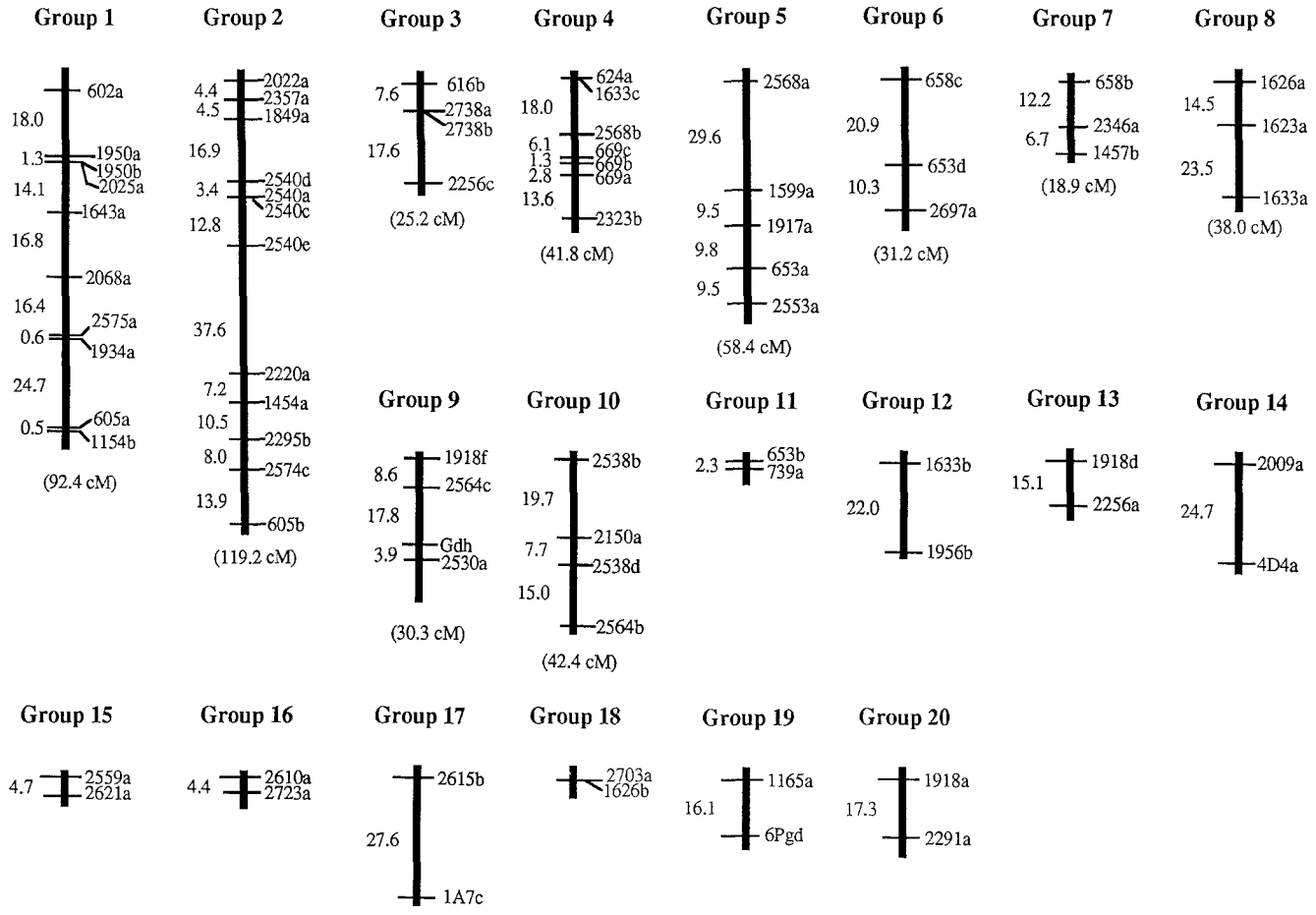

Fig. 2 Loblolly pine genetic map based on 73 RFLP loci and 2 isozyme loci. *Numbers to the right of bold lines* refer to genetic loci; *numbers to the left* refer to distances between markers in centiMorgans (cM)

An interesting result from this study concerns the occurrence of multigene families. The existence of families of related DNA sequences has been well established in other organisms. These sequences appear to be widely dispersed in the conifer genome. Fifteen of the cDNA probes that were mapped in loblolly pine detected more than one segregating locus; many more probes detected multiple loci but were not mapped because of the complexity of their hybridization patterns. These results suggest that conifer genome size is in part a result of large multigene families. Kinlaw et al. (1990) observed a complex multibanded pattern when probing *P. taeda* and *P. radiata* with alcohol dehydrogenase (ADH) cDNA clones, in contrast to relatively simple patterns characteristic of other plant species, e.g., maize (Dennis et al. 1985). The additional closely homologous sequences may be pseudogenes that have been altered to prevent transcription of functional messenger RNAs or translation of functional proteins. Analysis of the gene families identified by these RFLP probes has been initiated and should provide interesting insights into the evolution of the conifer genome.

Acknowledgements The authors thank the North Carolina State University – Industry Cooperative Tree Improvement Program and C. Williams and L. Burris of Weyerhaeuser Company for providing the germ plasm used in this study. We also thank B. Sherman for computer assistance, K. Marshall for graphics, and G. Moran and M. Roose for their reviews of the manuscript. This study was funded by USDA Forest Biology Grant No. 88-33520-4074.

References

Adams WT, Joly RJ (1980a) Genetics of allozyme variants in loblolly pine. J Hered 71:33–40

Adams WT, Joly RJ (1980b) Linkage relationships among twelve allozyme loci in loblolly pine. J Hered 71:199–202

Conkle MT, Hodgskiss PD, Nunnally LB, Hunter SC (1982) Starch gel electrophoresis of conifer seeds: a laboratory manual. USDA For Ser Gen Tech Rep PSW-64, pp 1–18

Dennis ES, Sachs MM, Gerlach WL, Finnegan EJ, Peacock WJ (1985) Molecular analysis of the alcohol dehydrogenase 2 (*Adh2*) gene of *Z. mays*. Nucleic Acids Res 13:727–743

Devey ME, Jermstad KD, Tauer CG, Neale DB (1991) Inheritance of RFLP loci in a loblolly pine three-generation pedigree. Theor Appl Genet 83:238–242

Dhillon SS (1987) DNA in tree species. In: Bonga JM, Durzan DJ (eds) Cell and tissue culture in forestry. Vol 1: general principles and biotechnology. Marinus Nijhoff Publ, Dordrecht, The Netherlands, pp 298–313

Dorman KW (1976) Genetics and breeding of southern pines. Agric Handb 471. US Dept Agriculture, Washington, D.C.

Green P, Falls K, Crooks S (1988) Documentation of CRI-MAP. Collaborative Research

Hamrick JL, Mitton JB, Linhart YB (1981) Levels of genetic variation in trees: the influence of life history characteristics. In: Conkle MT (ed) Proc Symp Isozymes North Am For Trees For Insects. Gen Tech Rep PSW-48. Pacific Southwest Forest and Range Experiment Station, Forest Service, US Dept Agriculture, Berkeley, Calif., pp 35–41

Kinlaw CS, Harry DH, Sederoff RR (1990) Isolation and characterization of alcohol dehydrogenase cDNAs from *Pinus radiata*. Can J For Res 20:1343–1350

Kriebel HB (1985) DNA sequence components of the *Pinus strobus* nuclear genome. Can J For Res 15:1–4

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

Lathrop GM, Lalouel JM (1984) Easy calculation of lod scores and genetic risks on small computers. Am J Hum Genet 36: 460–465

Liu BH, Knapp SJ (1990) GMENDEL: a program for Mendelian segregation and linkage analysis of individual or multiple progeny populations using log-likelihood ratios. J Hered 8:407

Miksche JP (1985) Recent advances of biotechnology and forest trees. For Chron 61:449–453

Neale DB, Williams CG (1991) Restriction fragment length polymorphism mapping in conifers and applications to forest genetics and tree improvement. Can J For Res 21:545–554

Neale DB, Weber JC, Adams WT (1984) Inheritance of needle tissue isozymes in Douglas-fir. Can J Genet Cytol 26:459–468

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