Disclaimer: unless otherwise agreed by the Council of UPOV, only documents that have been adopted by the Council of UPOV and that have not been superseded can represent UPOV policies or guidance.

This document has been scanned from a paper copy and may have some discrepancies from the original document.

Avertissement: sauf si le Conseil de l’UPOV en décide autrement, seuls les documents adoptés par le Conseil de l’UPOV n’ayant pas été remplacés peuvent représenter les principes ou les orientations de l’UPOV.

Ce document a été numérisé à partir d’une copie papier et peut contenir des différences avec le document original.

Allgemeiner Haftungsausschluß: Sofern nicht anders vom Rat der UPOV vereinbart, geben nur Dokumente, die vom Rat der UPOV angenommen und nicht ersetzt wurden, Grundsätze oder eine Anleitung der UPOV wieder.

Dieses Dokument wurde von einer Papierkopie gescannt und könnte Abweichungen vom Originaldokument aufweisen.

Descargo de responsabilidad: salvo que el Consejo de la UPOV decida de otro modo, solo se considerarán documentos de políticas u orientaciones de la UPOV los que hayan sido aprobados por el Consejo de la UPOV y no hayan sido reemplazados.

Este documento ha sido escaneado a partir de una copia en papel y puede que existan divergencias en relación con el documento original.
INTERNATIONAL UNION FOR THE PROTECTION OF NEW VARIETIES OF PLANTS

GENEVA

WORKING GROUP ON BIOCHEMICAL AND MOLECULAR TECHNIQUES
AND DNA-PROFILING IN PARTICULAR

Third Session
Wageningen, Netherlands, September 19 to 21, 1995

THE “RAPD-MEGAGAMETOPHYTE” APPROACH: AN EFFICIENT TECHNIQUE FOR GENETIC ANALYSIS IN CONIFER SPECIES

GENOMIC MAPPING IN PINUS PINASTER (MARITIME PINE) USING RAPD AND PROTEIN MARKERS

GENOMIC ANALYSIS IN MARITIME PINE (PINUS PINASTER) - COMPARISON OF TWO RAPD MAPS USING SELFED AND OPEN-POLLINATED SEEDS OF THE SAME INDIVIDUAL

Document prepared by experts from France
The "RAPD-megagametophyte" approach: an efficient technique for genetic analysis in conifer species

Plomion C, Bahman N, Kremer A
INRA, Station de Recherches Forestieres, BP45, 33610 Cestas, France

Most genetic analysis in plants and animals has been carried out with codominant markers. These types of markers are common in isozymes, RFLPs and many PCR based markers. RAPD markers are dominant and therefore are usually not informative when it is necessary to distinguish heterozygotes from homozygotes. This property has caused many geneticists to avoid RAPD markers. However, during the past four years, genetic analysis with RAPDs has been tremendously enhanced in gymnosperms species. This development is due to a special feature of the reproductive biology of conifers: the haploid megagametophyte tissue in the seeds. Megagametophytes are mitotic derivatives of a single haploid megaspore and are derived from the same megaspore that gives rise to the maternal gamete. Therefore they are genetically equivalent to maternal gametes.

We investigated the use of RAPD markers to construct a genetic map in maritime pine (Pinus pinaster Ait.) by analyzing their segregations in 124 megagametophytes of F2 selfed seeds from a single hybrid individual (Plomion et al. 1995a; 1995b). The dominance mode of inheritance was not an issue for genetic mapping, because informative RAPD markers segregated 1:1 as in a backcross progeny. We carefully screened for RAPD polymorphisms that were repeatable across four replicate sets of 31 independent megagametophytes. A total of 436 RAPD markers chosen on the basis of repeatability, generated a saturated map of the maritime pine genome. These markers segregating in megagametophytes could also be detected in the diploid tissue of both the F1 individual and its two parents, as well as in the F2 progeny. They should be valuable genetic markers for future mapping studies (e.g., QTL analysis in the same pedigree) provided that care is taken to use identical conditions for carrying out the PCR protocols.

RAPD markers provided a fast, efficient and reliable way to construct this map over a period of 6 months. The advantages of the RAPD technique are the requirement for small amount of DNA (5-20ng per PCR), the rapidity to screen for polymorphisms, the efficiency to generate a large number of markers for genomic mapping and the potential automation of the technique. In addition, no prior knowledge of sequence is required. Since primers can be chosen arbitrarily, any individual can be mapped with the same set of primers. These advantages make RAPD markers far easier to work with than RFLP's and thus very attractive for breeding application. Therefore, one large impact of the RAPD technique has been to increase the species amenable to mapping activities. It is particularly true in forest trees where genetic map has been constructed in slash pine, loblolly pine, white spruce, norway spruce, scots pine and douglas-fir.

In the large genome of maritime pine (24 pg/C), we showed that RAPD markers essentially amplified from highly repetitive chromosomal regions that is mostly non coding DNA. The next step of our research is to localize new molecular markers corresponding to coding sequences in the existing "RAPD-megagametophytes-based" map. As an begining, 44 protein loci revealed by two dimensional gel electrophoresis have been incorporated in the map.

References:
Genomic mapping in *Pinus pinaster* (maritime pine) using RAPD and protein markers

C. PLOMION†*, N. BAHRMAN†, C.-E. DUREL† & D. M. O’MALLEY‡
‡INRA, Laboratoire de Génétique et Amélioration des arbres forestiers, BP 45, F-33610 Cestas, France and †Forest Biotechnology Group, Department of Forestry, North Carolina State University, Box 8008 Raleigh, NC 27695, U.S.A.

A detailed genomic map was constructed for one F1 individual of maritime pine, using randomly amplified polymorphic DNA (RAPD) and protein markers scored on megagametophytes of germinated seeds. Proteins allowed the localization of exclusively coding DNA in the large genome of this *Pinus* species, mapped with RAPD markers that essentially fall within repetitive (i.e. mostly noncoding) DNA. Dot blots experiments of 53 RAPD fragments showed that 89 per cent amplified from highly repetitive chromosomal regions. The map comprised 463 loci, including 436 RAPDs amplified from 142 10-mer oligonucleotide primers and 27 protein loci. Twelve major and one minor linkage groups were identified using a LOD score ≥ 5 and a recombination fraction ≥ 0.30. A framework map was ordered with an interval support ~4, covering 1860 cM which provided almost complete coverage of the maritime pine genome. The average distance between two framework markers was 8.3 cM; only one interval was larger than 30 cM. Protein loci were well distributed throughout the map. Their potential use as anchor points to join RAPD-based maps is discussed. Finally, the genomic maps of *Arabidopsis* and maritime pine were compared. Linkage groups were shown to have similar total map lengths on a chromosomal basis, despite a 57-fold difference in DNA content.

**Keywords:** 2-D electrophoresis, linkage map, *Pinus pinaster*, protein, RAPDs.

**Introduction**

The first linkage studies of *Pinus* were based on segregation of isozymes extracted from megagametophytes. More than 10 species were studied for about 15 loci (reviewed by Tulsieram *et al.*, 1992). Conkle (1981) located more loci but the number of markers that were resolved and analysed was still too low for applications that required a broader genome coverage (e.g. quantitative trait dissection studies). Two-dimensional electrophoresis (2D-PAGE) of megagametophyte proteins identified a larger number of loci. Bahrman & Damerval (1989) reported linkage analysis for 119 loci and Gerber *et al.* (1993) reported a 65 loci linkage map covering 530 cM of the maritime pine genome. Both isozymes and proteins correspond to coding DNA. Devey *et al.* (1994) presented linkage groups in loblolly pine for 80 RFLPs detected using cDNA probes. RFLPs typically sample genetic variation in coding regions or directly adjacent to coding regions of the genome, and use low copy probes. The RAPD method (Williams *et al.*, 1990) permits identification of a large number of polymorphic DNA markers distributed throughout the genome, including both coding and noncoding regions (Williams *et al.*, 1990). RAPDs have been used for genomic mapping in several conifer species (Neale & Sederoff, 1991; Tulsieram *et al.*, 1992; Nelson *et al.*, 1993; Binelli & Bucci, 1994). *Pinus* species have a large genome (Ohri & Khoshoo, 1986; Wakamiya *et al.*, 1993) characterized by a high proportion of repetitive DNA (Miksche & Hotta, 1973; Rake *et al.*, 1980; Kriebel, 1985). Therefore, RAPDs and markers that are based on coding sequences could provide different coverage of the genome in pine.

Each type of marker has advantages and limitations and many factors can influence choice of marker systems for a given purpose. Marker based technologies are being used for linkage map construction, quantitative traits dissection experiments, germplasm evaluation, genetic fingerprinting and manipulation of genes. Genomic maps can also be used to locate and clone genes of interest. In addition, they may provide information for understanding genome structure and evolution (Neale & Williams, 1991). RFLP methods...
are well suited for species maps because the same hybridization probes can be used in comparisons among species. Ahuja et al. (1994) showed that mapped DNA probes from loblolly pine can be used to construct RFLP maps for other members of Pinaceae, thus presenting an opportunity to compare the genome in related pine species. Forest trees exhibit generally high levels of genetic diversity and are highly outcrossed. As a result, linkage disequilibrium should be low. Thus, alleles at quantitative trait loci (QTLs) and alleles at marker loci should be randomly associated in different genotypes (Strauss et al., 1992). Therefore the ability to create maps for individual trees, and to assess marker genotypes on hundreds of progenies, is essential for breeding experiments that aim to use marker-assisted selection. The RAPD technology appears to be well suited for developing single-tree maps. The major advantage of this technique is the rapidity of screening for polymorphisms, the identification of a large number of markers and its potential automation (Sobral & Honeycutt, 1993). However, synteny of linkage groups with other species could be difficult to established with RAPD markers, as noticed by Torres et al. (1993).

We located 2-D protein polymorphisms that represent coding DNA on a RAPD-based framework map of maritime pine (Pinus pinaster Ait., 2n = 24). Protein markers were used because (i) they only reveal gene products, (ii) they have been well studied in that species, and (iii) they provide interesting tools, as candidate genes, for genetic dissection of vigour and adaptative traits in the frame of the maritime pine breeding programme. The objectives of this contribution are threefold: (i) genetic map construction based on markers of coding and noncoding chromosomal regions, (ii) estimation of genome size of maritime pine, and (iii) characterization of RAPD fragments’ internal sequences for copy number in the genome. The map and markers will be used for QTL mapping of height growth and adaptative traits in an F₂ progeny of the mapped tree.

Materials and methods

Plant material

The material used was 124 megagametophytes from seeds of one inter-racial hybrid (parent F₁; Corsican × Landes) of maritime pine. Individual megagametophytes were harvested after 2 weeks of germination. Each megagametophyte was cut in two parts, with one-quarter freeze-dried for RAPD assay, and three-quarters stored at -80°C and devoted to protein analysis.

DNA extraction and RAPD procedure

DNA samples were prepared from needles of the Corsican and Landes grandparents (accessions C10 and L146, respectively), and the F₁ hybrid parent (accession H12), as well as from the megagametophytes from H12. DNA was extracted using the CTAB procedure described by Bousquet et al. (1990). The DNA extracted from needle samples was purified by centrifugation in a CsCl-ethidium bromide density gradient. The pine DNA was diluted to a working concentration of approximately 1 ng/μL, by comparison with the fluorescence of lambda DNA concentration standards on ethidium bromide-stained agarose gel. RAPD reactions were performed using the method of Williams et al. (1990) with 5–10 ng template; 20 ng of ten-base primers from Operon Technologies (Alameda, CA), kits A–Z; 8 μg/μL non-acetylated bovine serum albumin and 1 unit of Taq DNA polymerase. The mixture was covered with 50 μL of mineral oil and amplifications were carried out in 96-well microtitre plates using an MJ Research PT-100 thermal cycler (MJ Research, Watertown, MA). Amplification products were separated by electrophoresis on 2 per cent agarose gels, detected by staining with ethidium bromide, and the gels were silver-stained according to Damerval et al. (1993). The gels were silver-stained for 1 h with one-quarter freeze-dried for RAPD assay, and

Protein extraction and electrophoresis

A sample of 34 megagametophytes was individually crushed in 6 μL/mg UKS buffer (9.5 M urea, 5 mM K₂CO₃, 1.25 per cent SDS, 0.5 per cent dithiothreitol, 2 per cent pharmalyte pH 3–10 and 6 per cent Triton X-100). Thirty-five microlitres of the supernatant was submitted to electrofocusing in the first dimension, followed by a second dimension electrophoresis (Bahram & Thiellement, 1987). The gels were silver-stained according to Damerval et al. (1987) in the

apparatus described by Granier & de Vienne (1986) and dried.

**Scoring and nomenclature of RAPD and protein markers**

Segregation of RAPD markers was recorded in four sets of 31 different megagametophytes from H12. DNA extraction, reaction mixture preparations, gel analysis and genotype scoring were performed independently for each set. This replicated design provided a control that aimed to retain RAPD markers that amplified consistently in the studied population. RAPD fragments were named by the OPERON primer code, followed by their molecular size in base pairs. Because protein analysis requires elaborate laboratory techniques, protein markers were recorded for only 34 randomly chosen megagametophytes among the mapping sample of 124 megagametophytes. The dried gels were visually scored on an illuminated box by superimposition. Three kinds of variations were scored in 2-D protein patterns: position (V), presence/absence (P) and staining intensity (I) variations. The name of each marker included the grandparental origin ('+' denoted markers inherited from the Corsican grandparent, '−' denoted markers inherited from the Landes grandparent).

**Linkage analysis**

A total of 463 genetic markers were tested for departure from the 1:1 Mendelian ratio of presence/absence of band. The linkage relationships of the markers were analysed with the Macintosh MAPMAKER v2.0 computer program (Lander et al., 1987). Markers were considered to be linked when their LOD score was ≥ 5.0 and recombination fraction $\Theta \leq 0.30$. A subset of markers that could be ordered with an interval support ≥ 4 (i.e., difference in log likelihood between the best and alternative orders ≥ 4), provided a framework map. Accessory markers that could not meet this ordering criterion were located to the closest framework markers. Recombination distances of accessory markers to the nearest framework markers were incorporated in the marker names. Recombination fractions were converted to map distances using Kosambi's mapping function.

**Results**

**Identification of polymorphic markers**

We scored 35 protein markers in 2-D protein patterns obtained from germinated megagametophytes. Twenty spots belonging to 10 polypeptides corresponded to allelic products of structural genes varying in position, 22 spots concerned presence/absence variations and three spots involved staining intensity variations (see Bahrman & Damerval, 1989; Gerber et al., 1993, for genetic analysis of each variation). They all segregated in a Mendelian fashion ($\alpha = 0.01$). From 520 OPERON primers screened for polymorphisms, 142 showed amplification and segregating RAPD markers present in one grandparent as well as in the hybrid parent and absent in the other grandparent. Out of the 142 primers, 113 and 29 were used to produce RAPD markers on four and two replicates of 31 megagametophytes, respectively. Segregation of 470 RAPD markers was scored in the whole experiment (Fig. 1). A total of 437 RAPDs were repeatable among the mapping replications. On average, one primer produced three polymorphic RAPD fragments. Fragment sizes ranged from 194 to 2627 base pairs. They all conformed to Mendelian segregation ($\alpha = 0.01$). The nonreproducible RAPD bands were discarded from further analysis. They were typically very faint, and often had a molecular weight > 2000 bp or < 200 bp.

**Construction of the genomic map**

Grouping and initial ordering of markers were carried out at LOD ≥ 5.0 and $\Theta \leq 0.30$. Out of 471 markers (436 RAPDs and 36 proteins), 463 loci (436 RAPDs and 27 proteins) were assigned to 13 linkage groups. Eight of the protein markers were not linked to any other locus when lowering the statistical stringency. However, lowering the LOD score to 3.0 and keeping $\Theta$ to 0.30 would result in the merging of linkage group 1 and 13 as indicated by a faint line between markers R10_767/− and O18_1207/+ (Fig. 2). Few additional mapped markers would be needed to fill the gap between these two groups. Local mapping based techniques (Reiter et al., 1992) should facilitate this objective. The number of major groups corresponded to the 12 expected based on the known karyotype of maritime pine (Saylor, 1964).

A framework map was established using the RIPPLE command, to identify a subset of loci that could be locally ordered with an interval support ≥ 4. Approximately 53 per cent of the markers were placed on the framework map defining a total of 244 loci and 1860 cM of map distance. The size of linkage groups ranged from 177.9 cM to 16.6 cM. The average distance between two framework markers was 8.3 cM, with only a few gaps exceeding 20 cM. Only one interval between two markers located in linkage group 5 was larger than 30 cM. However, the LOD score for this interval was above the threshold 5. The majority of the intervals (72 per cent) were < 10 cM. Most of the accessory markers were placed within 5 cM of the nearest framework
marker. The limited number of megagametophytes did not allow a precise estimate of the recombination distances and ordering of tightly linked markers. In a sample of 62 or 124 megagametophytes, approximately 95 per cent of the gametes will show no recombination in a 5 cM interval and provide no information on order. Assuming the same error rate in genotyping, we also observed that the number of accessory RAPD markers placed at distances 6 cM from the nearest framework marker was not much higher for 62 (two sets of 31 megagametophytes) than for 124 megagametophytes (four sets): 22 per cent and 18 per cent amplified from samples of 62 and 124 megagametophytes, respectively. Thus, there was little increase in precision on the relative position of accessory markers when genotyping 124 individuals instead of 62.

Assuming that the order of framework markers was correct, we used the show raw command to identify double recombinants that involved flanking loci. True double recombination events should be rare, and an excess of double recombination could indicate potential scoring errors. Double crossovers were systematically re-examined on gel photos. When dubious data points were found they were treated as missing data. Then the ordering analysis was performed again. Errors in genotyping were mostly from weak amplification of a specific band, smearing problems or artefacts from the loading of wells in the gel. This data quality control, in combination with the framework map construction procedure, should provide a high confidence for map length and loci order.

The 27 mapped protein loci were well distributed throughout the genomic map (Fig. 2). A total of four, three, two and one protein loci were mapped in two, two, five and three linkage groups, respectively. The three types of proteins (see Materials and methods) were represented on the map. Protein loci were genotyped on only 34 megagametophytes, which did not allow a precise estimation of two-point recombination fractions. This could explain the high proportion of unlinked protein markers (23 per cent) and the fact that almost all proteins could not meet the local order criterion used for framework map construction (interval support ≥ 4). Therefore, most protein loci were placed as accessory markers.

Copy number of RAPD fragments

Out of 53 RAPD fragments, 11 per cent did not show detectable hybridization or gave a faint signal in the 20 µg dilution. They were classified as amplifying from low-copy to moderately repetitive chromosomal regions (Fig. 3a,b). Twenty per cent gave a signal in the 20 µg and 2 µg dilutions and were classified as amplifying from highly repetitive regions (Fig. 3c); 69 per cent gave a signal in all dilutions and were classified as amplifying from very highly repeated regions (Fig. 3d,e). So, 89 per cent of the RAPD fragments were

Fig. 2  Linkage map of maritime pine 'H12' hybrid. Loci are listed on the right (named according to the text) and recombination distances (cM) are listed on the left of each linkage group. Markers were grouped with a LOD ≥ 5 and θ ≤ 0.30. Framework markers have been ordered with an interval support ≥ 4. Accessory markers that could not be ordered with equal confidence (interval support < 4) are listed on the right side of the framework markers. Protein loci are boxed and indicated by arrows.
Thus, coding DNA may not represent more than a few per cent of the pine genome. This result agrees with the average size of a gene of 1981). If the very high and consistent amount of DNA per nucleus (Ohri & Khosho, 1986; Wakamiya et al., 1993). The technique of DNA reassociation kinetics applied to Pinus species (Mikshe & Hotta, 1973; Rake et al., 1980; Kriebel, 1985) showed that 25 per cent of total DNA is low- to single-copy, 75 per cent being middle to highly repetitive. Thus, the vast majority of the DNA in the pine genomes is arranged in repeated sequence families. Most of this repeated DNA does not encode proteins (Thompson & Murray, 1981). If 60 000 genes are expressed during the life cycle of a plant (Kamalay & Goldberg, 1980), given an average size of a gene of 2000 bp (exons only) and the size of maritime pine genome as $24 \times 10^8$ kbp, 0.5 per cent of the genome is likely to be coding DNA. This could be an underestimate because a significant number of genes occur in multigene families (Kinlaw & Gerttula, 1993; Ahuja et al., 1994; Devey et al., 1994). Thus, coding DNA may not represent more than a few per cent of the pine genome. This result agrees with the estimated fraction of coding regions in plant species (Goldberg et al., 1978; Thompson & Murray, 1981).

The RAPD primers used for mapping consisted of random sequences that should not discriminate coding and noncoding chromosomal regions. Therefore, and at least in conifer species, most RAPD loci are likely to fall within noncoding DNA. The characterization of the internal sequence of 53 RAPD fragments for copy number in the maritime pine genome showed that, although RAPD fragments mapped to unique genomic sites, most of them contained highly repeated sequences. Conversely, protein markers sample regions of coding DNA. Our results showed that mapped protein markers were well distributed throughout the genome of maritime pine.

**Discussion**

**Pine genome organization**

Gymnosperm species are characterized by: (i) their antiquity (conifers appeared 140 millions years before the first angiosperm) and their longevity, (ii) the absence of ploidy level and chromosome number evolution (reviewed by Neale & Williams, 1991), and (iii) the very high and consistent amount of DNA per nucleus (Ohri & Khosho, 1986; Wakamiya et al., 1993). The technique of DNA reassociation kinetics applied to Pinus species (Mikshe & Hotta, 1973; Rake et al., 1980; Kriebel, 1985) showed that 25 per cent of total DNA is low- to single-copy, 75 per cent being middle to highly repetitive. Thus, the vast majority of the DNA in the pine genomes is arranged in repeated sequence families. Most of this repeated DNA does not encode proteins (Thompson & Murray, 1981). If 60 000 genes are expressed during the life cycle of a plant (Kamalay & Goldberg, 1980), given an average size of a gene of 2000 bp (exons only) and the size of maritime pine genome as $24 \times 10^8$ kbp, 0.5 per cent of the genome is likely to be coding DNA. This could be an underestimate because a significant number of genes occur in multigene families (Kinlaw & Gerttula, 1993; Ahuja et al., 1994; Devey et al., 1994). Thus, coding DNA may not represent more than a few per cent of the pine genome. This result agrees with the estimated fraction of coding regions in plant species (Goldberg et al., 1978; Thompson & Murray, 1981).

A species consensus map of markers and traits could be difficult to use for breeding applications in allogamous species with a wide genetic base, such as forest trees (Grattapaglia & Sederoff, 1994). Marker:trait associations are likely to be in linkage equilibrium in early generations of the breeding population and will probably have to be established for each cross independently. Mapping of individual trees using markers specific to only one cross provides a powerful approach to genetic analysis of quantitative and complex qualitative traits within families. However, genomic maps of individuals using RAPD markers can not readily be combined to make a consensus species map because the migration distance of a RAPD fragment is not sufficient information to identify uniquely a specific locus across a species. Similar problems exist for RFLP probes that recognize several bands (e.g. Tanksley et al., 1988; Song et al., 1991; Devey et al., 1994), a problem addressed by using probes that yield only one band (e.g. Beavis & Grant, 1991). Furthermore, many individuals could be homozygous and the marker would not be available for mapping in many crosses, depending on gene frequency. The criteria for establishing synteny using RAPD markers, or multiple band RFLP markers must be more stringent, perhaps requiring parallel linkage groups having several markers in the same order in different individuals. The identity of some allozyme or protein markers (Gerber et al., 1993) should be useful for establishing the correspondence of linkage groups in RAPD maps from different trees. The distribution of the 27 mapped protein loci throughout the genomic map of maritime pine is encouraging for that objective. A further advantage to using proteins as genetic markers for the mapping of quantitative or qualitative traits is that the polymorphism of a specific gene product could potentially be responsible for the mapped quantitative effect (Damerval et al., 1994). Alternatively, a small number of hypervariable microsatellite markers could be...
assayed in each cross to establish the correspondence of linkage groups.

**Genome size of maritime pine**

The protein data produced by Bahrman & Damerval (1989) and Gerber et al. (1993) suggested a genome size of approximately 2000 cM for maritime pine (Gerber & Rodophe, 1994). However, framework map procedures had not been used to construct these two protein-based maps. This may lead to overestimates for genetic distances and total map length. Maritime pine has 12 metacentric chromosomes of approximately equal size (Saylor, 1964). Linkage groups 1-12 had approximately the same length (about 155 cM) and therefore should provide almost complete coverage of the genome. In addition, the genome size of the presented framework map (1860 cM) agreed with what has been found for other dense linkage maps of loblolly pine, constructed with approximately 400 RAPD markers (H. Amerson & P. Wilcox, personal communication). The relationship between recombination rates and genome size has been a matter of speculation for many years. Grant (1958) predicted that plants with long generation times, such as pine, will have genetic systems that promote recombination. Short-lived annual plants, such as Arabidopsis, should have genetic systems that restrict recombination. One mechanism to promote recombination could be an increased number of chromosomes. Grant (1958) also speculated that long-lived organisms such as pine might have a higher chiasma frequency to promote recombination. Our data for pine, however, do not support this idea. The total map distance per chromosome was approximately 1.55 Morgans for pine and 1.30 Morgans for Arabidopsis (Reiter et al., 1992). Maritime pine and Arabidopsis have approximately 2 pg and 0.03 pg of DNA per chromosome, respectively (Ohri & Khosoo, 1986; Arumuganathan & Earle, 1991). Thus, on a chromosomal basis, maritime pine has approximately 57-fold more DNA per cM than Arabidopsis. However, the number of crossovers per chromosome was almost equivalent and did not seem to be really affected by the DNA content and the proportion of coding DNA. Although large and small genomes could differ in the organization and structure of genomic DNA (John & King, 1980; Flavell et al., 1985; Brown & Sundaresan, 1991), the mechanism of crossing-over must be highly conserved on a chromosomal basis and independent of physical map size and the fraction of coding DNA. This observation is consistent with other results showing that recombination per chromosome was approximately constant despite large differences in DNA amount (Rees & Durrant, 1986; Tanksley et al., 1988).

**Acknowledgements**

The RAPD assay was completed in the Forest Biotechnology Group at North Carolina State University. We thank Dario Grattapaglia for assistance with the characterization of genomic sequence complexity of RAPD markers and spending much time discussing genetics. We thank Antoine Kremer for critical reading of the manuscript. This work was supported by a grant from the Institut National de la Recherche Agronomique (AIP Génome des plantes cultivées 1994) and from the USDA Forest Service (Coop Agreement No 29-1009).

**References**


C. Plomion · D. M. O’Malley · C. E. Durel

Genomic analysis in maritime pine (Pinus pinaster). Comparison of two RAPD maps using selfed and open-pollinated seeds of the same individual

Abstract Two genomic maps were constructed for one individual tree of maritime pine, Pinus pinaster Ait., using a common set of 263 RAPD markers (random amplified polymorphic DNA). The RAPD markers were chosen from a larger number of polymorphic RAPD fragments on the basis of repeatability and inheritance in a three-generation pedigree. The maps were constructed from two independent mapping samples of 62 megagametophytes (In) from a self cross and from an open-pollinated cross. The markers were grouped (LOD≥4: θ≤0.25) and assigned to 13 major and 5 minor linkage groups. Two framework maps were constructed using the ordering criterion of interval support≥3. Comparison of the two framework maps suggested that the locus order was incorrect for 2% of the framework markers. A bootstrap analysis showed that this error rate was representative for our data set. The results showed that framework maps constructed using RAPD markers were repeatable and that differences in locus order for maps of different genotypes or species could result from chance. The total map distance was 1380 cM and the map provided coverage of approximately 90% of the genome.

Key words RAPD markers · Linkage map · Map comparison · Locus ordering · Pinus pinaster

Introduction

Comparison of genomic maps from different species or genera can provide insight on plant evolution and genome structure (Tanksley et al. 1988; Bonierbale et al. 1988; Gebhardt et al. 1991; Tanksley et al. 1992; Whitkus et al. 1992). In allogamous plant species, map-based comparison of segregation distortion from different crosses of the same individual genotype could reveal chromosomal segments that contain genes affecting fertilization or viability (Gebhardt et al. 1991; Bradshaw and Stettler 1994). Because the same locus order is expected, the comparison of maps from the same individual genotype also provides a way to evaluate the repeatability of genomic map construction. Experimental comparison of maps from the same individual genotype or different genotypes within species has only rarely been reported (e.g., Beavis and Grant 1991).

Two maps from the same individual genotype should closely resemble each other if the experimental methods used to produce the genetic markers and the statistical methods for constructing the genomic maps are sufficiently rigorous (i.e., repeatable). The comparison of genomic maps depends upon the accurate determination of locus order. The “ordering problem” is difficult because of the large number of possible locus orders (n!/2 for n loci) and because customary likelihood ratio tests cannot be carried out (reviewed by Ott 1991). Framework maps are constructed for a chosen subset of markers ordered with an interval support≥3, a widely employed criterion (e.g., Reiter et al. 1992; Kesseli et al. 1994; Grattapaglia and Sederoff 1994) recommended by Keats et al. (1991). Interval support is obtained by subtracting the log likelihood for the linkage group with the best locus order from the log likelihood for the same linkage group with a different local order (usually alternative permutations of three adjacent markers). The order for framework loci should be more certain than the order of closely linked loci on a comprehensive genomic map (no local support criteria for locus order), but framework locus order probably depends on the grouping criteria and the algorithm used.

Here we report a comparison of two maps that were constructed for one individual of maritime pine (Pinus pinaster Ait. 2x=2n=12) using RAPD (random amplified polymorphic DNA) markers (Williams et al. 1990; Welsh and McClelland 1990). The only genetic markers known in this
species before this study were proteins revealed by two-dimensional gel electrophoresis (Bahrman and Damerval 1989; Gerber et al. 1993). We used a replicated design involving four sets of 31 different individuals to choose RAPD markers that were highly repeatable and easily scored in megagametophytes. A map was constructed with two sets of 31 megagametophytes from a self family (SELF map) and compared to a map constructed with two sets of 31 megagametophytes from an open-pollinated family (OP map). Conifers are believed to have a large number of recessive embryonic lethal genes that could result in segregation distortion (Sorensen 1967; Strauss and Conkle 1986). The comparison of two maps constructed for the same individual provided an opportunity to test the reliability of markers and the robustness of the linkage groupings, as well as to screen for segregation distortion. The individual that was mapped (twice) is part of a breeding program for genetic improvement of maritime pine in France and is an F1 hybrid between the Landes and Corsican races. The genetic markers could be used to introgress stem straightness and good branching habit from the Corsican race into the widely planted Landes race if these traits are oligogenic.

Materials and methods

Plant material and DNA extraction

DNA samples were prepared from needles of the Corsican and Landes grandparents (accessions CI0 and L146, respectively) and the inter-racial hybrid parent (accession H12), as well as from the megagametophytes of selfed and open-pollinated seeds from H12. The seeds were germinated following standard methods. After emergence and just before the seed coat was cast off, the megagametophyte was collected from the seedling and freeze-dried or stored at -80°C. Megagametophyte tissue frozen in liquid nitrogen was ground to a fine powder in a 1.5-ml microfuge tube. Freeze-dried needles (4 g) of both grandparents and the hybrid parent were ground under liquid nitrogen using a prechilled mortar and pestle and transferred to 1.5-ml microfuge tubes. DNA was then extracted using the CTAB method of Doyle and Doyle (1987). The DNA extracted from these older needle samples was purified further by centrifugation in a CsCl-ethidium bromide density gradient. The pine DNA was diluted to a working concentration of approximately 1 μg/μl by comparison with the fluorescence of lambda DNA concentration standards on an ethidium bromide-stained agarose gel.

DNA amplification by polymerase chain reaction (PCR)

The method of Williams et al. (1990) was used to PCR-amplify polymorphic DNA fragments to be used as genetic markers. The volume of the reaction mixture was 15 μl and contained 8 mg/ml non-acetylated bovine serum albumin. The mixture was covered with 50 μl of mineral oil, and amplification were carried out in 96-well microtitre plates using a MJ Research PT-100 thermal cycler (MJ Research, Watertown, Mass.). The DNA fragments were separated by standard electrophoretic methods on 2% horizontal agarose TBE gels. The DNA fragments were separated by standard electrophoretic methods on 2% horizontal agarose TBE gels (Watertown, Mass.). The DNA fragments were separated by standard electrophoretic methods on 2% horizontal agarose TBE gels (Watertown, Mass.). The DNA fragments were separated by standard electrophoretic methods on 2% horizontal agarose TBE gels (Watertown, Mass.). The DNA fragments were separated by standard electrophoretic methods on 2% horizontal agarose TBE gels (Watertown, Mass.).

Identification of RAPD markers

RAPD polymorphisms that are good genetic markers should be easily repeatable in PCR amplification reactions carried out on different days. We carefully screened for RAPD polymorphisms that were repeatable across four replicate sets of 31 independent megagametophytes. Two sets comprised the SELF mapping sample, and two sets comprised the OP mapping sample. This replicated design allowed us to choose RAPD polymorphisms that were repeatable across replicates within and between mapping samples. Some markers had co-migrating bands or were difficult to classify as presence or absence in some of the replicates. Each photo was scored twice and the individual phenotypes compared. When two scores disagreed, the lane was scored as missing for that sample.

Linkage analysis of RAPD markers

The linkage relationships of the markers were analyzed with MAP-MAKER (Lander et al. 1987) version 2.0 for the Macintosh provided by S. Tingey (DuPont, Wilmington, De.). The genetic model for conifer megagametophyte segregation data for individual trees is analogous to a testcross with the parental linkage phase unknown (O'Malley et al. 1986). The MAPMAKER Macintosh HAPLOID model assumes that all markers are in the coupling phase and consequently does not recognize linkages for markers in repulsion. The assignment of coupling and repulsion phases is arbitrary for a testcross model, and repulsion phase linkages can be detected by analyzing recoded data (i.e., presence recoded to absence, and vice versa) together with the original data set. Analysis of the combined data yielded twice the expected number of linkage groups, corresponding to the two homologs for each chromosome. The two homologous groups contained the same markers in the same exact order. Markers were assigned to linkage groups using a LOD ≥ 4.0 and recombination fraction of 0 ≤ 0.25. The order of the markers was approximated using FIRST ORDER (a matrix correlation procedure).

Framework maps were then constructed by comparing the likelihood of all permutations of all adjacent triplets using RAPD-1. Individual markers were dropped from each linkage group until a marker sequence was obtained that had an order at least 1000 times better than other orders (i.e., log likelihood difference ≥ 3). The markers that were dropped were placed on the framework map as accessory markers and located to the closest framework markers (Fig. 1). Recombination fractions were converted to map distances using the Kosambi mapping function.

Comparative mapping between the SELF and the OP maps

Statistical evaluation of differences in locus order is intractable due to the large number of possible orders when more than a few loci are considered. A comparison of two genomic maps therefore must assume loci that have the same locus order on both maps. Homogeneity of individual two-point recombination fraction estimates was tested using a G-statistic approach implemented in GEMINDEL 2.0 (Liu and Knapp 1992). This test can be expressed as follows: $G_{\text{homogeneity}} = G_{\text{SEL}} + G_{\text{OP}} = G_{\text{POOL}}$, where the G-statistics are $1 df$ tests for the independent assortment of a pair of markers in the SELF, OP, and POOL maps. For a global test that combined all intervals (i.e., summed the G-statistics), P values were approximated following Beavis and Grant (1991) and Lander and Botstein (1989). The significance level on the whole experiment is approximated by $\alpha = 1 - (1 - \alpha')^n$, where $\alpha'$ is the nominal significance level for each interval, and $n$ is the number of intervals.
Identification and inheritance of RAPD markers

The 102 primers that revealed polymorphisms among the two grandparents and the F₁ parent were used to amplify DNA fragments from megagametophytes of the F₁ individual. RAPD reactions yielded a total of 374 DNA fragments that showed polymorphisms in at least one of the four replicate sets of 31 different individuals. RAPD fragments that amplified in only one replicate set were dropped from further analysis. Nonrepeatable polymorphisms were typically faint bands and had a molecular weight of more than 2000 bp or less than 200 bp. There were 303 RAPD polymorphisms scored in both replicate sets of the SELF mapping sample and 289 in both replicate sets of the OP mapping sample. Some of the RAPD polymorphisms were repeatable only in the SELF mapping sample and some were repeatable only in the OP mapping sample, but most (263) were repeatable between the two mapping samples and these were used as markers for mapping. The similarity index (Sorensen 1948) for the two lists of repeatable polymorphisms for the SELF and OP mapping samples was 88.3. The screening of the grandparents and the F₁ parent yielded only 146 candidate polymorphisms, but more polymorphisms were detected from the segregation analysis because heterozygous and homozygous dominant grandparental phenotypes could not be distinguished.

The segregation ratio of most RAPD polymorphisms did not depart significantly from 1:1, the expected Mendelian ratio in megagametophytes. Polymorphisms that showed the strongest departures from the 1:1 segregation (P < 0.002) were later shown to involve co-migrating polymorphic bands. There were six such cases specific to the SELF mapping sample, and two cases specific to the OP mapping sample. From the 263 repeatable polymorphisms identified as genetic markers common to both SELF and OP mapping samples, 7 markers out of 526 showed significant departure from the 1:1 segregation (0.002 < P < 0.01). The number of departures from 1:1 was close to that expected due to chance, and the departures did not repeat between the two mapping samples. RAPD fragments ranged in size from 194 bp to 2326 bp, with an average of 874 ± 416 bp. There were 7 putative codominant markers. The 102 primers identified by screening ultimately yielded 2.6 markers per primer.

Results

Screening for RAPD polymorphisms

RAPD polymorphisms that should segregate in the megagametophytes of the F₁ hybrid individual were identified by screening with genomic DNA samples taken from needles of the two grandparents and the F₁ individual. RAPD fragments that were present in only one of the two grand-

parents and also present in the F₁ should be coded by a heterozygous locus in the F₁. Of the 520 oligonucleotide primers that were screened, 31 (6.0%) failed to amplify any DNA fragments, 387 (74.4%) did not yield any polymorphisms and 102 (19.6%) amplified at least 1 scorable polymorphism (146 polymorphisms in total).

Linkage analysis and locus ordering

Grouping and ordering of markers were carried out using a LOD > 4.0 and θ < 0.25. Of 263 markers, 251 markers were assigned to 13 large linkage groups (Fig. 2A), with the remaining 12 assigned to five doublets and triplets (not
When the grouping criteria were relaxed (LOD ≥4.0, θ≤0.30), groups 3.a and 3.b in the OP map were joined (indicated by a faint line on Fig. 2A), as they were already in the SELF map. Group 5 in the OP map was divided into groups 5.a and 5.b in the SELF map. The grouping criteria would have to be relaxed to LOD≤1.5 and θ≥0.40 to join 5.a and 5.b. The linkage group assignments were generally stable for 3≤LOD≤6, with θ≤0.25. Using the RIPPLE command, we constructed framework maps for the SELF (173 markers) and OP (152 markers) mapping samples using an interval support ≥3. Tightly linked markers were dropped one by one until the framework criterion was met. These “accessory” markers were generally located 0≤5 cM from the closest framework marker. The locus order of the framework maps obtained with MAPMAKER 2.0 (matrix correlation method) and with GMEN-DEL 2.0 (simulated annealing, θ≤0.25 and P≤0.0001) were almost identical. The exceptions were permutations of closely linked markers.

For the framework maps, the average spacing between markers was 9.0±5.8 cM in the SELF and 10.3±6.3 cM in the OP, with a maximum gap between consecutive markers of 26.1 cM and 25.9 cM, respectively. The size of the large linkage groups ranged from 33.1 cM to 183.2 cM, with an average size of approximately 96 cM. The mean number of markers per 20- or 25-cM interval was 3.5 and 4.2, respectively, including both accessory and framework markers. A chi-square (5 df) goodness-of-fit test for departure from a Poisson distribution provided no evidence that the markers were clustered (P<0.35).

Homogeneity of recombination fraction

Homogeneity of recombination fraction was tested for 2 marker pairs adjacent to an apparent break between linkage groups 5.a and 5.b that occurred in the OP map but not in the SELF map (Fig. 1). The pairwise combinations of the 4 markers (G4_838/− and P10_324/+ against C1_1363/− and G10_626/+ ) were tested for departure from homogeneity among the SELF and OP mapping samples
TABLE 1  G-statistics test (df=1) for homogeneity of recombination fraction among the selfed (SELF) and open-pollinated (OP) mapping samples for four RAPD markers flanking a break in linkage group 5 (G_{SELF} and G_{OP} G-statistics for linkage for the SELF and OP mapping samples, respectively. G_{homo}, G-statistics for homogeneity, K, recombination fraction)

<table>
<thead>
<tr>
<th>RAPD markers</th>
<th>P10_324</th>
<th>C1_1363</th>
<th>G10_626</th>
</tr>
</thead>
<tbody>
<tr>
<td>G4_838</td>
<td>G_{SELF} 38.40***</td>
<td>0.40</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>R        1.94</td>
<td>45.88</td>
<td>56.65</td>
</tr>
<tr>
<td></td>
<td>G_{OP}   61.00***</td>
<td>29.40***</td>
<td>25.80***</td>
</tr>
<tr>
<td></td>
<td>R        0.00</td>
<td>14.99</td>
<td>16.96</td>
</tr>
<tr>
<td></td>
<td>G_{homo} 1.20</td>
<td>11.50***</td>
<td>8.20**</td>
</tr>
<tr>
<td>P10_324</td>
<td>G_{SELF} 2.80</td>
<td>1.70</td>
<td></td>
</tr>
<tr>
<td></td>
<td>R        59.01</td>
<td>41.67</td>
<td></td>
</tr>
<tr>
<td></td>
<td>G_{OP}   33.20***</td>
<td>26.70***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>R        14.76</td>
<td>16.66</td>
<td></td>
</tr>
<tr>
<td></td>
<td>G_{homo} 8.40**</td>
<td>7.90**</td>
<td></td>
</tr>
<tr>
<td>C1_1363</td>
<td>G_{SELF} 49.60***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>R        4.92</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>G_{OP}   55.00***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>R        1.70</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>G_{homo} 0.10</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

** Significant at the 0.005 level; *** Significant at the 0.001 level

(Table 1). These markers segregated in Mendelian proportions in both populations and had no missing data. A significant departure from homogeneity (P<0.005) among OP and SELF was observed between marker pairs flanking the break. By means of the G-statistics provided in GENEMAP (Kesseli et al. 1994), the data quality for the RAPD framework maps appeared to be high; the observed and expected numbers of "double recombinants" were similar. Misclassification of band phenotypes results in apparent "double recombination" events (discussed by Ott 1991). Assuming the order of framework markers was correct, we estimated the frequency of double recombinants by multiplying together the recombination fraction for adjacent intervals on the framework maps. For a sample of 2 large linkage groups, there were no significant differences between the observed and expected numbers of apparent double recombinants, and none of them had ambiguous band phenotypes.

The RAPD markers that we identified showed few cases of segregation distortion in the SELF and OP mapping samples, so there was no evidence for genetic load in the F1 individual.

Discussion

Repeatability, quality, and segregation of RAPD markers

RAPD markers provided a fast, efficient, and reliable way to construct genomic maps in maritime pine. The two genomic maps were constructed over a period of 6 months for an individual tree using megagametophytes from open-pollinated seeds (OP map) and from selfed seeds (SELF map). RAPD markers for genomic mapping were chosen on the basis of repeatability, inheritance, and expression using genomic DNA from needles. The markers segregating in megagametophytes could be detected in the diploid tissue of both the F1 individual and its two parents, as well as in the F2 progeny, with a few exceptions. These RAPD polymorphisms should be valuable genetic markers for future mapping studies provided that care is taken to use identical conditions for carrying out the PCR protocols, as noted by Penner et al. (1993).

There were no unusual mapping problems for the RAPD markers in maritime pine, in contrast to those described in lettuce by Kesseli et al. (1994). The data quality for the RAPD framework maps appeared to be high; the observed and expected numbers of "double recombinants" were similar. Misclassification of band phenotypes results in apparent "double recombination" events (discussed by Ott 1991). Assuming the order of framework markers was correct, we estimated the frequency of double recombinants by multiplying together the recombination fraction for adjacent intervals on the framework maps. For a sample of 2 large linkage groups, there were no significant differences between the observed and expected numbers of apparent double recombinants, and none of them had ambiguous band phenotypes.

The RAPD markers that we identified showed few cases of segregation distortion in the SELF and OP mapping samples, so there was no evidence for genetic load in the F1 individual.

Framework map comparison

Tests for homogeneity of recombination fraction over all markers (for regions where locus order was identical) did not reveal evidence of heterogeneity. One marker pair, however, showed a significant departure from independent assortment (i.e., linkage) for one map, but not for the other map. Homogeneity of recombination fraction was rejected for this pair of markers and resulted in a large linkage group in the OP map being "broken" in the SELF map (Fig. 1). The analysis of an additional 40 megagametophytes for this marker pair confirmed independent assortment in the SELF and linkage in the OP, suggesting that the original result of no linkage for the SELF mapping sample was not spurious. We could not find a biological explanation for an increased recombination fraction in the SELF versus the OP map for that marker pair.

The locus order for the two maritime pine framework maps was different for 7 out of 129 direct comparisons.

Genome size estimation

The total map distance was estimated following Hulbert et al. (1987). For the maritime pine marker data, the number of informative meioses per map was 62. The number of framework loci was n=173 and n=152 for the SELF and OP maps, respectively. Linked markers were determined by a minimum LOD threshold of T=5.0 and a recombination fraction of θ=0.25. According to Chakravarti et al. (1991), we set the parameter X of Hulbert et al. (1987) to the maximum cM distance between linked markers: 26.1 cM and 25.9 cM for the SELF and OP map, respectively. The TWO-POINT command of MAPMAKER was used to determine the number K of informative marker pairs each within X cM and linked with LOD score≥2. The total distances estimated by this method were 1336 cM (K=581) for the SELF map and 1357 cM (K=438) for the OP map. The same calculation taking into account the 263 mapped markers gave an estimate of 1223 cM (K=1470) and 1236 cM (K=1444) for the SELF and OP map, respectively.
The maritime pine RAPD maps were constructed for the genetic analysis of quantitative traits, thus the accuracy of locus order and the marker density is adequate for our purposes (Darvasi et al. 1993). A 2% error rate could be important for map comparisons where differences in locus order suggest genetic rearrangements. For example, in humans, Higgins et al. (1990) reported a case where the order determined by physical mapping was different from the order inferred by genetic mapping. For a comparison of genomic maps of maize and sorghum, Whitkus et al. (1992) attributed 9 out of 14 differences in locus order to chromosomal rearrangements, with the remaining 5 cases (3% error rate) attributed to uncertainties in locus ordering methods. Thus, our result confirms their suspicion that a small number of locus order differences should be expected by chance due to map construction methods. Accurate locus order is also important for gene isolation by map-based cloning. A better quality of locus order could have been obtained either by increasing the sample size while holding the number of framework markers at approximately 150, or by holding the sample size at 62 and raising the interval support criterion for choosing framework loci.

Genomic analysis in maritime pine

In forestry, restriction fragment length polymorphism (RFLP) markers have been used for studies of genome structure and evolution as well as for analysis of quantitative genetic variation (Neale and Williams 1991; Groover et al. 1994; Devey et al. 1994). Forest trees are generally genetically heterogeneous and highly outcrossed, thus anonymous markers such as RAPD can be readily detected despite their dominant pattern of inheritance. RAPD markers are an efficient first step towards establishing a genomic map for previously unstudied species (Tulsieram et al. 1992; Nelson et al. 1993; Grattapaglia and Sederoff 1994). Our result demonstrates that a high quality genomic map that covers 90% of the genome can be constructed from RAPD markers. Additional genetic markers (e.g., isozymes, proteins, RFLPs, sequence-tagged sites) will be needed to establish synteny with other species. This genomic map of maritime pine will facilitate quantitative trait dissection studies and marker-assisted breeding. The RAPD map can be supplemented by known and unknown genes to further characterize the maritime pine genome.
As a part of our mapping project, 27 protein loci have been located on the maritime pine map described in this paper (Plomion et al. 1995).

Acknowledgements We thank Dario Grattapaglia, Jose Chaparro, Ross Weir, and Ben Liu for friendship, support and suggestions. We wish to thank Ron Sederoff for inviting C.P. to conduct research during the first year of his Ph.D. with the Forest Biotechnology Group, North Carolina State University, Raleigh. We thank Bruce Weir for helpful suggestions. This work was partly supported by grants from (Bureau des Ressources Génétiques from the Ministère de la Recherche et de l'Espace), the U.S. Department of Agriculture (NRCGP No. 92-37300-7549) and the USDA Forest Service (Coop Agreement No. 29-1099).

References