Genetic recombinational and physical linkage analyses on slash pine

R. L. Doudrick

Southern Institute of Forest Genetics, USDA Forest Service, Saucier, Mississippi, 39574-9344, USA

SUMMARY

Slash pine is native to the southeastern USA, but is commercially valuable world-wide as a timber-, fiber- and resin-producing species. Breeding objectives emphasize improving planting stock through selection of open-pollinated families that exhibit low percentage infections using composites of spore collections of *C. q. fusiforme* (Zobel et al., 1971). Unfortunately, this approach provides little genetic information. Artificial inoculation studies suggest specific genetic factors in host and pathogen contribute to disease development (Snow et al., 1975; Griggs and Walkinshaw, 1982; Kinloch and Walkinshaw, 1991; Doudrick and Nelson, 1993; Nelson et al., 1993a). Identification of markers linked to factors conditioning specific resistance in slash pine should expand our knowledge of the micro-coevolution of host and pathogen and assist breeders (Nance et al., 1992).

INTRODUCTION

Slash pine (*Pinus elliottii* Engelm. var. *elliottii*) is native to the southeastern United States. It is a major forest component on flatwoods and along streams and the edges of swamps and bays, wherever soil moisture is ample (Lohrey and Kossuth, 1990). Slash pine is planted commonly inside and outside its natural range (Boyer and South, 1984), as well as in exotic plantings in Asia (Pan, 1989), South America (Picchi and Barrett, 1967), Africa and Australia (Mullin et al., 1978). It is commercially valuable world-wide as a timber-, fiber- and resin-producing species.

In the United States, fusiform rust, caused by *Cronartium quercuum* (Berk.) Miyabe in Shirai f. sp. *fusiforme*, continues to be the most damaging disease on slash pine (USDA Forest Service, 1994). Control measures focus primarily on improving planting stock through selection of open-pollinated stock that exhibit low percentage infections using composites of spore collections of *C. q. fusiforme* (Zobel et al., 1971). Unfortunately, this approach provides little genetic information. Artificial inoculation studies suggest specific genetic factors in host and pathogen contribute to disease development (Snow et al., 1975; Griggs and Walkinshaw, 1982; Kinloch and Walkinshaw, 1991; Doudrick and Nelson, 1993; Nelson et al., 1993a). Identification of markers linked to factors conditioning specific resistance in slash pine should expand our knowledge of the micro-coevolution of host and pathogen and assist breeders (Nance et al., 1992).

GENETIC RECOMBINATIONAL LINKAGE ANALYSES

Towards this end a molecular genetic recombinational linkage

staining are being used to integrate the genetic recombinational maps. A karyotype and ideogram have been prepared for slash pine (2n=2x=24); metaphase chromosome preparations show 11 pairs of long metacentric chromosomes and one shorter pair of submetacentric chromosomes. Patterns of fluorescence in situ hybridization to genes for the large and small rRNA subunits and fluorochrome banding patterns using the GC-base-specific chromomycin A3 (CMA) and AT-base-specific 4',6-diamidino-2-phenylindole (DAPI) allowed all twelve pairs of chromosomes to be identified and a standard karyotype established.

A family of sequences associated with (TTTAGGG)n related repeats has been identified in slash pine using a labeled synthetic oligonucleotide probe. Fluorescence in situ hybridization shows a weak signal at telomeres and significantly stronger intensity at non-telomeric sites. The most common non-telomeric location was in the pericentric regions of chromosomes; interstitial sites of hybridization were relatively common.

Microsatellite DNAs, an abundant retrotransposon-like element, and total genomic in situ hybridization and species and chromosome specific DNAs are being evaluated for analyses of interspecific hybrids and chromosome evolution between related species. Interest in low and single copy sequences is increasing.

Key words: Genetic mapping, Pinus elliottii, Telomere, In situ hybridization, Karyotype, rDNA
map of the nuclear genome in a mature slash pine tree was constructed using random amplified polymorphic DNA (RAPD) markers (Nelson et al., 1993b). The tree was hypothesized to possess at least four resistance loci; apparently homozygous dominant for resistance at one locus, homozygous recessive at a second locus and heterozygous at two additional loci (Doudrick and Nelson, 1993; Nelson et al., 1993a). The total number and allelism of resistance loci in slash pine are unknown, and dependent on identifying additional pathotypes of C. q. fusiforme capable of conditioning differential interactions.

The map was constructed using a source of segregating haploid tissue unique to conifers, the female megagametophyte. This megagametophyte is nutritive endosperm resulting from mitotic divisions of a single meiotic product; the product giving rise also to the embryo of the same seed. Since the DNA prepared for analysis from the megagametophyte is haploid and strictly maternally inherited, segregation and recombination can be evaluated in a sample of wind-pollinated seeds from a single tree (Guries et al., 1978; Conkle, 1981) and the RAPD markers observed as either present or absent (Tulsieram et al., 1993).

RAPD reaction protocols were described by Welsh and McClelland (1990), and Williams et al. (1990) and modified for slash pine by Nelson et al. (1993b). To identify clear and repeatable polymorphisms, arbitrary (sequence 50-80% C+G content), 10 base, oligonucleotide RAPD primers were first screened for amplification of segments within eight megagametophyte DNA samples. These RAPDs were characterized further within a sample of 68 megagametophytes from the same tree. Segments segregating in a Mendelian ratio of 1:1 (α=0.05) by chi-square analysis (χ², present-to-absent, were classified and mapped using multi-point linkage analysis (MAPMAKER II version 1.9; Lander et al., 1987).

The analysis revealed 13 linkage groups of three or more loci, ranging in size from 28 to 68 recombinational units (centiMorgans, cM), and nine linked pairs of loci. The 22 groups and pairs included 73 RAPD markers that covered a genetic map distance of ~782 cM; genome size estimates ranged from ~2,880 to 3,560 cM. With a physical size of >20,000 Mbp (Dhillon, 1980; Ohri and Khosho, 1986; Greilhuber, 1988; Wakamiya et al., 1993) and this linkage data, values of 6.0-6.9×10⁶ bp/Cm are obtained for comparative and map-based analyses. Using a 30 cM map scale and including the 24 unlinked markers and the ends of the 13 linkage groups and nine linked pairs, these RAPD markers account for ~2,160 cM or 64-75% of the genome. Because of this low percentage coverage, 80 additional RAPD markers were mapped (C. D. Nelson, personal communication), permitting placement of 131 loci total into 20 linkage groups of three or more loci, nearly doubling the coverage in the groups to a genetic map distance of ~1,347 cM (Fig. 1).

Data are preliminary on a map of another single mature slash pine; this tree is hypothesized to have a different complement of specific resistance loci. Analyses are incomplete with only 32 megagametophyte DNAs scored, but show 151 polymorphic segments segregating 1:1 (T. L. Kubisiak, personal communication).

In similar fashion, a single mature slash pine tree selected for high wood specific gravity and its progeny screened for heritability of fusiform rust disease reaction was RAPD mapped (Byram et al., 1992). The RAPD primers selected for mapping produced 185 scoreable markers with segregation ratio 1:1. The 185 loci were placed into 19 linkage groups of three or more loci and four pairs. The resulting map covered ~1,776 cM, about 59% of the genome.

RAPDs were used also to simultaneously construct linkage maps of the mature slash pine and longleaf pine (Pinus palustris Mill.) parent trees of an interspecific F1 family (Kubisiak et al., 1995). RAPD primers were screened against the two parents and six F1 progenies. A total of 247 segregating loci (233 at 1:1; 14 at 3:1) and 87 polymorphic (between-parents), but non-segregating, loci were identified; 101 loci segregating 1:1 for the slash pine parent. The 101 marker loci were placed in 13 groups and six pairs for ~952.9 cM using multi-point linkage analysis (MAPMAKER/EXP version 3.0; Lincoln et al., 1992). When a weighted average distance between groups and unlinked markers was added with the mapped markers in the groups and pairs, genome coverage was estimated at ~1,462.9 cM. For a 30 cM map scale, this is approximately 61.7% of the total genome length, the estimated length somewhat smaller for slash pine by this method than from the maps above, approximately 2,300 cM compared to about 3,000 cM. Although Moran et al. (1983) reported sexually related regional variations in meiotic recombination in Pinus radiata D. Don, greater recombination rates for female vs male meiocytes is common in plants (Rhoades, 1941) and may explain the above differences. Linkage of the 3:1 loci to 1:1 loci in each of the parental maps was used to infer further linkages within maps, as well as potentially homologous counterparts between maps. Three of the longleaf-pine linkage groups appear to have potentially homologous counterparts to four different slash-pine linkage groups identified by Nelson et al. (1993b).

All the current maps of mature slash pine trees are of medium density and incomplete (more groups than chromosomes). The extent of genomic coverage in the maps does not allow for efficient mapping of genes conditioning specific resistance for C. q. fusiforme in this species. Obviously more markers are required to bridge gaps between current groupings and expand coverage towards 100%. Estimates of the number of markers required to obtain 90% coverage of the pine genome (average spacing of 20-30 cM) suggest that approximately 200 to 300 additional markers will be required (Nelson et al., 1993b). Although automated preparation of RAPD reactions showed improvements in speed (Nance and Schumate, 1992) and repeatability, at the current levels of genomic coverage (60-85%), screening more primers for additional polymorphisms may not be cost effective. The very high bp/Cm average ratio in slash pine will require very high resolution mapping <0.1 cM average spacing for map-based cloning applications (Nelson et al., 1994). Increasing sample size and applying bulked segregant analysis (Michelmore et al., 1991) should help identify polymorphisms near termini of the groups and remaining unlinked markers.

Sampling other regions of the genome to supplement RAPD mapping efforts may be best served by integrating other types of markers (e.g. isozymes, proteins, restriction fragment length polymorphisms (RFLPs), sequence characterized sites), markers commonly used to analyze quantitative genetic variation and study genome structure and evolution. The additional markers will be especially useful to evaluate synteny with other species. Conkle (1981) using isozymes identified allelic
MOLECULAR CYTOGENETIC ANALYSES

A detailed karyotype could be used to efficiently assign recombinational map linkage groups to physical chromosomes, integrate the various maps, and complement studies on the molecular and quantitative genetics of slash pine. The lack, however, of morphological variability in chromosomes historically has impeded progress in karyotyping species of Pinus (Pederick, 1967). Most Pinus species, including slash pine (as P. elliottii, Mergen, 1958; as Pinus caribaea Morelet, Mehra and Khosho, 1956; Ogoshi and Nakata, 1955), have 12 pairs of chromosomes and are characterized by 10 or 11 pairs of long, metacentric chromosomes and one or two pairs of shorter submetacentric chromosomes. DNA-DNA in situ hybridization is the key to developing high-resolution techniques that can be applied reproducibly to samples, and linking molecular results with chromosomes (Heslop-Harrison et al., 1991).

Doudrick et al. (1995) have prepared a karyotype and ideogram for slash pine (2n=2x=24) using mitotic metaphase cells from root-tips of seedlings; each metaphase had 11 pairs of long metacentric chromosomes and one shorter pair of submetacentric chromosomes. Patterns of fluorescence in situ hybridization using the coding sequences of the genes for 18S-

variation and linkage in five Pinus spp. The loci consistently mapped in the same order and with similar map distances among the species from different subsections of the genus. Similar data on slash pine would be immediately informative. Unique to slash pine are the studies reported by Michelozzi et al. (1990). They showed simple inheritance in 85 slash pine clones for high terpene content in needles, a cortical monoterpene, and ~-phellandrene, and ~-pinene, respectively. These RFLPs provide another opportunity to compare the genomes of related pine species through hybridization patterns and possibly construct RFLP maps (Ahuja et al., 1994).

Fig. 1. RAPD genetic recombinational linkage map of a mature slash pine tree (Southern Institute of Forest Genetics clone 87). One thousand three hundred map units (cM) are present in linkage groups with three or more loci, indicated by letters A through T. Marker names are given on the right-hand side of the linkage groups and Haldane (Haldane function; Haldane 1919) cM distances are given on the left-hand side. The marker names contain the primer ID and approximate length in base pairs of the mapped in the same order and with similar map distances.
5.8S-25S rRNA (18S-25S rDNA) and the complete gene for 5S rRNA and accompanying application of fluorochrome banding allowed identification of all twelve pairs of chromosomes. Fluorescence in situ hybridization using biotin-labeled pTa71 (from *Triticum aestivum* L.; Gerlach and Bedbrook, 1979) showed seven pairs of chromosomes with intercalary sites (six major sites and one of lesser intensity and therefore possibly fewer copies) for 18S-25S rDNA and one pair with a paracentromeric site (Fig. 2). The pXV1 (from *Beta vulgaris* L.; Schmidt et al., 1994) digoxigenin-labeled probe localized a major 5S rDNA hybridization site on another pair of chromosomes (Fig. 3). Two minor sites occurred on other chromosomes, one with and one without a site of 18S-25S rDNA. In both cases, a heterologous probe of DNA from an angiosperm was used as probe, thus only the sites where the rRNA coding sequence is present were identified.

After in situ hybridization, chromosomes were stained with the AT-base-specific fluorochrome 4',6-diamidino-2-phenylindole (DAPI) and the GC-base-specific fluorochrome chromomycin A3 (CMA) (Figs 2, 3). Positive CMA bands at intercalary sites were detected only on two pairs of chromosomes, paracentromeric sites only on three pairs of chromosomes, and bands at both regions on four pairs. Many bands of CMA showed at sites of hybridization of 18S-25S rDNA but one major band occurred at a centromere without an rDNA site. After staining with DAPI, bands appeared at AT-rich intercalary and/or centromeric regions of nearly all chromosomes. A characteristic double band of DAPI occurred near the centromere of one pair of chromosomes.

Patterns of fluorescence in situ hybridization and fluorochrome banding allowed all twelve pairs of chromosomes in slash pine to be identified (Fig. 4; Table 1). As reported for other species of pines (Sax and Sax, 1933), previously only a single chromosome pair, the shortest and only nonmetacentric chromosomes, could be identified unequivocally by size and centromere index alone. All 12 chromosome pairs were identified consistently after probing for the 18S-25S and 5S rDNA loci, and fluorochrome staining. The staining results using CMA and DAPI compare favorably with banding patterns reported for *P. banksiana* Lamb. and *P. nigra* Arnold (MacPherson and Filion, 1981), *P. densiflora* Sieb. et Zucc.

**Fig. 2.** In situ hybridization of rDNA and fluorochrome banding to mitotic metaphase chromosomes of root-tips of seedlings of slash pine: (a) in situ hybridization showing the eight pairs of sites of 18S-25S rDNA detected red. (b) CMA staining showing GC-enriched chromosome segments (bright yellow). Most sites are co-incident with sites of 18S-25S rDNA, but a centromeric pair on chromosome 4 is prominent but not rDNA-associated (arrows). (c) DAPI staining showing the 24 chromosomes and bright, AT-enriched bands at most centromeres. Chromosome 10 shows double bands on either side of the centromere (arrows). (d) Overlay of images a, b and c. Note also banding and hybridization sites in the adjacent interphase nucleus.
Fig. 3. In situ hybridization of rDNA and fluorochrome banding to spreads of metaphase chromosomes of root-tips of seedlings of slash pine: (a) in situ hybridization showing the eight pairs of sites of 18S-25S rDNA detected red; (b) CMA staining showing GC-enriched chromosome segments (bright yellow); (c) in situ hybridization using a probe of 5S rDNA, detected yellow-green, showing the major (chromosome 9) and two minor pairs of 5S sites (chromosomes 1 and 2; one chromosome 1 site, arrowhead, is obscured underneath another chromosome); (d) DAPI staining showing the morphology of the 24 chromosomes.

(Hizume et al., 1989), and P. luchuensis Mayr (Hizume et al., 1990), but are dissimilar for P. thunbergii Parl. (Hizume et al., 1989), and P. koraiensis Sieb. et Zucc., P. resinosa Ait., and P. strobus L. (MacPherson and Filion, 1981). The slash pine karyotype is proposed therefore to be the standard karyotype.

Table 1. Criteria to identify metaphase chromosomes in spreads from root-tips of seedlings of slash pine

<table>
<thead>
<tr>
<th>Chromosome</th>
<th>Important identification criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>1*</td>
<td>Minor, terminal 5S rDNA site at end of long arm.</td>
</tr>
<tr>
<td>2*</td>
<td>18S-25S rDNA site on short arm. Minor terminal 5S rDNA site on the long arm.</td>
</tr>
<tr>
<td>3</td>
<td>18S-25S rDNA site on short arm. One of the largest chromosomes.</td>
</tr>
<tr>
<td>4*</td>
<td>18S-25S rDNA site on long arm. Unique, strong CMA band at centromere.</td>
</tr>
<tr>
<td>5*</td>
<td>18S-25S rDNA site on short arm. Unique in not having DAPI band at centromere.</td>
</tr>
<tr>
<td>6</td>
<td>Weak 18S-25S rDNA on short arm, weak CMA band at centromere. Not one of largest chromosomes.</td>
</tr>
<tr>
<td>7*</td>
<td>Unique 18S-25S rDNA site at centromere.</td>
</tr>
<tr>
<td>8</td>
<td>18S-25S rDNA site on short arm. Not one of largest chromosomes.</td>
</tr>
<tr>
<td>9*</td>
<td>Unique major 5S rDNA site at intercalary position.</td>
</tr>
<tr>
<td>10*</td>
<td>Unique major DAPI bands each side of the centromere.</td>
</tr>
<tr>
<td>11</td>
<td>18S-25S rDNA site on short arm. Small chromosome.</td>
</tr>
<tr>
<td>12*</td>
<td>Smallest chromosome, unique in having unequal arms.</td>
</tr>
</tbody>
</table>

*Chromosomes that normally can be identified by inspection and without reference to other chromosomes in the karyotype. Only invariable and major bands are described. All chromosomes, except 12, are metacentric so arms with markers may reverse (see Fig. 4).

Telomeres are the extreme termini of chromosomes and contain tandemly repeated copies of the sequence (T/A)$_{1-4}$G$_1-8$ in all eukaryotes studied (Zakian, 1989; Blackburn, 1991). Characterizing the slash pine telomeres would further contribute to unifying the genetic recombinational linkage maps and provide anchor points for the ends of linkage groups. In situ hybridization studies showed that sequences homologous to the Arabidopsis thaliana (L.) Heyn. telomere repeat (TTTAGGG)$_n$ are located in the terminal regions of other plant chromosomes (Schwarzacher and Heslop-Harrison, 1991). In situ hybridization studies showed that on a number of different plant chromosomes the TTAGGG repeat is associated with repetitive sequences that do not appear to be localized in the terminal regions (Cheung et al., 1994).

A family of sequences related to the telomeric repeats of A. thaliana has been identified in slash pine. The probe was a heterogeneous population of biotinylated molecules containing repeat arrays of various lengths produced by concatenation of the simple monomers (TTTAGGG)$_2$ and (CCCTAAA)$_2$ using PCR without a template (Ijdo et al., 1991; Cox et al., 1993). Although this family of sequences contains motifs homologous to the repeats of A. thaliana telomeres, in situ hybridization shows extremely weak signal at telomeres (identified only by using computer enhanced quantitative pixel analysis; BDS-Image version 1.5, Oncor, MD), and significantly stronger intensity at non-telomeric sites (Fig. 5). The most common non-telomeric sites were pericentric and interstitial regions of chromosomes. Variation in signal intensity was interpreted as differences in copy number at sites of hybridization group designations should serve as the framework for homologous group designations of chromosomes in related Pinus spp. and possibly for the Pinaceae and other genera of conifers.
Fig. 4. Ideogram of slash pine. Darkest sites and bands show most intense fluorescence, shading less fluorescence and dots least. Band width suggests relative size of site or band. Relative lengths are percentages of the length of the total haploid chromosome complement.

(Schwarzacher and Heslop-Harrison, 1991). Relatively large amounts of the sequence were present within regions of constitutive heterochromatin.

CONCLUSIONS AND FUTURE DIRECTION

The ecologic and economic importance of slash pine have prompted the construction of genetic recombinational linkage maps based principally on RAPDs (Byram et al., 1992; Nelson et al., 1993b; Kubisiak et al., 1995). Genetic mapping efforts summarized, are beginning to reveal details about the organization of coding and adjacent regions and the level of gene duplication within the genomes, thus providing a framework for comparative gene mapping.

Although such genetic maps are informative, they furnish an incomplete picture of genome organization. Molecular characterization and fluorescence in situ hybridization of non-coding sequence probes are providing some information on genomic structure. Synthetic oligonucleotides of tandem arrays of simple sequence repeats were used to detect variation in slash pine and identified polymorphic regions of the genome that distinguished slash pine from related species (A. Kamm, personal communication). A Tyl-copia retroelement was shown to be highly amplified in slash pine and widely dispersed throughout the genome (A. Kamm, personal communication). Considering the large size of the genome of slash pine and the undoubtedly high proportion of noncoding DNA, genetic mapping tasks would be very difficult without the guidance afforded by physical mapping. Physical mapping can help in the generation and use of high-resolution genetic maps by permitting: (i) the mapping of regions experiencing little or no recombination; (ii) comparison of the degree of similarity (and thus application of) genetic maps between different genotypes; (iii) assessment of multilocus probes for use in RFLP mapping; (iv) the ordering of contiguous, large clones such as YAC and cosmid inserts for the
isolation of genes located on RFLP and RAPD linkage maps; and perhaps (v) enable analysis of chromosomal recombination in lines derived from interspecific hybrids.

Dense maps based on DNA markers can be constructed by segregation analysis. However, in high-resolution genetic linkage mapping, the smaller the map distance between loci, the larger the segregating population needed to ensure, at an acceptable level of probability, the proper ordering of loci. Large numbers of segregating individuals from controlled crosses may not be available, because of lack of polymorphisms or the time and space required to generate families, whereas to physically map loci requires seed from only one individual. With in situ hybridization, several probes can be detected independently and ordered along chromosomes in a single experiment, potentially making genome mapping a quicker and more efficient process than segregation analysis (Brown et al., 1993). Combining the two techniques of genetic mapping, Mendelian recombination and cytogenetic localization, holds the most promise for breeders and pathologists to address the immediate problem of fusiform rust disease, and more broadly the genetics of disease resistance genes improving longer-term breeding strategies.

I thank R. C. Schmidttling, T. L. Kubisiak, and A. Kamm for their reviews of earlier versions of this manuscript. I thank J. S. Heslop-Harrison for assistance in preparation of the final manuscript.

REFERENCES


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


