

# A Consensus Map for Loblolly Pine (*Pinus taeda* L.). I. Construction and Integration of Individual Linkage Maps From Two Outbred Three-Generation Pedigrees

Mitchell M. Sewell<sup>\*,1</sup> Bradley K. Sherman<sup>\*,2</sup> and David B. Neale<sup>\*,†</sup>

<sup>\*</sup>*Institute of Forest Genetics, USDA Forest Service, Pacific Southwest Research Station, Placerville, California 95667 and*

<sup>†</sup>*University of California, Department of Environmental Horticulture, Davis, California 95616*

Manuscript received July 22, 1998

Accepted for publication September 28, 1998

## ABSTRACT

A consensus map for loblolly pine (*Pinus taeda* L.) was constructed from the integration of linkage data from two unrelated three-generation outbred pedigrees. The progeny segregation data from restriction fragment length polymorphism, random amplified polymorphic DNA, and isozyme genetic markers from each pedigree were recoded to reflect the two independent populations of parental meioses, and genetic maps were constructed to represent each parent. The rate of meiotic recombination was significantly greater for males than females, as was the average estimate of genome length for males [1983.7 cM [Kosambi mapping function (K)]] and females [1339.5 cM(K)]. The integration of individual maps allows for the synthesis of genetic information from independent sources onto a single consensus map and facilitates the consolidation of linkage groups to represent the chromosomes ( $n = 12$ ) of loblolly pine. The resulting consensus map consists of 357 unique molecular markers and covers  $\sim 1300$  cM(K).

A genetic linkage map is a fundamental organizational tool for genomic research. For forest trees, the most important applications of genetic maps are toward (1) a basic knowledge of genome organization and evolution (KINLAW and NEALE 1997), (2) the localization of monogenic traits (*e.g.*, DEVEY *et al.* 1995) and oligogenic traits (*e.g.*, GROOVER *et al.* 1994), and (3) studies of genetic diversity (MITTON 1994). Therefore, for any given species, individual genetic maps are often constructed with a specific goal in mind, thereby generating multiple maps for a single species that feature novel markers and genetic information. The information contained within these individual maps can be further enhanced when these maps are synthesized into a single consensus map to represent a given species.

Consensus maps have been constructed for a number of plant species (*e.g.*, *Arabidopsis thaliana*, HAUGE *et al.* 1993; *Brassica oleracea*, KIANIAN and QUIROS 1992; *Helianthus annuus*, GENTZBITTEL *et al.* 1995; *Hordeum vulgare*, QI *et al.* 1996; *Zea mays*, BEAVIS and GRANT 1991). Mapping with multiple populations provides several advantages over mapping based on a single population. In particular, a larger number of loci can be placed onto a single map. This is especially important when attempting to map specific genes of interest (*e.g.*, morpho-

logical markers or candidate genes for trait mapping) that are unlikely to segregate within a single mapping population. This also provides for greater genomic coverage. These multipopulation mapping studies have provided evidence for chromosomal rearrangements (BEAVIS and GRANT 1991; KIANIAN and QUIROS 1992) and gene duplication (KIANIAN and QUIROS 1992; GENTZBITTEL *et al.* 1995), have assisted in the assignment of linkage groups to chromosomes (BEAVIS and GRANT 1991), and have provided the basis for comparative studies among related species and subspecies (KIANIAN and QUIROS 1992; HAUGE *et al.* 1993; GENTZBITTEL *et al.* 1995).

Loblolly pine (*Pinus taeda* L.) is a member of the genus *Pinus*, which encompasses  $\sim 100$  species and is among the world's most widely distributed and economically important group of trees. Pines are characterized by their longevity, outbred mating system, and high genetic variability relative to other plant species (HAMRICK and GODT 1990). Because of its wide range, rapid growth, and versatility as a source for wood products, loblolly pine is the principal commercial pine species in North America. All members of the genus *Pinus* have a haploid chromosome number of 12. Compared to other plant species, pine genomes are large (*e.g.*, the estimated *C*-value for loblolly pine is 21–23 pg; WAKAMIYA *et al.* 1993) and are composed of complex gene families (KINLAW and NEALE 1997). Nevertheless, initial investigations suggest that gene order has been conserved among pines (DEVEY *et al.* 1996b), and cytogenetic studies conclude that *Pinus* has been cytologically very stable over evolutionary time (PEDERICK 1970).

For the majority of plant species of academic and

Corresponding author: David B. Neale, Institute of Forest Genetics, USDA Forest Service, 1 Shields Ave., Department of Environmental Horticulture, University of California, Davis, CA 95616.  
E-mail: dneale@dendrome.ucdavis.edu

<sup>1</sup> Present address: USDA Forest Service, Southern Research Station, P.O. Box 2680, Asheville, NC 28802.

<sup>2</sup> Present address: Mendel Biotechnology, Inc., 21375 Cabot Blvd., Hayward, CA 94545.

agronomic interest, pedigrees involving backcrosses (BCs) or  $F_2$ s from inbred lines are constructed for mapping purposes. This pedigree structure simplifies the mapping procedures because the genetic segregation observed in the mapping population is the result of the meiotic recombination from a single genotype (the  $F_1$ ). Therefore, only two alleles segregate in these mapping populations, and the resulting genetic map represents the recombination of the  $F_1$  parent. Furthermore, because two homozygous inbred lines are used to generate the  $F_1$  parent, linkage phase among the alleles of the mapping population is known *a priori*.

In contrast, high genetic load typically prevents the construction of inbred lines in coniferous forest tree species (FRANKLIN 1970). Consequently, both parents of an outbred pedigree are typically highly heterozygous and can possess different pairs of alleles at each locus (*i.e.*, as many as four alleles can segregate for any given marker). Therefore, the genetic segregation observed in such mapping populations is the result of meiotic recombination from both parents, and any given marker can segregate in two (1:1), three (1:2:1), or four (1:1:1:1) genotypic classes within a single mapping population. In addition, phase relationships among alleles are not known *a priori*, but instead must be determined either from the inheritance of alleles within a three-generation pedigree structure or from progeny segregation data.

With a cross between two highly heterozygous parents, the progeny data can be subdivided into two independent data sets that separately contain the meiotic segregation data from each parent. Independent linkage maps can then be constructed for each parent. This "single-parent" approach is an extension of the strategy used to construct maternal linkage maps from random amplified polymorphic DNA (RAPD) markers and haploid megagametophytes in conifers (TULSIERAM *et al.* 1992). Furthermore, the codominant and multiallelic nature of restriction fragment length polymorphism (RFLP) markers provides orthologous markers for aligning the linkage information from each parental data set to produce a sex-average map (DEVNEY *et al.* 1994).

As an initial step toward synthesizing the genetic information available for loblolly pine, the present study integrates the linkage data from two independent outbred pedigrees into a single consensus map. This consensus map serves as a reference genetic map for loblolly pine and as a foundation for studies of genome organization and evolution in conifers.

## MATERIALS AND METHODS

### Data collection

**Mapping populations:** Two three-generation outbred pedigrees of loblolly pine were constructed from unrelated first-

generation selections of the North Carolina State University-Industry Cooperative Tree Improvement Program. These pedigrees were constructed and maintained by the Weyerhaeuser Company. The first pedigree (referred to as the *base* pedigree) consists of 95 full-sib progeny and was previously used to construct a genetic map from 90 RFLP and 6 isozyme loci (DEVNEY *et al.* 1994). The second pedigree (referred to as the *qtl* pedigree) consists of 172 full-sib progeny. GROOVER *et al.* (1994) previously used a selective genotyping strategy to construct the initial genetic map for the *qtl* pedigree. Therefore, a subset of 48 extreme-phenotype progeny was used as the preliminary mapping population. Genotypic data for the 124 intermediate-phenotype progeny were subsequently obtained for selected markers throughout the genome. A total of 155 RFLP loci were used for this previous map construction (GROOVER *et al.* 1994).

**Genetic markers:** Three sources of genetic markers (RFLPs, RAPDs, and isozymes) were used for each mapping population. Methods pertaining to RFLP analysis for loblolly pine were described in DEVNEY *et al.* (1991). The primary source of RFLP probes was a loblolly pine complementary DNA (cDNA) library constructed by DEVNEY *et al.* (1991). Probes from a genomic DNA library were also used (DEVNEY *et al.* 1991). In addition, various colleagues have contributed a number of gene probes characterized from loblolly pine and *P. sylvestris*. The RAPD analysis followed methods described in NELSON *et al.* (1993). Only RAPD markers in a BC configuration for either parent were used in linkage analyses. Methods for protein extraction from vegetative bud tissue for isozyme analyses were described in NEALE *et al.* (1984). Gel electrophoresis and protein staining followed the method of CONKLE *et al.* (1982). The protein stain for cinnamyl alcohol dehydrogenase followed the method of O'MALLEY *et al.* (1992).

**Data organization:** The original *base* and *qtl* maps were constructed independently, and little effort was made to assign common names to redundant probes or to identify orthologous markers. The positive identification of orthologous RFLP markers is not straightforward in conifers because of the presence of multiple alleles and multigene families, yet it is fundamental to the effective utilization and synthesis of genetic data from independent sources. Therefore, to organize and integrate these independent data sets, the raw segregation data were completely rescored, and band migration distances were visually compared as a means to identify orthologous markers and alleles within and among each pedigree data set. This identification procedure was twofold: (1) to identify sets of redundant probes (the final data set includes only a single representative from each set of redundant probes); and then (2) among members of a multigene family (revealed either by nonredundant or redundant probes), to identify putative orthologous markers among individual mapping populations. Once these data sets were organized and the orthologous markers were identified, the probes that were analyzed in only one pedigree were applied to the other pedigree as a means to identify additional orthologous markers.

### Segregation and linkage analyses

**Genotypic classification and coding:** Raw genotypic data were initially recorded with no previous knowledge of phase relationships or inheritance. By using the three-generation pedigree structure, the raw data were converted into an "ancestry-known" data format that resembles that of an inbred pedigree. First, the alleles for each progeny and parent genotype were reoriented to reflect their inheritance from the previous generation. Second, the data were divided into parental lines, and two gametic segregation data sets were created to represent the populations of segregating meioses from each parent.

In situations where the ancestry could not be determined (*i.e.*, when a pair of grandparents were heterozygous for the same pair of alleles or when grandparental data were missing), the data remained in an "ancestry-unknown" format.

Data sets were then coded for linkage analyses with MAPMAKER (Macintosh version 2.0; LANDER *et al.* 1987), MAPMANAGER (Macintosh version 2.6.1; MANLY 1993), and JOINMAP (UNIX version 1.4; STAM 1993). When marker data from an outbred pedigree are oriented to reflect ancestry for each parental line, the coding designations (*i.e.*, A and H) for progeny segregation data represent alleles that were inherited from the grandmother or grandfather, respectively, rather than "homozygote" or "heterozygote." Therefore, the ancestry-known markers can be analyzed within a phase-known model. To accommodate the remaining ancestry-unknown data within a phase-known model, these markers were "reciprocally" coded to arbitrarily represent both possible ancestry options (NELSON *et al.* 1993). During the linkage analyses, the correct phase of the ancestry-unknown markers was established relative to the ancestry-known markers. The original data set was then updated to include the correct phase coding for the ancestry-unknown markers, producing a complete data set that is both ancestry and phase known.

Because JOINMAP manages phase-unknown data internally, the reciprocal coding step described above is not necessary. The disadvantage of this internal manipulation is that the phase orientation that is eventually used is not conveniently made known to the user. In addition, because MAPMAKER and MAPMANAGER cannot simultaneously analyze markers in both a BC and intercross (IC) configuration without a loss of data points, the markers in an IC configuration were removed from each data set for MAPMAKER and MAPMANAGER analyses.

**Concurrent linkage analyses of multiple populations:** Chi-square values were calculated for individual markers to detect deviation ( $P \leq 0.05$ ) of gametic segregation from Mendelian ratios (1:1). Linkage analyses for each mapping population were initially performed using MAPMAKER. Linked markers were first placed into linkage groups using the "group" command with  $\text{LOD} \geq 5.0$  and recombination fraction ( $\theta$ )  $\leq 0.30$ . A preliminary estimation of marker order was obtained using the "first order" command. These preliminary linkage groups from each mapping population were then visually aligned via orthologous markers and coalesced into sets of linkage group homologues. Linkage data within individual mapping populations were consequently joined into individual linkage groups by  $\text{LOD} \geq 5.0$  and  $\theta \leq 0.30$  and/or orthologous markers. Additional markers were then placed onto these linkage groups at  $\text{LOD} \geq 3.0$  and  $\theta \leq 0.40$ . This strategy ensures confidence in the formation of each set of linkage group homologues, yet it allows the placement of a maximum number of markers as well as the proper orientation of linkage fragments. The probable order of all markers for each linkage group was determined using the "compare" command and verified using the "ripple" command. The marker segregation data for these orders were visually inspected using the program MAPMANAGER. The IC markers were then placed relative to the established order of BC markers using JOINMAP and the "fixed sequence" command.

**Integration of linkage data:** An integrated linkage analysis was performed using all segregation data (*i.e.*, both BC and IC markers) from the four mapping populations using JOINMAP. Pairwise estimates of  $\theta$  and corresponding LOD scores were calculated from segregation data for each population. When orthologous markers defined an interval for two or more populations, JOINMAP replaced the individual values of  $\theta$  with a weighted average value (STAM 1993). Using the fixed sequence command, these data were integrated relative to the estab-

lished order of BC markers ascertained from the individual MAPMAKER analyses. Map units in centimorgans were derived from the Kosambi (K) mapping function and were used to construct all linkage maps from both MAPMAKER (results not shown) and JOINMAP analyses. Map figures were constructed using the computer program Linkage Group Visualization Tool (P. SKAGGS, unpublished results).

**Comparison of meiotic recombination rate among parents:** Orthologous marker pairs were chosen for sequential intervals spaced evenly along linkage groups. A minimum of 40 progeny were used in pairwise analyses. A test for a global difference in maternal and paternal recombination frequency among parental data sets was performed from averages of marker pair intervals from each pedigree using a  $z$ -test with the delta method to estimate the associated standard errors (BISHOP *et al.* 1975).

**Estimation of genome length and map coverage:** ECHT and NELSON (1997) identified the need for adopting uniform criteria when estimating genome length. Following these criteria, each data set was reduced to a progeny size of 48 to avoid missing data points but to still include a random sample of all markers. Estimations of genome length,  $E(G)$  (HULBERT *et al.* 1988), and expected genome coverage,  $E(C_e)\%$  (BISHOP *et al.* 1983), were calculated from pairwise segregation data for marker pairs above threshold LODs of 3, 4, and 5. Confidence intervals for genome length were calculated following the method of GERBER and RODOLPHE (1994). The observed genome coverage,  $C_{\text{obs}}$ , was calculated from the results of map construction using a 30-cM scale (NELSON *et al.* 1994). These procedures are detailed in ECHT and NELSON (1997).

**Nomenclature and database:** Locus nomenclature followed that outlined on the Genome Resources page in Dendrome, the forest tree genome informatics project website (<http://dendrome.ucdavis.edu/>). A modification of the alphanumeric coding was used for the locus identifier field for RFLP markers. A lowercase letter was used for RFLP markers in the *base* pedigree, whereas a number was used for RFLP markers in the *qtl* pedigree. For RFLP markers found in both pedigrees, an uppercase letter was used to indicate commonality and facilitate map integration. An additional prefix may be added to a locus or mapping population name, *e.g.*, LG2/2986.A or LG4/*qtl*-pat, respectively, to indicate a specific linkage group.

The constituent and *consensus* maps (SEWELL and NEALE 1998) and all information pertaining to map construction are located in the *Dendrome* and/or *TreeGenes* databases (<http://dendrome.ucdavis.edu/Treegenes/>). The cDNA clones used for RFLP analyses are freely distributed and may be obtained by submitting a request to dneale@dendrome.ucdavis.edu via electronic mail.

## RESULTS

**Genetic markers and single-locus segregation analysis:** From 168 RFLP probes, 42 RAPD primers, and 12 isozyme stains, a total of 581 genetic markers that segregated among the two independent outbred pedigrees of loblolly pine were identified. These marker data were subdivided into four data sets: *base*-maternal (*base*-mat), *base*-paternal (*base*-pat), *qtl*-maternal (*qtl*-mat), and *qtl*-paternal (*qtl*-pat). Of these 581 markers (430 cDNA RFLPs, 36 genomic RFLPs, 87 RAPDs, and 28 isozymes), a total of 418 unique markers (300 cDNA RFLPs, 23 genomic RFLPs, 80 RAPDs, and 15 isozymes) were identified. The remaining 163 markers were orthologous complements found in two or more of the constituent

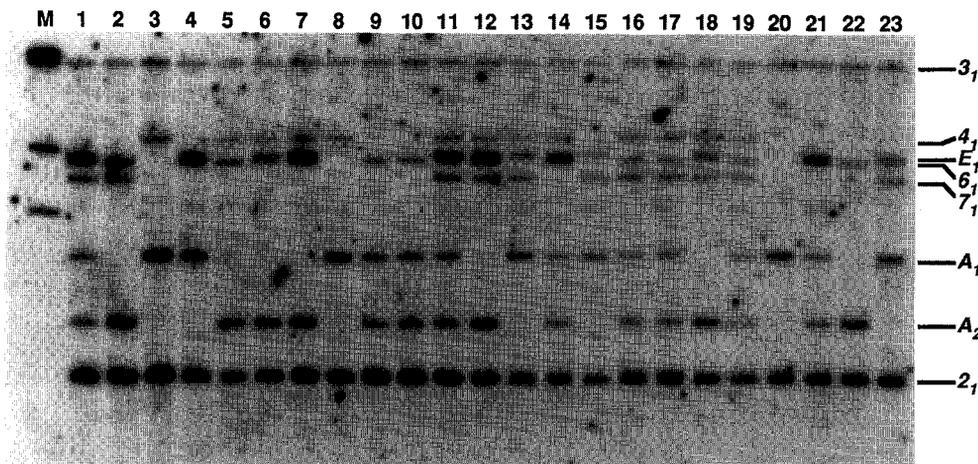


FIGURE 1.—Autoradiogram from a Southern blot analysis of the *qtl* pedigree. Genomic DNAs from 23 progeny were hybridized to cDNA probe PtIFG\_2899. RFLP fragments 4<sub>1</sub>, E<sub>1</sub>, 6<sub>1</sub>, and 7<sub>1</sub> were scored as *present vs. absent*, and they are treated as dominant markers. Loci 2899\_A, 2899\_4, 2899\_E, and 2899\_7 are dispersed throughout the genome (LG12, LG1, LG3, and LG5, respectively; see Figure 2). Locus 2899\_6 is in an IC configuration and was not used in linkage analyses. 2899\_2 and 2899\_3 are putative loci that did not segregate in either the *base* or *qtl* pedigrees.

data sets. On average among these two pedigrees, each marker type yielded 1.88 unique segregating loci (1.89 per cDNA probe, 2.55 per genomic probe, 1.90 per RAPD primer, and 1.25 per isozyme stain). The maximum number of scorable segregating loci per marker type was eight (cDNA probe), six (genomic probe), three (RAPD primer), and two (isozyme stain).

Allelic associations among RFLP fragments could not be identified for some loci (Figure 1). Consequently, these loci were scored as presence *vs.* absence for allele fragments. This phenomenon has been observed in other plant species and in other conifers (*e.g.*, *Arabidopsis thaliana*, LIU *et al.* 1996; *B. oleracea*, KIANIAN and QUIROS 1992; *H. annuus*, GENTZBITTEL *et al.* 1995; *P. radiata*, DEVEY *et al.* 1996a; *Pseudotsuga menziesii*, JERMSTAD *et al.* 1994). Several factors could account for this. The “null” allele might go undetected as a result of comigration with another fragment or limitations of the technique of analysis. Alternatively, the locus might possess a null allele and/or be regarded as hemizygous.

Eight markers (1.4%, two from each mapping population) exhibited significant distortion ( $P \leq 0.05$ ) from expected Mendelian ratios based on chi-square tests

(results not shown). This observed single-locus segregation distortion does not appear to result from selection for or against the alleles of a particular grandparent and is likely due to chance alone.

**Construction of individual linkage maps and multilocus segregation analyses:** The four data sets—*base-mat*, *base-pat*, *qtl-mat*, and *qtl-pat*—contained 109, 106, 191, and 175 genetic markers, respectively, of which 97 (89.0%), 79 (74.5%), 179 (93.7%), and 153 (87.4%) were linked at  $\text{LOD} \geq 3.0$  (Table 1). These 508 (87.5%) mapped markers consisted of 415 RFLPs, 73 RAPDs, and 20 isozymes. Among these markers, 108 were common to two or more populations (Table 2; SEWELL and NEALE 1998) and, thus, were used to identify and align linkage group homologues among mapping populations.

The maps generated by JOINMAP, which includes all markers from both BC and IC configurations, cover 699, 591, 908, and 956 cM (K) for each individual mapping population (*base-mat*, *base-pat*, *qtl-mat*, and *qtl-pat*, respectively; Table 1). These maps (see SEWELL and NEALE 1998) were constructed from the same marker order as defined for the BC markers by the MAPMAKER analyses

TABLE 1  
Number of genetic markers used for map construction

Mapping population	Total markers	Mapped markers <sup>a</sup>	Map coverage <sup>b</sup>	
			JOINMAP	MAPMAKER
<i>base-mat</i>	109	97 (89.0)	699.2	769.2
<i>base-pat</i>	106	79 (74.5)	591.3	603.4
<i>qtl-mat</i>	191	179 (93.7)	907.8	1017.3
<i>qtl-pat</i>	175	153 (87.4)	956.1	1026.6
Population total	581	508 (87.4)	—	—
Consensus	—	357	1359.4	—

<sup>a</sup> Percentage of mapped markers in parentheses.

<sup>b</sup> Map units in centimorgans using Kosambi mapping function [cM(K)].

**TABLE 2**  
Number of orthologous markers  
used for map integration

Mapping population	Orthologous markers		Total <sup>a</sup>
	Intrapedigree	Interpedigree	
<i>base</i> -mat	35	38	55
<i>base</i> -pat	35	44	57
<i>qtl</i> -mat	53	51	81
<i>qtl</i> -pat	53	35	65
Consensus	—	—	108

<sup>a</sup> Total number of orthologous markers represented in intra- and/or interpedigree categories.

(maps not shown). The maps constructed from JOINMAP are slightly shorter than those from MAPMAKER (Table 1) even though the JOINMAP analyses included the additional IC markers. This difference was also observed in barley and is attributed to how each program calculates map distance when the actual interference differs from that assumed (QI *et al.* 1996).

Orthologous markers among constituent linkage group homologues were compared for collinearity of marker order. With minor exceptions, complete collinearity was observed among the orthologous markers of these linkage groups. The minor exceptions included LG5a/*base*-pat, for which markers mapped to the same general area as those from other linkage group homologues, although the marker order is slightly different (SEWELL and NEALE 1998). This region contained a relatively high frequency of double crossovers that made conclusive ordering of markers difficult. Also, several IC markers, which are less precisely ordered because of the presence of ambiguous heterozygous progeny classes, appear inverted among intrapedigree linkage

**TABLE 3**  
Global test for deviation between maternal  
and paternal recombination rates ( $\theta$ )

Pedigree	No. of intervals	Mean difference in $\theta$ (%)	SE (%)	$z^a$
<i>Base</i>	22	26.12	5.61	4.653*
<i>QTL</i>	29	26.33	6.46	4.077*
<i>Combined</i>	51	26.23	4.30	6.105*

<sup>a</sup>  $z$  test with the delta method (BISHOP *et al.* 1975).

\*  $P \leq 0.0001$ .

groups (*e.g.*, LG2/*base*-2885\_B and LG9/*qtl*-Got-4). These comparisons among individual maps from within a species are consistent with those among a self and  $F_1$  map from an individual tree (PLOMION *et al.* 1995), suggesting that the reported discrepancies are caused by minor statistical errors and are not biological in nature. Therefore, we found no significant evidence of minor intraspecific chromosomal rearrangement among the four loblolly pine parent trees.

The availability of orthologous markers also allowed for a direct comparison of the rate of meiotic recombination among the four mapping populations. A comparison of mean recombination frequencies among intervals from throughout the genome revealed a significantly greater male than female meiotic recombination rate in both the *base* (26.12%) and *qtl* (26.33%) pedigrees (Table 3). This is consistent with that reported by GROOVER *et al.* (1995) from a smaller subset of marker pairs for these same pedigrees.

Estimates of genome length were calculated from pairwise segregation data for each mapping population (Table 4). The average estimates of genome length in cM(K) and their 95% confidence intervals (in parenthe-

**TABLE 4**  
Estimated genome length and coverage

Mapping population	No. of loci	$E(G)^a$	$E(C_n)\%^b$	$C_{obs}^c$	$C_{obs}\%^d$
<i>base</i> -mat	88	1315.5 (K)	79.80	918.1	69.79
		1579.5 (H)		967.3	73.53
<i>base</i> -pat	79	2129.1 (K)	55.40	1178.3	55.34
		2554.7 (H)		1152.9	54.15
<i>qtl</i> -mat	157	1363.4 (K)	91.48	1101.0	80.76
		1634.0 (H)		1213.5	89.01
<i>qtl</i> -pat	147	1838.2 (K)	84.12	1541.9	83.88
		2225.6 (H)		1534.2	83.46

<sup>a</sup> Average estimated genome length in centimorgans using Kosambi (K) and Haldane (H) mapping functions.

<sup>b</sup> Percentage of expected genome coverage.

<sup>c</sup> Observed genome coverage [cM(K)] calculated from linkage maps constructed with JOINMAP (above) and MAPMAKER (below).

<sup>d</sup> Percentage of observed genome coverage from JOINMAP (above) and MAPMAKER (below).





ses) at LOD threshold 4 were as follows: 1315.5 (1117.9–1612.7), *base-mat*; 2129.1 (1696.4–2845.7), *base-pat*; 1363.4 (1241.1–1526.6), *qtl-mat*; and 1838.2 (1608.3–2075.3), *qtl-pat*. The average estimates for each maternal genome length were nearly identical, and those for the paternal genomes were not significantly different. However, the paternal estimates were significantly greater than the maternal estimates within each pedigree (38.2% difference within *base* and 25.8% within *qtl*) and also among pedigrees for *qtl-mat* and *base-pat* (and nearly so between *base-mat* and *qtl-pat*).

**Construction of integrated consensus map:** The linkage data from these four independent mapping populations contained 108 orthologous markers that were common to at least two mapping populations. The *base-mat*, *base-pat*, *qtl-mat*, and *qtl-pat* data sets contained 55, 57, 81, and 65 orthologous markers, respectively (Table 2). The resulting consensus map (Figure 2) contains 357 (278 RFLPs, 67 RAPDs, and 12 isozymes) unique genetic markers (Table 1). *Consensus* linkage groups 1–12 were integrated from each of the four independent mapping populations and cover 1175 cM(K) of the genome. *Consensus* groups 13 and 14 were integrated only within the *qtl* pedigree and account for an additional 52 cM(K). The remaining six small “single-population” linkage groups cover 133 cM(K). These 20 linkage groups potentially provide a total genome coverage of 1359 cM(K). Without the presence of orthologous markers on these single-population linkage groups, however, it is undeterminable whether this 133 cM(K) is already covered by the integrated regions of the *consensus* map. Thus, a conservative estimate of the total genome coverage for the *consensus* map from unique linkage segments is 1227 cM(K). With a few minor exceptions (*e.g.*, LG7/1636.1; SEWELL and NEALE 1998), the marker order on the *consensus* map was not different from that defined by each constituent map.

## DISCUSSION

The primary goal for the construction of this consensus map was to place, relative to one another, as many genetic markers as possible onto a single map. Therefore, the concern is more towards obtaining a general order and distance among these markers rather than the fine resolution of order and distance. Furthermore, the *consensus* map provides a means to consolidate linkage groups from the *base* and *qtl* pedigrees (Figure 2; SEWELL and NEALE 1998). With 357 unique genetic markers from cDNA and genomic probes, RAPDs, and isozyme markers, this map is an excellent resource from which markers may be selected for future mapping projects within loblolly pine and for comparative studies among other pines (*e.g.*, DEVEY *et al.* 1996b). For example, combining information from multiple pedigrees is necessary if important traits do not segregate within a single population. Often it is not practical or necessary

to construct complete genetic maps to identify the genomic locations of these traits (*e.g.*, DEVEY *et al.* 1995). However, such studies can still be related to the entire genome by selecting markers suitable for superimposing the detailed region onto the *consensus* map. In a similar manner, the *consensus* map can also serve as a reference for considering the pines as a single genetic system, as similarly proposed for the grasses (BENNETZEN and FREELING 1993), in which genetic tools and information from one pine species immediately become available for studying other pines.

A significant difference between female and male meiotic recombination frequencies was detected among map intervals defined by linked orthologous markers (Table 3). Although reciprocal crosses were not investigated, these results support those from *P. radiata* for which reciprocal crosses were analyzed (MORAN *et al.* 1983). The higher rate of male recombination was observed throughout the genome and resulted in the fragmentation of paternal linkage groups (*e.g.*, LG1/*base-pat* and LG1/*qtl-pat*; SEWELL and NEALE 1998). This difference also resulted in a lower percentage of mapped markers within the paternal mapping populations relative to the maternal populations, where a similar number of markers was used (Table 1). BEAVIS and GRANT (1991) raised the question of whether data should be pooled for construction of a composite map if unequal recombination is detected among constituent populations. However, they pointed out that such a composite map is still a valuable reference tool for selecting markers and comparing the relative location of traits in different genetic backgrounds.

This difference among female and male recombination frequencies directly influenced the estimates of genome length calculated for loblolly pine because these estimates are, in part, based on the number of linked marker pairs relative to the total number of marker pairs within a mapping population. The estimated genome length for loblolly pine was 1315 and 1363 cM(K) when using maternal segregation data and 1838 and 2129 cM(K) when using paternal data (Table 4). ECHT and NELSON (1997) used the same criteria to estimate the genome length of *P. palustris*, *P. pinaster*, and *P. strobus* from maternal segregation data using RAPD markers. The reported genome length for each of these species, which represent both subgenera (*Pinus* and *Strobus*) of the genus *Pinus*, is ~2000 cM(K). Based on comparisons of the reported confidence intervals, the maternal estimate for genome length of loblolly pine is significantly different from the maternal estimates reported by ECHT and NELSON (1997). Loblolly pine is closely related to longleaf pine; therefore, taxonomic relatedness does not appear to be a factor regarding the difference among these estimates. A possible factor that might contribute to this difference is the type of marker used for each of these estimates. Putatively, RAPD markers are more randomly dispersed

throughout the genome than RFLP markers (detected from cDNA probes) and isozymes. If clustering does exist among the majority of markers used in the loblolly pine study, then genome length would be underestimated.

Nonuniform marker distributions are observed as gaps among many of the linkage groups (Figure 2; SEWELL and NEALE 1998). These gaps may represent a non-random sampling of the genome resulting from an underrepresentation of markers from these regions. Alternatively, the markers may be random along the physical length of the chromosome, and these gaps may represent an uneven distribution of recombination events along the chromosome (TANKSLEY *et al.* 1992). It is possible that the distribution of expressed genetic probes (*i.e.*, cDNAs and isozymes) is not representative of the entire genome. RAPD markers were included in the analysis in an attempt to add markers from throughout the genome. However, the addition of RAPD markers did not bridge any gaps. This evidence suggests that the genome has been randomly sampled and the gaps represent nonuniform recombination events. Furthermore, the observation that gaps can be found in the same general region in each of the four independent maps (*e.g.*, linkage groups 1 and 5; SEWELL and NEALE 1998) suggests that these gaps represent some sort of cytogenetic phenomenon (*e.g.*, the presence of centromeres; TANKSLEY *et al.* 1992) rather than a simple genome sampling error. Notwithstanding, ~13% of the markers remain unlinked among these mapping populations.

Southern blot analyses using cDNA and gene probes have revealed genes that are found in duplicate, and occasionally multiple, copies in many plant species (*e.g.*, BERNATSKY and TANKSLEY 1986; HELENTJARIS *et al.* 1988). However, the majority of cDNA and gene probes in loblolly pine detect genes in multigene families (*e.g.*, Figure 1; KINLAW and NEALE 1997). These gene families are often significantly larger than their counterparts in angiosperms (KINLAW *et al.* 1994; PERRY and FURNIER 1996) and they are found both dispersed throughout the genome (*e.g.*, Figure 1) or as tightly linked clusters (*e.g.*, LG10/1635; Figure 2).

Polyploidization has played a role in the evolution of gene families in plant genomes. For example, the genetic map of maize revealed many syntenic regions that are duplicated throughout the genome, which are believed to be the result of an ancient polyploidization event, followed by genomic rearrangements (HELENTJARIS *et al.* 1988). Cytogenetic studies of pines have generated conflicting conclusions regarding polyploidization; MEHRA and KHOSHOO (1956) and MIROV (1967) find no evidence for polyploidization, whereas DREWRY (1988) supports this possibility. Currently, the consensus map provides no evidence for the presence of duplicated syntenic regions. Furthermore, gene family sizes in pines vary considerably among different species and

probes (AHUJA *et al.* 1994), suggesting that gene amplification is an ongoing process in pines.

Alternatively, one of the possible mechanisms of gene amplification in conifers is reverse transcription of mRNA templates resulting in processed retropseudogenes (DROUIN and DOVER 1987). KAMM *et al.* (1996) identified a *Ty1-copia*-like retroelement that was highly amplified throughout the pine genome. KVARNHEDEN *et al.* (1995) found evidence for six retropseudogenes in the *cdc2* gene family in *Picea abies*. PICHERSKY (1990) notes that most postulated mechanisms of gene duplication do not result in an "instant," fully diploidized duplication. A plant's ability to self increases the likelihood of establishing the novel duplication in the diploid state, whereas in a highly outbred species this likelihood is decreased. Therefore, the process of gene amplification could produce unlinked, hemizygous nulls within the conifer genome (*e.g.*, Figure 1).

An often overlooked advantage of mapping with forest trees is their longevity. Once the initial time and money is invested, relatively little additional effort is required to maintain this germplasm. Consequently, it is feasible to preserve these pedigrees in multiple environments and score multiple phenotypic traits over time. The genetic information pertaining to this consensus map, based on the immortalized *base* and *qtl* pedigrees, is made publicly available through the *Dendrome* web server and *TreeGenes* database for genome projects in loblolly and other conifers.

This work has benefited from the contributions of many colleagues. In particular, the authors thank J. Cairney, P. Gustafsson, D. Harry, S. Jansson, S. Karpinski and colleagues, C. Loopstra, D. O'Malley, and R. Whetten for contributing characterized gene probes; L. Burris and N. Wheeler for maintaining pedigrees and providing germplasm; D. Bassoni, M. Devey, A. Groover, P. Hodgskiss, K. Jermstad, K. Kiehne, J. Lee, and P. Wilcox for providing technical and/or analytical support; C. Echt, S. Young, and three anonymous reviewers for providing helpful editorial comments. This research was supported by U.S. Department of Agriculture/National Research Initiative Plant Genome Grant 96-35300-3719.

## LITERATURE CITED

- AHUJA, M. R., M. E. DEVEY, A. T. GROOVER, K. D. JERMSTAD and D. B. NEALE, 1994 Mapped DNA probes from loblolly pine can be used for restriction fragment length polymorphism mapping in other conifers. *Theor. Appl. Genet.* **88**: 279-282.
- BEAVIS, W. D., and D. GRANT, 1991 A linkage map based on information from four F<sub>2</sub> populations of maize (*Zea mays* L.). *Theor. Appl. Genet.* **82**: 636-644.
- BENNETZEN, J. L., and M. FREELING, 1993 Grasses as a single genetic system: genome composition, collinearity and compatibility. *Trends Genet.* **9**: 259-261.
- BERNATSKY, R., and S. D. TANKSLEY, 1986 Majority of random cDNA clones correspond to single loci in the tomato genome. *Mol. Gen. Genet.* **203**: 8-14.
- BISHOP, Y. M., S. E. FIENBURG and P. W. HOLLAND, 1975 *Discrete Multivariate Analysis: Theory and Practice*, pp. 492-494. MIT Press, Cambridge, MA.
- BISHOP, D. T., C. CANNINGS, M. SKOLNICK and J. A. WILLIAMSON, 1983 The number of polymorphic DNA clones required to map

- the human genome, pp. 181–200 in *Statistical Analysis of DNA Sequence Data*, edited by B. S. WEIR. Marcel-Dekker, New York.
- CONKLE, M. T., P. D. HODGSKISS, L. B. NUNNALLY and S. C. HUNTER, 1982 Starch gel electrophoresis of conifer seeds: a laboratory manual. Gen. Tech. Rep. PSW-64. Pacific Southwest Forest and Range Experiment Station, USDA Forest Service, Berkeley, CA.
- DEVEY, M. E., K. D. JERMSTAD, C. G. TAUER and D. B. NEALE, 1991 Inheritance of RFLP loci in a loblolly pine three-generation pedigree. *Theor. Appl. Genet.* **83**: 238–242.
- DEVEY, M. E., T. A. FIDDLER, B.-H. LIU, S. J. KNAPP and D. B. NEALE, 1994 An RFLP linkage map for loblolly pine based on a three-generation outbred pedigree. *Theor. Appl. Genet.* **88**: 273–278.
- DEVEY, M. E., A. DELFINO-MIX, B. B. KINLOCH and D. B. NEALE, 1995 Efficient mapping of a gene for resistance to white pine blister rust in sugar pine. *Proc. Natl. Acad. Sci. USA* **92**: 2066–2070.
- DEVEY, M. E., J. C. BELL, D. N. SMITH, D. B. NEALE and G. F. MORAN, 1996a A genetic linkage map for *Pinus radiata* based on RFLP, RAPD, and microsatellite markers. *Theor. Appl. Genet.* **92**: 673–679.
- DEVEY, M. E., M. M. SEWELL and D. B. NEALE, 1996b A comparison of loblolly and radiata pine genomes using RFLP markers, pp. 478–480 in *Tree Improvement for Sustainable Tropical Forestry*, edited by M. J. DIETERS, A. C. MATHESON, D. G. NIKLES, C. E. HARWOOD and S. M. WALKER. Queensland Forestry Research Institute, Gympie, Queensland, Australia.
- DREWRY, A., 1988 The G-banded karyotype of *Pinus resinosa* Ait. *Silvae Genet.* **37**: 218–221.
- DROUIN, G., and G. A. DOVER, 1987 A plant processed pseudogene. *Nature* **328**: 557–558.
- ECHT, C. S., and C. D. NELSON, 1997 Linkage mapping and genome length in eastern white pine (*Pinus strobus* L.). *Theor. Appl. Genet.* **94**: 1031–1037.
- FRANKLIN, E. C., 1970 Survey of mutant forms and inbreeding depression in species of the family Pinaceae. Research Paper SE-61. Southeastern Forest Experiment Station, Asheville, NC.
- GENTZBITTEL, L., F. VEAR, Y.-X. ZHANG, A. BERVILLÉ and P. NICOLAS, 1995 Development of a consensus linkage RFLP map of cultivated sunflower (*Helianthus annuus* L.). *Theor. Appl. Genet.* **90**: 1079–1086.
- GERBER, S., and F. RODOLPHE, 1994 An estimation of the genome length of maritime pine (*Pinus pinaster* Ait.). *Theor. Appl. Genet.* **88**: 289–292.
- GROOVER, A., M. DEVEY, J. LEE, R. MEGRAW, T. MITCHELL-OLDS *et al.*, 1994 Identification of quantitative trait loci influencing wood specific gravity in an outbred pedigree of loblolly pine. *Genetics* **138**: 1293–1300.
- GROOVER, A. T., C. G. WILLIAMS, M. E. DEVEY, J. M. LEE and D. B. NEALE, 1995 Sex-related differences in meiotic recombination frequency in *Pinus taeda*. *J. Hered.* **86**: 157–158.
- HAMRICK, J. L., and M. J. GODT, 1990 Allozyme diversity in plant species, pp. 43–63 in *Plant Population Genetics, Breeding, and Genetic Resources*, edited by A. H. D. BROWN, M. T. CLEGG, A. L. KAHLER and B. S. WEIR. Sinauer Associates, Sunderland, MA.
- HAUGE, B. M., S. M. HANLEY, S. CARTINHO, J. M. CHERRY, H. M. GOODMAN *et al.*, 1993 An integrated genetic/RFLP map of the *Arabidopsis thaliana* genome. *Plant J.* **3**: 745–754.
- HELENTJARIS, T., D. WEBER and S. WRIGHT, 1988 Identification of the genomic locations of duplicate nucleotide sequences in maize by analysis of restriction fragment length polymorphisms. *Genetics* **118**: 353–363.
- HULBERT, S. H., T. W. ILOTT, E. J. LEGG, S. E. LINCOLN and R. W. MICHELMORE, 1988 Genetic analysis of the fungus, *Bremia lactucae*, using restriction fragment length polymorphisms. *Genetics* **120**: 947–958.
- JERMSTAD, K. D., A. M. REEM, J. R. HENIFIN, N. C. WHEELER and D. B. NEALE, 1994 Inheritance of restriction fragment length polymorphisms and random amplified polymorphic DNAs in coastal Douglas fir. *Theor. Appl. Genet.* **89**: 758–766.
- KAMM, A., R. L. DOUDRICK, J. S. HESLOP-HARRISON and T. SCHMIDT, 1996 The genomic and physical organization of *Tyl-copia*-like sequences as a component of large genomes in *Pinus elliotii* var. *elliotii* and other gymnosperms. *Proc. Natl. Acad. Sci. USA* **93**: 2708–2713.
- KIANIAN, S. F., and C. F. QUIROS, 1992 Generation of a *Brassica oleracea* composite RFLP map: linkage arrangements among various populations and evolutionary implications. *Theor. Appl. Genet.* **84**: 544–554.
- KINLAW, C. S., and D. B. NEALE, 1997 Complex gene families in pine genomes. *Trends Plant Sci.* **2**: 356–359.
- KINLAW, C. S., S. M. GERTTULA and M. C. CARTER, 1994 Lipid transfer protein of loblolly pine are members of a complex gene family. *Plant Mol. Biol.* **26**: 1213–1216.
- KVARNHEDEN, A., K. TANDRE and P. ENGSTRÖM, 1995 A *cdc2* homologue and closely related processed retroseudogenes from Norway spruce. *Plant Mol. Biol.* **27**: 391–403.
- LANDER, E. S., P. GREEN, J. ABRAHAMSON, A. BARLOW, M. J. DALY *et al.*, 1987 MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics* **1**: 174–181.
- LIU, Y.-G., N. MITSUKAWA, C. LISTER, C. DEAN and R. F. WHITTIER, 1996 Isolation and mapping of a new set of 129 RFLP markers in *Arabidopsis thaliana* using recombinant inbred lines. *Plant J.* **10**: 733–736.
- MANLY, K. F., 1993 A Macintosh program for storage and analysis of experimental genetic mapping data. *Mamm. Genome* **4**: 303–313.
- MEHRA, P. N., and T. N. KHOSHOO, 1956 Cytology of conifers. *Int. J. Genet.* **54**: 165–185.
- MIROV, M. T., 1967 *The Genus Pinus*. The Ronald Press Company, New York.
- MITTON, J. B., 1994 Molecular approaches to population biology. *Annu. Rev. Ecol. Syst.* **25**: 45–69.
- MORAN, G. F., J. C. BELL and A. J. HILLIKER, 1983 Greater meiotic recombination in male vs. female gametes in *Pinus radiata*. *J. Hered.* **74**: 62.
- NEALE, D. B., J. C. WEBER and W. T. ADAMS, 1984 Inheritance of needle tissue isozymes in Douglas fir. *Can. J. Genet. Cytol.* **26**: 459–468.
- NELSON, C. D., W. L. NANCE and R. L. DOUDRICK, 1993 A partial genetic linkage map of slash pine (*Pinus elliotii* Engelm. var. *elliotii*) based on random amplified polymorphic DNAs. *Theor. Appl. Genet.* **87**: 145–151.
- NELSON, C. D., T. L. KUBISIAK, M. STINE and W. L. NANCE, 1994 A genetic linkage map of longleaf pine (*Pinus palustris* Mill.) based on random amplified polymorphic DNAs. *J. Hered.* **85**: 433–439.
- O'MALLEY, D. M., S. PORTER and R. R. SEDEROFF, 1992 Purification, characterization, and cloning of cinnamyl alcohol dehydrogenase in loblolly pine (*Pinus taeda* L.). *Plant Physiol.* **98**: 1364–1371.
- PEDERICK, L. A., 1970 Chromosome relationships between *Pinus* species. *Silvae Genet.* **19**: 171–180.
- PERRY, D. J., and G. R. FURNIER, 1996 *Pinus banksiana* has at least seven expressed alcohol dehydrogenase genes in two linked groups. *Proc. Natl. Acad. Sci. USA* **93**: 13020–13023.
- PICHERSKY, E., 1990 Nomad DNA—a model for movement and duplication of DNA sequences in plant genomes. *Plant Mol. Biol.* **15**: 437–448.
- PLOMION, C., D. M. O'MALLEY and C.-E. DUREL, 1995 Genomic analysis in maritime pine (*Pinus pinaster*): comparison of two RAPD maps using selfed and open-pollinated seeds of the same individual. *Theor. Appl. Genet.* **90**: 1028–1034.
- QI, X., P. STAM and P. LINDHOUT, 1996 Comparison and integration of four barley genetic maps. *Genome* **39**: 379–394.
- SEWELL, M. M., and D. B. NEALE, 1998 A consensus map for loblolly pine. Dendrome database, <http://dendrome.ucdavis.edu/Data/>.
- STAM, P., 1993 Construction of integrated genetic linkage maps by means of a new computer package: JOINMAP. *Plant J.* **3**: 739–744.
- TANKSLEY, S. D., M. W. GANAL, J. P. PRINCE, M. C. DE VICENTE, M. W. BONIERBALE *et al.*, 1992 High density molecular linkage maps of the tomato and potato genomes. *Genetics* **132**: 1141–1160.
- TULSIERAM, L. K., J. C. GLAUBITZ, G. KISS and J. E. CARLSON, 1992 Single tree genetic linkage mapping in conifers using haploid DNA from megagametophytes. *Bio/Technology* **10**: 686–690.
- WAKAMIYA, I., R. J. NEWTON, J. S. JOHNSTON and H. J. PRICE, 1993 Genome size and environmental factors in the genus *Pinus*. *Am. J. Bot.* **80**: 1235–1241.