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, Bradley K. Sherman and David B. Neale

*Institute of Forest Genetics, USDA Forest Service, Pacific Southwest Research Station, Placerville, California 95667 and
†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 (1359.5 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 ~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, Haage et al. 1993; Brassica oleracea, Kianian and Quiros 1992; Helianthus annuus, Gentzblett 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., morphological 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; Gentzblett 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; Haage et al. 1993; Gentzblett et al. 1995).

Loblolly pine (Pinus taeda L.) is a member of the genus Pinus, which encompasses ~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 1996). 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; Waki-miya et al. 1993) and are composed of complex gene families (Kianian 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
agronomic interest, pedigrees involving backcrosses (BCs) or F2s 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 F2). Therefore, only two alleles segregate in these mapping populations, and the resulting genetic map represents the recombination of the F2 parent. Furthermore, because two homozygous inbred lines are used to generate the F2 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 (DEVEY 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 (DEVEY et al. 1994). The second pedigree (referred to as the g1t 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 g1t pedigree. Therefore, a subset of 88 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 DEVEY et al. (1991). The primary source of RFLP probes was a loblolly pine complementary DNA (cDNA) library constructed by DEVEY et al. (1991). Probes from a genomic DNA library were also used (DEVEY 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. (1995). 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 NIALE et al. (1984). Gel electrophoresis and protein staining followed the method of CONKLE et al. (1982). The protein stain for cinnamaldehyde dehydrogenase followed the method of O'MALLEY et al. (1992).

Data organization: The original base and g1t 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 reasured, and band migration distances were visually compared as a means to identify orthologous markers and assign them to 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 (identified 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 ancestry could not be determined (i.e., when a pair of grandparents were heterozygous for the same allele pair), the data remained in an "ancestry-unknown" format.

Data sets were then coded for linkage analyses with MAPMAKER (McIntosh version 2.0; LANDER et al. 1987), MAP-_MANAGER (McIntosh version 2.0; LANDER et al. 1987), and JOIN- MAP (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 MAP-_MANAGER 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 MAP-MANAGER analyses.

Concurrent linkage analyses of multiple populations: Chi-square values were calculated for individual markers to detect deviation (P ≤ 0.05) of genetic 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 LOD ≥ 5.0 and recombination fraction (θ) ≤ 0.50. 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 LOD ≥ 5.0 and θ ≤ 0.30 and/or orthologous markers. Additional markers were then placed onto these linkage groups at LOD ≥ 3.0 and θ ≤ 0.40. This strategy ensures confidence in the formation of each set of linkage group homologues; however, it does allow 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 θ 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 θ with a weighted average value (Stam 1995). Using the fixed sequence command, these data were integrated relative to the established order of BC markers ascertainment 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 pairs 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: Eich 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) (Hulebert et al. 1988), and expected genome coverage, E(C)% (Bishop et al. 1985), 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, Cobs, was calculated from the results of map construction using a 30-cM scale (Nelson et al. 1994). These procedures are detailed in Eich 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 gtl 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., LG2/2986A or LG4/gtlpat, respectively, to indicate a specific linkage group. The constitutive and consensuz map (Swellell and Nalle 1998) and all information pertaining to map construction are located in the Dendrome and/or TreeGenus databases (http://dendrome.ucdavis.edu/TreeGenus/). The cDNA clones used for RFLP analyses are freely distributed and may be obtained by submitting a request to dNeal@delendrome.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 (basemat), base-paternal (basepat), gtl-maternal (gtlmat), and gtl-paternal (gtlpat). Of these 581 markers (430 CDNA RFLPs, 56 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 103 markers were orthologous complements found in two or more of the constituent
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).

Alicic 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. olerek*, Kianian and Quiros 1992; *H. annus*, Gentzbittel et al. 1995; *P. radiata*, Devey et al. 1996a; *Pseudotsuga menziesii*, Jerstad et al. 1994). Several factors could account for this. The “null” allele might go undetected as a result of conigration 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*, *q1f-mat*, and *q1f-pat*—contained 109, 106, 191, and 173 genetic markers, respectively; i.e., 97 (89.0%), 79 (74.5%), 170 (93.7%), and 153 (87.4%) were linked at 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*, *q1f-mat*, and *q1f-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

<table>
<thead>
<tr>
<th>Population</th>
<th>Total markers</th>
<th>Mapped markers*</th>
<th>Map coverage†</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>JOINMAP</td>
</tr>
<tr>
<td><em>base-mat</em></td>
<td>109</td>
<td>97 (89.0)</td>
<td>699.2</td>
</tr>
<tr>
<td><em>base-pat</em></td>
<td>106</td>
<td>79 (74.5)</td>
<td>591.3</td>
</tr>
<tr>
<td><em>q1f-mat</em></td>
<td>191</td>
<td>179 (93.7)</td>
<td>907.8</td>
</tr>
<tr>
<td><em>q1f-pat</em></td>
<td>175</td>
<td>153 (87.4)</td>
<td>956.1</td>
</tr>
<tr>
<td>Population total</td>
<td>561</td>
<td>508 (87.4)</td>
<td>1359.4</td>
</tr>
<tr>
<td>Cenemas</td>
<td>561</td>
<td>357</td>
<td></td>
</tr>
</tbody>
</table>

*Percentage of mapped markers in parentheses.
†Map units in centimorgans using Kosambi mapping function [cM(K)].
TABLE 2
Number of orthologous markers used for map integration

<table>
<thead>
<tr>
<th>Mapping population</th>
<th>Orthologous markers</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Intrapedigree</td>
<td>Interpedigree</td>
<td>Total*</td>
</tr>
<tr>
<td>base-mat</td>
<td>35</td>
<td>38</td>
<td>55</td>
</tr>
<tr>
<td>base-pat</td>
<td>35</td>
<td>44</td>
<td>57</td>
</tr>
<tr>
<td>qtl-mat</td>
<td>53</td>
<td>51</td>
<td>81</td>
</tr>
<tr>
<td>qtl-pat</td>
<td>53</td>
<td>35</td>
<td>65</td>
</tr>
<tr>
<td>Consensus</td>
<td>—</td>
<td>—</td>
<td>108</td>
</tr>
</tbody>
</table>

*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 groups (e.g., LG2/base-2885.B and LG9/qtlGot-4). These comparisons among individual maps from within a species are consistent with those among a self and F1 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 parenthesis.

TABLE 3
Global test for deviation between maternal and paternal recombination rates (θ)

<table>
<thead>
<tr>
<th>Pedigree</th>
<th>No. of intervals</th>
<th>Mean difference in θ (%)</th>
<th>SE (%)</th>
<th>z*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Base</td>
<td>22</td>
<td>26.12</td>
<td>5.61</td>
<td>4.055*</td>
</tr>
<tr>
<td>QTL</td>
<td>29</td>
<td>26.33</td>
<td>6.46</td>
<td>4.077*</td>
</tr>
<tr>
<td>Combined</td>
<td>51</td>
<td>26.23</td>
<td>4.30</td>
<td>6.105*</td>
</tr>
</tbody>
</table>

* z test with the delta method (Bishop et al. 1975).

p < 0.0001.

TABLE 4
Estimated genome length and coverage

<table>
<thead>
<tr>
<th>Mapping population</th>
<th>No. of loci</th>
<th>E(G)*</th>
<th>E(G)%*</th>
<th>Cexp*</th>
<th>Cexp%*</th>
</tr>
</thead>
<tbody>
<tr>
<td>base-mat</td>
<td>88</td>
<td>1515.5 (K)</td>
<td>70.80</td>
<td>918.1</td>
<td>69.79</td>
</tr>
<tr>
<td>base-pat</td>
<td>79</td>
<td>1579.5 (H)</td>
<td>907.3</td>
<td>73.53</td>
<td></td>
</tr>
<tr>
<td>qtl-mat</td>
<td>157</td>
<td>2129.1 (K)</td>
<td>55.40</td>
<td>1178.3</td>
<td>55.34</td>
</tr>
<tr>
<td>qtl-pat</td>
<td>147</td>
<td>2554.7 (H)</td>
<td>1152.9</td>
<td>54.15</td>
<td></td>
</tr>
<tr>
<td>qtl-pat</td>
<td>157</td>
<td>1363.4 (K)</td>
<td>91.48</td>
<td>1101.0</td>
<td>80.76</td>
</tr>
<tr>
<td>qtl-pat</td>
<td>147</td>
<td>1634.0 (H)</td>
<td>1213.5</td>
<td>89.01</td>
<td></td>
</tr>
<tr>
<td>qtl-pat</td>
<td>147</td>
<td>1886.2 (K)</td>
<td>84.12</td>
<td>1541.9</td>
<td>83.88</td>
</tr>
<tr>
<td>qtl-pat</td>
<td>147</td>
<td>2225.6 (H)</td>
<td>1254.2</td>
<td>83.46</td>
<td></td>
</tr>
</tbody>
</table>

* Average estimated genome length in centimorgans using Kosambi (K) and Haldane (H) mapping functions.
1 Observed genome coverage (cM(K)) calculated from linkage maps constructed with JOINMAP (above) and MAPMAKER (below).
2 Observed genome coverage (cM(K)) calculated from linkage maps constructed with JOINMAP (above) and MAPMAKER (below).
Figure 2.—A consensus map for loblolly pine. The constituent maps from the base and qtl pedigrees (Sewell and Neale 1998) are located on the World Wide Web at http://dendrom.ucdavis.edu/Data/.
ses) at LOD threshold 4 were as follows: 1315.5 (1117.9–1612.7), base-mat; 2129.1 (1966.4–2845.7), base-pat; 1363.4 (1241.1–1526.6), gfl-mat; and 1838.2 (1608.3–2075.3), gfl-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 gfl) and also among pedigrees for gfl-mat and base-pat (and nearly so between base-mat and gfl-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, gfl-mat, and gfl-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 gfl pedigree and account for an additional 52 cM (K). The remaining six small “single-population” linkage groups cover 135 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 135 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 gfl 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., Devy 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., Devy 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/gfl-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 recombinational 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 nonrandom 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., Bernatzky and Tanksley 1986; Heijtenjars 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 (Heijtenjars et al. 1988). Cytogenetic studies of pines have generated conflicting conclusions regarding polyploidization; Mehra and Krashevo (1956) and Mrnov (1967) find no evidence for polyploidization, whereas Drozywa (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 retroseudogenes (Drouin and Dover 1987). Kamis et al. (1996) identified a Ty1-copia-like retroelement that was highly amplified throughout the pine genome. Kvarnbreden et al. (1995) found evidence for six retroseudogenes in the ode2 gene family in Pinus abies. Pitchersky (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 gil 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. Gustafson, D. Harry, S. Jansson, S. Karpinski and colleagues, C. Loopstra, D. O'Malley, and R. Whetten for contributing characterized gene probes; L. Burns and N. Wheeler for maintaining pedigrees and providing germplasm; D. Basnuni, M. Devvy, A. Groover, P. Hodgkinson, K. Jerstad, K. Kielne, J. Lee, and P. Wilcox for providing technical and/or analytical support; C. Echt, J. 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


Bishop, D. T., C. Cannings, M. Skolnick and J. A. Williamson, 1988. The number of polymorphic DNA clones required to map