

An AFLP linkage map for Douglas-fir based upon multiple full-sib families

Nicholas K. Ukrainetz · Kermit Ritland ·
Shawn D. Mansfield

Received: 12 October 2006 / Revised: 22 April 2007 / Accepted: 1 May 2007 / Published online: 28 June 2007
© Springer-Verlag 2007

Abstract An amplified fragment length polymorphism (AFLP) linkage map for coastal Douglas-fir (*Pseudotsuga menziesii*) was constructed from eight full-sib families each consisting of 40 progeny. These families were part of the British Columbia Ministry of Forests second-generation progeny test program and represent typical family sizes used in progeny trials. For map construction, ten primer pairs using *EcoRI*+3 and *MseI*+4 were employed to identify and assay AFLP loci that segregated in backcross configurations. A new technique was used to obtain a single recombination rate for each pair of marker loci: for each locus pair, a recombination rate and log-odd value were estimated across all segregating families using a joint maximum likelihood function that considered the full dataset of segregating genotypes. The resulting matrix of recombination rates between all pairs of loci was used to construct an integrated linkage map using JoinMap. The final map consisted of 19 linkage groups spanning 938.6 cM at an average distance of 9.3 cM between markers. The simultaneous integration of data from multiple families may provide an effective way to construct a linkage map, using the genetic resources inherent in most tree improvement programs, where progeny tests of small size are conducted. The statistical property of number of families used is briefly discussed. For our data, at least

three to four families greatly increased the chance of obtaining an informative locus in at least one family. Families as small as ten are adequate for closely linked loci (<10 cM), while the size used in our study (40) is adequate for loci within 30 cM.

Keywords AFLP · Linkage map · Douglas-fir

Introduction

Douglas-fir (*Pseudotsuga menziesii* (Mirb.) Franco) is arguably one of the most important commercial tree species along the West Coast of Canada and the United States. Its native range extends from the Rocky Mountains to the Pacific Ocean and from central British Columbia to Mexico (USDA 2000). There are two varieties of Douglas-fir: coastal Douglas-fir (var. *menziesii*) and interior Douglas-fir (var. *glauca*). The coastal variety occurs mainly west of the Cascade and Coastal Mountains, and its wood is highly valued for lumber production due to its inherent high density and exceptional strength properties (Chantre et al. 2002; Koshy and Lester 1994; Loo-Dinkins and Gonzalez 1991; Loo-Dinkins et al. 1991; St. Clair 1994; Vargas-Hernandez and Adams 1991). The interior variety is also an important commercial species (Nigh et al. 2004) where it is used to produce high quality veneer (Hesterman and Gorman 1992) and lumber (Wagner et al. 2002).

Conifer genomes are large and complex, which make sequencing difficult and unlikely (Krutovsky et al. 2004). A more suitable approach to the investigation of the structure, organization, and evolution of conifer genomes is the development of large-scale genomics and association genetics programs of unrelated populations. An important component of this process is the development of a linkage

Communicated by R. Johnson

N. K. Ukrainetz · S. D. Mansfield (✉)
Department of Wood Science, University of British Columbia,
#4030-2424 Main Mall,
Vancouver, BC, Canada V6T 1Z4
e-mail: shawn.mansfield@ubc.ca

K. Ritland
Department of Forest Science, University of British Columbia,
Vancouver, BC, Canada

and quantitative trait loci (QTL) map for the study of a species, which requires the development of molecular markers. One such technique, amplified fragment length polymorphisms (AFLP), has proven to be one of the most reliable molecular marker techniques for saturating linkage maps (Cabrita et al. 2001; Cervera et al. 2000; Jones et al. 1997; Vos et al. 1995) and has been used for genetic mapping of several tree species (Cervera et al. 2001; Chagne et al. 2002; Remington et al. 1999; Scalfi et al. 2004; Travis et al. 1998; Wu et al. 2000; Yin et al. 2003; Zhang et al. 2004). Currently, there is no AFLP linkage map for Douglas-fir, although several linkage maps have been developed utilizing random amplification of polymorphic DNAs (RAPDs) and restriction fragment length polymorphisms (RFLPs; Jermstad et al. 1998; Krutovskii et al. 1998).

Linkage maps in forest trees have been shown to be useful for identifying quantitative trait loci (QTLs; Jermstad et al. 2001a, b; Markussen et al. 2003; Sewell et al. 2000, 2002) for candidate gene mapping (Brown et al. 2003; Wheeler et al. 2005) and for comparative mapping between species (Chagne et al. 2003; Krutovsky et al. 2004). Jermstad et al. (2001a, b, 2003) employed a linkage map for Douglas-fir to identify QTLs for adaptive traits including bud flush, fall and spring cold hardiness, and growth initiation and cessation. This map was also used for comparative genome mapping with loblolly pine and other conifer species (Krutovsky et al. 2004).

Linkage analyses in outbred organisms such as conifers generally employ marker information derived from the offspring of a single cross to develop two linkage maps, one for each parent. These maps are then joined using markers that segregate (are heterozygous) in both parents (Chagne et al. 2002; Jermstad et al. 1998; Wu et al. 2000). Although codominant markers are ideal for map joining, dominant markers can also be employed, but are not as informative, as the progeny segregate at a 3:1 ratio. Combining two or more linkage maps, often with unrelated pedigrees, can be done by an ad hoc clustering of pairwise recombination rates averaged over pedigrees. Alternatively, a JoinMap procedure can be employed (Van Ooijen and Voorrips 2001), which combines pairwise recombination estimates from different experiments in proportion to their log-odd (LOD) values (Stam 1993). Hu et al. (2004), in contrast, proposed a joint maximum likelihood method involving analyses of all original datasets; however, this technique requires access to the original data of all pedigrees.

In this paper, we present an AFLP linkage map for Douglas-fir based on the assay of eight 40-member full-sib families from the British Columbia Ministry of Forests second generation progeny test program. We employed the joint likelihood function of Hu et al. (2004) to calculate the most likely recombination rates across families, generating the overall LOD scores that serve as an input into the

JoinMap linkage program. The simultaneous integration of data from multiple families has the potential to be an effective way to construct a linkage map. This study provides the first AFLP map for Douglas-fir, and the ensuing linkage map will serve as a base for a QTL analysis of these pedigrees (Ukrainetz et al. 2007, submitted as accompanying paper <http://dx.doi.org/10.1007/s11295-007-0097-x>).

Materials and methods

Plant material and DNA isolation

Plant material was collected from eight full-sib families from the British Columbia Ministry of Forests second-generation progeny test program for coastal Douglas-fir, planted in 1977. The families were grown on four sites in southwestern British Columbia, with ten random individuals (of a possible 16) per family per site and 40 individuals per family. In May 2004, branches with newly flushing buds were cut and collected from the upper part of the live crown of each individual. Buds were removed from the branches, placed into cryovials and frozen in a vapor tank for transport, and stored at -80°C .

DNA was isolated from bud material using a cetyl trimethyl ammonium bromide (CTAB) procedure adapted from (Doyle and Doyle 1987). Buds are superior for AFLPs, as earlier tests showed mature needles to give erratic fuzzy banding patterns compared to newly flushing buds. Frozen buds were ground by mortar and pestle under liquid nitrogen, after which, to each sample, 400 μl of concentrated buffer base [0.2 M hydroxymethylaminoethane (Tris), 0.04 M ethylenediaminetetraacetic acid (EDTA) and 2.8 M NaCl at pH 8.3], and 400 μl of 0.1 M CTAB were added and briefly vortexed. The buffer–bud mixture was then incubated at 65°C for 1 h with routine shaking every 10 min. Samples were spun at 10,000 rpm for 2 min in a 4°C centrifuge and the supernatant collected. RNase (10 μg) was added to each sample and incubated at 37°C for 30–45 min, after which, 700 μl of chloroform/isoamyl alcohol (24:1) was added and rotamixed for an additional 30 min. The mixture was spun for 10 min at 10,000 rpm in a 4°C centrifuge and the supernatant collected. Cold isopropanol (400 μl) was added to each tube, mixed gently, and placed at -20°C for 30 min. Tubes were again spun for 30 min at 10,000 rpm in a 4°C centrifuge and the supernatant discarded. The pellet was washed twice with 200 μl of ice-cold 70% ethanol and the DNA dried in a speedvac for 5 min, re-suspended in 100 μl of sterile water, and incubated at 65°C for 1 h. DNA from each sample was quantified on a spectrophotometer (Pharmacia Biotech Ultrospec 3000) at 260 and 280 nm, diluted to 100 ng/ μl , and stored at -20°C .

AFLP template preparation and reactions

Restriction double-digests were completed using *EcoRI*/*MseI* and *PstI*/*MseI*, as per the restriction–ligation (RL) protocol of Vos et al. (1995). Previously, Paglia and Morgante (1998) reported that AFLP profiles generated using *EcoRI* have a low signal to noise ratio and limited number of discernable polymorphic bands in conifers. As a result, both *EcoRI* and *PstI* AFLP profiles were generated and compared for polymorphisms and ease of scoring. *EcoRI* produced an abundance of polymorphic loci, whereas *PstI* was rather monomorphic (data not shown). Consequently, *EcoRI* was used as the frequent cutter to generate AFLP profiles for this study. Approximately 1,000 ng of DNA were digested in a 10 µl solution of 5×RL buffer (50 mM Tris–HAc pH 7.5, 50 mM MgAc, 250 mM KAc, and 25 mM DTT), 8.04 U of *EcoRI*, 6.7 U of *MseI*, and incubated for 2–3 h at 37°C. A 2.5 µl solution containing 10 mM ATP, 5×RL buffer, 5 pmol/µl of *EcoRI* adapter, 50 pmol/µl of *MseI* adapter, and 0.5 U of T4 DNA ligase was added to 10 µl of the digested DNA product and incubated for 3 h at 37°C. After ligation, the RL product was diluted 1:10 with dH₂O and stored at –20°C. The adapter sequences were:

| | | |
|--------------|---------------|--------------------------|
| <i>EcoRI</i> | Top strand | 5' AACGACGACTGCGTACC 3' |
| | Bottom strand | 3' CTGCTGACGCATGGTTAA 5' |
| <i>MseI</i> | Top strand | 5' GACGATGAGTCCTGAG 3' |
| | Bottom strand | 3' TACTCAGGACTCAT 5' |

Amplification of the DNA template was adapted from Remington et al. (1999). Pre-amplification was carried out using *EcoRI* (E+AC) and *MseI* (M+CC) primers:

| | |
|--------------|------------------------|
| <i>EcoRI</i> | 5' GACTGCGTACCAATTC 3' |
| <i>MseI</i> | 5' GATGAGTCCTGAGTAA 3' |

A 7.5 µl solution containing 15 ng of E+AC, 15 ng of M+CC, 2 mM dNTP mix (with equal volumes of 0.4 mM dATP, dCTP, dGTP, and dTTP), 10×PCR buffer (Roche Diagnostics) and 0.6 U Taq polymerase (Roche Molecular Systems) was added to 2.5 µl of the RL product. The thermocycler conditions for PCR amplifications were as follows: 1 min denaturation at 94°C, 28 cycles of 30 s at 94°C, 30 s at 60°C, and 60 s at 72°C followed by a 5 min extension at 72°C. The pre-amplification product was diluted 1:40 with dH₂O before final amplification and stored at –20°C.

For final amplification, a solution of 10 µl containing 2.5 µl of pre-amplification product, 0.25 pmol of M13 labeled primer, 2.52 ng of E+3 tailed primer, 2.52 ng of

M+4 primer, 2 mM dNTP mix, 10×PCR buffer (Roche Diagnostics), and 0.6 U Taq polymerase was placed in a thermocycler. The primer sequences used were:

| | |
|--------------------|---|
| M13 labeled primer | 5' CACGACGTTGTAAAACGAC 3' |
| <i>EcoRI</i> | 5' CACGACGTTGTAAAACGACGACT GCGTACCAATTC 3' |
| <i>MseI</i> | 5' GATGAGTCCTGAGTAA 3' |

PCR amplifications were carried out as follows: 1 min denaturation at 94°C, three cycles of 30 s at 94°C, 30 s at 65°C, and 60 s at 72°C, 12 cycles of 30 s at 94°C, 30 s at 65°C (–0.7°C/cycle), and 1 min at 72°C, 23 cycles of 30 s at 94°C, 30 s at 56°C and 60 s at 72°C, and a 5 min extension step at 72°C. After amplification, 3 µl of formamide loading buffer was added to each sample and stored in aluminum foil at –20°C.

Detection and scoring of AFLP fragments

The AFLP products were resolved using a LiCor 4200 autosequencer with 25 cm plates. The gel was prepared with a 30 ml solution containing 7% Long Ranger polyacrylamide (FMC BioProducts), 7 M urea and 5×TBE (0.45 M Tris, 0.45 M boric acid, and 0.01 M EDTA). Ammonium persulfate (200 µl) was added to the urea solution and filtered. TEMED (*N,N,N,N*-tetramethyl-ethelenediamine; 15 µl) was added to the filtered solution to begin the solidification process. Forty-eight lane gels were used with 1 µl of sample in each well. Each gel contained one family with parents replicated three times (46 individuals) and two IRD-labeled molecular-weight markers (LiCor; size standard IRDye™) as standards. Two primer sets were loaded per gel, each containing a different M13 primer channel (700 or 800 nm). Electrophoresis was carried out using 1×TBE running buffer with motor speed 4, 2,000 V, 35 mA, 70 W and 50°C plate temperature, and 4 h running time.

Ninety-two primer combinations were screened using 12 individuals per primer set (three individuals from each family selected from Adam River, Gold River, and Squamish River). Five primer combinations were screened on each gel image separated by IRD-labeled standards. All E+3/M+4 and some E+3/M+5 combinations with an E+AC and M+CC pre-amplification were screened (Table 1). Primer pairs were first screened based on the number of polymorphic loci within families, then on their ease of scoring. Ten primer pairs were selected for final analysis (Table 2). Reproducibility tests were performed using ten individuals replicated four times on each gel with the same primer combination. Two primer pairs were used and the

Table 1 Number of polymorphic loci generated by each primer combination during primer screening

| Mse + | Eco + ACA | ACC | ACG | ACT | Total |
|-------|-----------------|-----------------|-----------------|-----|-------|
| CCAA | 19 | 0 | 20 | 2 | 41 |
| CCAC | 20 | 5 | 14 | 6 | 45 |
| CCAG | 38 | 28 | 34 | 18 | 118 |
| CCAT | 0 | 1 | 2 | 9 | 12 |
| CCCA | 4 | 1 | 17 ^a | 17 | 39 |
| CCCC | 26 | 15 | 25 | 13 | 79 |
| CCCG | 31 | 19 | 28 ^a | 51 | 129 |
| CCCT | 10 | 0 | 0 | 2 | 12 |
| CCGA | 35 | 9 | 22 ^a | 8 | 74 |
| CCGC | 32 | 37 ^a | 9 | 9 | 87 |
| CCGG | 48 ^a | 39 | 20 ^a | 24 | 131 |
| CCGT | 52 ^a | 41 | 22 ^a | 7 | 122 |
| CCTA | 8 | 7 | 3 | 8 | 26 |
| CCTC | 18 | 9 | 23 ^a | 14 | 64 |
| CCTG | 48 | 17 | 30 ^a | 20 | 115 |
| CCTT | 24 | 1 | 24 | 6 | 55 |
| CCAGA | 26 | 18 | 28 | 6 | 78 |
| CCAGC | 34 | 39 | 24 | 15 | 112 |
| CCAGG | 15 | 16 | 20 | 19 | 70 |
| CCAGT | 17 | 10 | 28 | 31 | 86 |
| CCCAA | 15 | 30 | 0 | 21 | 66 |
| CCGAA | 17 | 16 | 19 | 25 | 77 |
| CCGGA | 32 | 24 | 28 | 23 | 107 |
| Total | 569 | 382 | 440 | 354 | 1,745 |

^a Primer combinations selected for final analysis

data analyzed for missing or additional bands within replicated individuals for a given locus.

Scoring was completed by eye using the Saga (Generation 2) software program. Data was recorded as present (+) or absent (–) by the computer software. Loci that apparently segregated in a 1:1 ratio and 3:1 ratio were examined. If a segregating locus was detected in one family, the same locus was assessed and scored in all families. The markers were imported into JoinMap (Van Ooijen and Voorrips 2001) and checked for deviation from 1:1 ratios. Only loci that segregated 1:1 in all segregating families were included in the following analyses.

Linkage analysis

Average recombination rates and LOD scores for all marker pairs across families were calculated using the procedure of Hu et al. (2004) as implemented in a Fortran 95 program written by KR (available upon request). This procedure obtains a single maximum likelihood estimate of recombination between a pair of loci for all families that segregate for both markers; linkage phase is also inferred from the data. All pairwise estimates of recombination with LOD ≥ 0 were inputted into JoinMap (Van Ooijen and Voorrips 2001).

Grouping was initially carried out from LOD 2 to 10 with a step of one LOD, but in the end, the linkage map was grouped at LOD thresholds of 3 and subsequently filled in with those at 2; the lower end for trees (Cervera et al. 2001; Chagne et al. 2002; Jermstad et al. 1998; Scalfi et al. 2004; Travis et al. 1998; Wu et al. 2000), but deemed appropriate for this data set which employed a multiple family approach to linkage construction. All groups with ≥ 3 markers were retained for mapping. Maps were generated using the Kosambi mapping function with recombination rates < 0.45 and LOD ≥ 0 and a ripple with jump threshold of 5.

In some cases, due to the lack of adequate recombination (REC) and LOD information, complete maps for some groups could not be created. In these cases, markers within these groups were split based on linkage, and two or more maps were generated and linkage between smaller groups and larger groups assessed. When smaller groups linked within larger groups, the smaller groups were removed. A third round of marker addition was permitted, but without reordering first and second generation maps. Linkage groups with ≥ 4 loci were retained as major linkage groups to represent the Douglas-fir genome. We calculated total map distance that covered over all linkage groups and average distance between markers as total map distance divided by the number of mapped markers.

Marker distribution along the map

AFLP markers should resemble a random sample from the genome (Vos et al. 1995). The distribution of AFLP markers throughout the linkage map was tested using the method of Remington et al. (1999), which examines marker clustering by testing each linkage group for the expected number of markers, λ_i :

$$\lambda_i = \frac{mG_i}{\sum_i G_i}$$

Table 2 Backcross, intercross, and the total number of markers generated by each primer combination in the final AFLP analysis (only backcross markers were used in the map)

| Eco | Mse | Backcross | Intercross | Total |
|-----|------|-----------|------------|-------|
| ACA | CCGG | 145 | 39 | 184 |
| | CCGT | 167 | 54 | 221 |
| ACC | CCGC | 85 | 24 | 109 |
| ACG | CCCA | 100 | 31 | 131 |
| | CCCG | 53 | 12 | 65 |
| | CCGA | 79 | 29 | 108 |
| | CCGG | 109 | 44 | 153 |
| | CCGT | 104 | 35 | 139 |
| | CCTC | 108 | 36 | 144 |
| | CCTG | 72 | 36 | 108 |

where m is the total number of markers and G_i is the map distance of each linkage group adjusted for chromosome ends as $G_i = M_i + 2s$ where M_i is the map distance between terminal markers and s is the average marker spacing for the linkage map. The observed number of markers per group was tested using a two-tailed cumulative Poisson distribution with parameter λ_i (p values < 0.025 are significant at $\alpha = 0.05$).

Family contributions to the map

Ideally, each family should contribute equally to the linkage map. However, variability in the number of informative loci among families will result in unequal contributions. To assess this effect, family contribution to the linkage map was compiled and deviations from the average contribution evaluated with a chi-square test (SAS 2003). In order for marker information to contribute to the linkage data of a linkage group, two markers must be present from a common family. Therefore, if a family contributed just one marker to a linkage group, it was disregarded.

Genome coverage by multiple families

To ascertain how using multiple families aids in the construction of a comprehensive genetic map that incorporates all polymorphic markers in the population, the existing dataset was re-sampled over n entire families with replacement, for $n = 1$ to 8, and evaluated (1) the number of families with a heterozygous parent, averaged over loci, and (2) the probability that at least one family has a heterozygous parent, averaged over loci. This bootstrap-like procedure evaluated the statistical properties of using multiple families.

Results

AFLP Polymorphisms

Primer screening revealed many polymorphic loci for the E+3/M+4 primer combinations (Table 1). Combinations of markers with E+ACG, E+ACC, and E+ACA produced the best bands and were selected for the final analysis. E+ACA generated the most polymorphic loci followed by E+ACG, while both E+ACC and E+ACT produced the lowest number of polymorphic loci (Table 1). Of the M+4 primers, those with guanine at their 3' end generally produced the most polymorphic loci. In total, there were 1,745 polymorphic loci detected after the initial screening. Ten primer combinations were ultimately chosen for linkage mapping based on the number of polymorphic loci and profile clarity (shown in Table 2). For each primer combination, there

were two to four times more backcross markers present than intercross markers (Table 2).

Map results

A total of 531 markers that segregate in a 1:1 ratio in any of the eight full-sib families were used to calculate average pairwise LOD and REC data. The initial grouping with JoinMap was able to group 244 markers at LOD thresholds of 2 and 3. The final map contained 120 markers mapped on 19 linkage groups (Fig. 1). On average, there were 6.3 markers per group with an average of 9.3 cM between markers. The total map distance covered by the 19 linkage groups was 938.6 cM. Another 63 markers were mapped in groups of three on 21 linkage groups that covered a distance of 505 cM, with an average distance of 12.0 cM between markers. A total of 61 markers that were initially grouped were rejected in the final map.

Marker distribution

There was a significant excess of markers in linkage groups 1, 2, and 11 (Table 3), with two of these groups (1, 2) showing high statistical significance for the two-tailed p value ($p > 0.99$). There are generally fewer than expected markers in the smaller groups, and more than the expected markers in the larger groups, as illustrated by Fig. 2.

Family contributions

The test for family contribution to the linkage map suggests that families did not contribute equally to the linkage map. Table 4 shows that families 7, 26, 62, and 151 occur throughout the linkage map as expected under the chi-squared distribution (chi-square values less than 3.84), while families 2, 38, and 92 contribute excessively; family 75 showed little contribution. At the linkage group level, we found that individual linkage groups may contain markers provided by several families; however, in most cases, the linkage contribution is biased to a small number of families, often one or two. For example, linkage group 7 is composed of markers derived from family 2, whereas linkage group 1 contains markers from six of the eight families, but is biased towards markers being contributed by families 2, 62, and 75.

Genome coverage by multiple families

The results of resampling the current data for a specified number of families (crosses) shows that as more families are employed in the analysis, the number of heterozygous (informative) parents across all families increases in a strictly linear fashion (Fig. 3a). Figure 3b shows that the

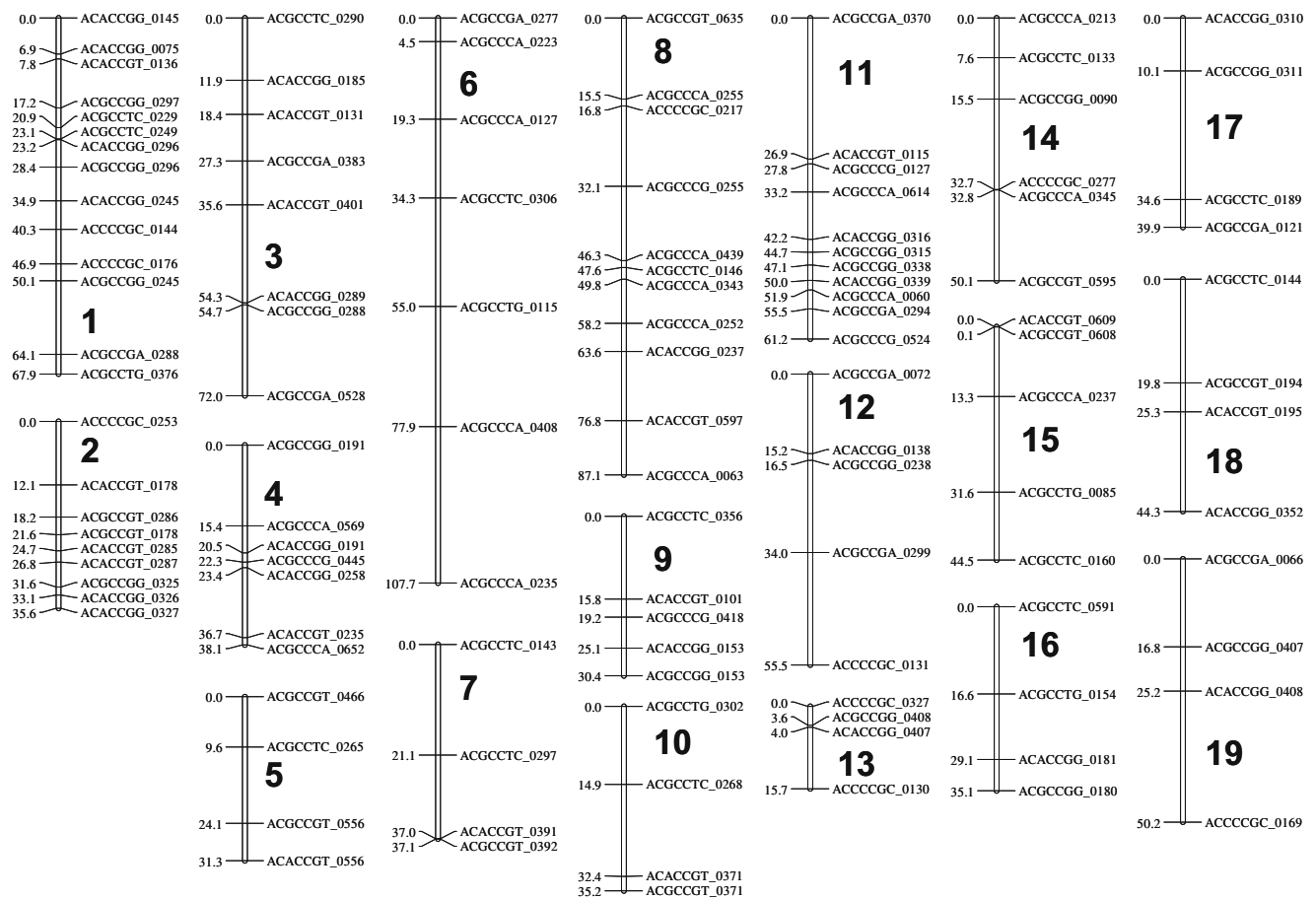


Fig. 1 AFLP linkage map for integrated segregation data from eight full-sib families of coastal Douglas-fir. Marker names are the E+3/M+4 primer combination and size. Map distances are in cM estimated by the Kosambi mapping function

probability that at least one parent is heterozygous asymptotically approaches one, but the increase slows appreciably after three to four families. Even with eight families, 20% of the loci have no heterozygous parents and hence are not mappable.

Discussion

The present AFLP map is composed of 120 markers on 19 linkage groups covering 938.6 cM with an average of 9.3 cM between markers. The advantage of this linkage map over those generated in other studies is the integration of segregation data from eight full-sib families. Rather than relying on a single cross or family to generate a linkage map, our map is a synthesis of data across eight full-sib families and represents loci more commonly polymorphic, and hence, of use in other pedigrees for QTL mapping and marker assisted selection.

Many AFLP linkage maps have been constructed for tree species. Two of the most comprehensive maps were generated for loblolly pine (Remington et al. 1999) and

maritime pine (*Pinus pinaster* Ait.; Chagne et al. 2002). Both maps resulted in 12 linkage groups which correspond to the haploid chromosome complement in pine. Several other AFLP maps have been generated for conifers (Travis et al. 1998; Yin et al. 2003) and hardwoods (Cervera et al. 2001; Scalfi et al. 2004; Wu et al. 2000; Zhang et al. 2004). Generally, these conifer linkage maps result in greater than 12 linkage groups, more than the number of chromosomes. This is consistent with the map generated in the present study, with 19 major linkage groups (Douglas-fir has 13 chromosomes, unlike other members of the pine family).

One recent linkage map for coastal Douglas-fir employed both dominant (RAPDs) and co-dominant markers (RFLPs; Jermstad et al. 1998) and has been the basis for QTL analysis of adaptive traits (Jermstad et al. 2001a, b, 2003) and comparative mapping with loblolly pine (Krutovsky et al. 2004). The map presented in this paper is of similar size and density to the one generated by Jermstad et al. (1998) who created a linkage map with 141 markers distributed on 17 major linkage groups covering a distance of 1,062 cM and a density of 7.5 cM between markers, as well as a second linkage map consisting of 15

Table 3 Marker density per linkage group

| LG | m_i | M_i (cM) | G_i (cM) | λ_i | p value |
|-----|-------|------------|------------|-------------|-----------|
| 1 | 14 | 67.9 | 78.3 | 7.04 | 0.994 |
| 2 | 9 | 35.6 | 44.5 | 4.00 | 0.992 |
| 3 | 8 | 72.0 | 92.6 | 8.32 | 0.547 |
| 4 | 7 | 38.1 | 50.7 | 4.56 | 0.908 |
| 5 | 4 | 31.3 | 52.2 | 4.70 | 0.495 |
| 6 | 7 | 107.7 | 143.6 | 12.91 | 0.056 |
| 7 | 4 | 37.1 | 61.9 | 5.56 | 0.348 |
| 8 | 11 | 87.1 | 104.6 | 9.40 | 0.762 |
| 9 | 5 | 30.4 | 45.5 | 4.09 | 0.770 |
| 10 | 4 | 35.2 | 58.6 | 5.27 | 0.394 |
| 11 | 11 | 61.2 | 73.4 | 6.60 | 0.963 |
| 12 | 5 | 55.5 | 83.2 | 7.48 | 0.243 |
| 13 | 4 | 15.7 | 26.2 | 2.36 | 0.909 |
| 14 | 6 | 50.1 | 70.1 | 6.31 | 0.557 |
| 15 | 5 | 44.5 | 66.7 | 6.00 | 0.446 |
| 16 | 4 | 35.1 | 58.4 | 5.25 | 0.397 |
| 17 | 4 | 39.9 | 66.5 | 5.98 | 0.288 |
| 18 | 4 | 44.3 | 73.8 | 6.63 | 0.209 |
| 19 | 4 | 50.2 | 83.6 | 7.52 | 0.131 |
| Sum | 120 | 938.6 | 1334.4 | 120 | |

The expected number of markers (λ_i) and the observed number of markers (m_i) is not significantly different if the Poisson two-tailed p value is >0.025 .

LG Linkage group, m_i number of markers, M_i actual linkage group length, G_i inferred linkage group length, λ_i expected number of markers, P value Poisson two-tailed p value

linkage groups spanning 897 cM using 72 equally spaced markers (Jermstad et al. 2003).

One of the criticisms of linkage mapping is the limitation of the analysis to a single individual or pair of individuals (for a sex-averaged map), which has implications for applications of derived maps. However, in at least one major instance, a linkage map has been created for loblolly pine (*Pinus taeda* L.) integrating segregation data from two outbred, three-generation pedigrees using intercross markers to align maps (Sewell et al. 1999). The resulting consensus linkage map consisted of 357 markers (258 RFLPs, 67 RAPDs, and 12 isozymes) and covered ~1,359 cM over 20 linkage groups. Twelve linkage groups were integrated with map information from all four mapping populations (all four parents), two were specific to their QTL pedigree, and six were specific to single populations (Sewell et al. 1999).

How many families (crosses) should be used? If too many crosses are used, the contribution of each cross can become small. If the linkage phase of parents is known, even families of size one can contribute to estimates of recombination (Ott 1999). If linkage phase is to be inferred (as in this study), family sizes of at least 30–40 should be used to infer phase between loci 20 and 30 mu apart. Given the results of Fig. 3b, at the minimum, it seems three to four families are needed to ensure at least one family is segregating.

What is the minimum family size for inferring linkage phase? This may dictate family number, as total sample size might be fixed. For recombination rate r , the probability of the correct phase, given the data of k , apparent recombinants and $m-k$ apparent non-recombinants (family size is m) is

$$P(m, k, r) = \frac{1}{1 + (r/(1 - r))^{m-2k}}$$

(Ott 1999, Eq. 5.19). These are “apparent” in the sense that we assume $k < m - k$ always, although the reverse might be true for high r , should phase be incorrectly inferred. Considering the sampling distribution of k , the overall expected probability for any particular m and r is

$$\sum_{k=0}^m B(m, k, r)P(m, k, r)$$

where $B(m, k, r)$ is the binomial probability. Figure 4 shows the numerical evaluation of this expression for various values of m and r . There is a clear dependence of this probability on both linkage and family size. For closely linked loci ($r < 0.1$), family sizes of ten are adequate. Family sizes of 20–30 are adequate when $r < 0.25$, and a family size of 40 (our study) is adequate for $r < 0.30$.

Primer pair characteristics

AFLP markers are very reproducible, and the technique is capable of producing large numbers of polymorphic loci per PCR reaction. Primer combinations of E+3/M+4 with an

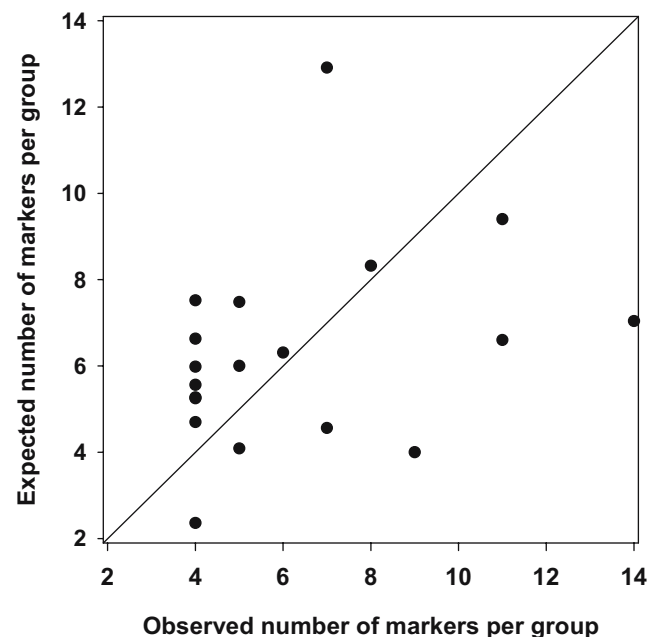


Fig. 2 Observed vs expected number of markers across linkage groups. There are fewer than expected markers in the smaller groups and more than the expected markers in the larger groups

Table 4 Contributions of individual families to the linkage map; total chi-square is 39.26 (7 df, $p < 0.001$)

| Family | Number of mapped markers | Contribution to chi-square |
|--------|--------------------------|----------------------------|
| 2 | 46 | 13.37 |
| 7 | 17 | 3.70 |
| 26 | 25 | 0.15 |
| 38 | 38 | 4.48 |
| 62 | 23 | 0.59 |
| 75 | 11 | 9.48 |
| 92 | 38 | 4.48 |
| 151 | 18 | 3.00 |
| Mean | 27 | |

E+2/M+2 pre-amplification gave the best results. The frequent cutter *PstI* was used to compare profiles and polymorphisms with *EcoRI* and found to lack polymorphic loci in Douglas-fir. Although Paglia and Morgante (1998) suggest that *PstI* is less prone to cutting within highly repetitive regions of the conifer genome, this restriction enzyme did not generate profiles of the same quality as *EcoRI* in the present study. As conifer genomes contain large amounts of repetitive DNA, *PstI* may result in profiles with higher signal to noise (Paglia and Morgante 1998). Our approach for mapping was more concerned with the ease of scoring than the number of polymorphic bands produced. As the signal to noise ratio between *PstI* and *EcoRI* profiles was similar, priority was given to maximizing the number of polymorphic markers produced per reaction.

Primer combinations using E+ACA produced the most polymorphic loci and were the easiest to score. *MseI* primers with guanine as the 3' selective nucleotide gave the best results. This is consistent with Remington et al. (1999) who report optimal results from primer combinations with one CpG unit in the selective region of *EcoRI* or *MseI* primers, and Paglia and Morgante (1998) who report more polymorphic loci in *PstI* profiles with CpG units within the selective region of final amplification primers. In contrast, our results suggest that primer combinations with E+ACG produce significant polymorphisms, but are difficult to score. The best combinations for ease of scoring and polymorphism production were E+ACA with M+CCGG and M+CCGT (Tables 1 and 2). Several E+3/M+5 primer combinations were also screened, but produced profiles that had darker backgrounds and were difficult to discern. Mismatches can occur when three or more selective nucleotides are added to the 3' end of a primer, resulting in inconsistent results and genotyping error.

Maps of relative low marker density are expected to cover a large proportion of the genome, as suggested by Travis et al. (1998) who report that ten primer combinations are enough to cover 72% of the genome at a density of

10 cM between markers and that 10–20 primer combinations are optimal for QTL analysis. For multiple-family mapping, Hu et al. (2004) suggested that 50 individuals from ten crosses are sufficient to generate linkage maps with the joint likelihood method. In this study, we used ten primer combinations from eight families with 40 individuals per family, which is slightly less than the number of crosses and individuals recommended for this technique and may be the cause of the low LOD values necessary for map development and the large number of unmapped loci. Although our map resulted in more than the expected number of linkage groups, it is expected to cover an adequate amount of the genome for QTL mapping.

For each primer combination, we observed that there were two to four times more backcross markers present than intercross markers (Table 2). Assuming Hardy–Weinberg genotype frequencies, this corresponds to a recessive gene

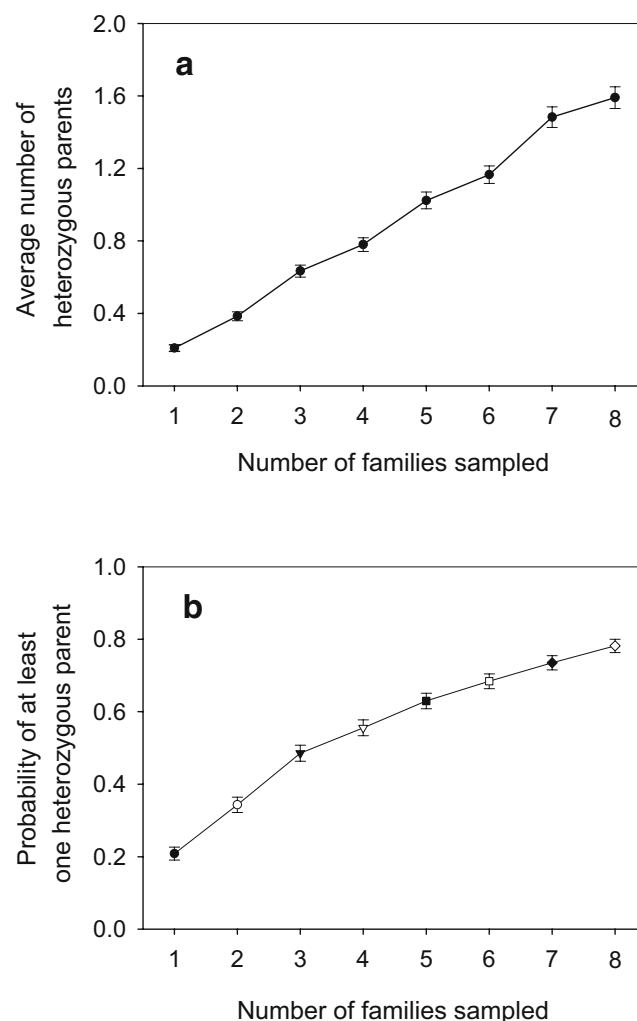


Fig. 3 Results of re-sampling the existing dataset for a specified number of families (*crosses*). **a** The number of heterozygous (informative) parents across all families increases as more families are used. **b** The probability that at least one parent is heterozygous asymptotically approaches one

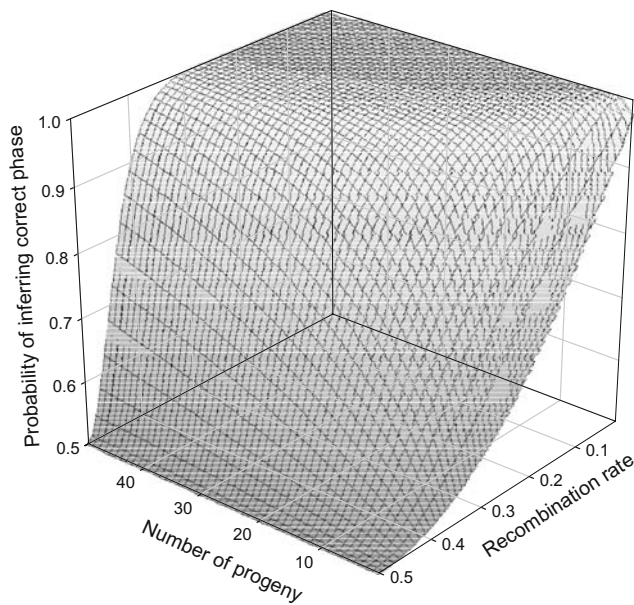


Fig. 4 The probability of inferring the correct phase of an individual heterozygous for two linked loci as a function of recombination rate and family size (number of progeny)

frequency of 1/5 to 1/3. This implies that informative markers involve recessives with rather low frequencies, where the frequency of the recessive bandless phenotype is 1/25 to 1/9.

The puzzle of unequal family contribution

In the absence of variation of inbreeding (unexpected in almost completely outcrossing conifers), the observed heterozygosity, and hence the information provided about linkage, should be the same among families. Yet, we observed rather dramatic differences among families for their contribution to the genetic map (Table 4). This may be related to DNA quality, which was variable between families, even when using freshly flushing buds. For example, families 2 and 38 consistently provided the cleanest AFLP profiles and were easy to score. Consequently, they contributed the most polymorphic markers to the linkage analysis, which likely reflects the number of easily scorable markers rather than heterozygosity. Alternatively, it is possible that the genetic distance of the parents of individual crosses differs significantly and consequently results in greater heterozygosity in the offspring, and therefore, has a higher proportion of polymorphic AFLP markers that are useable for mapping.

AFLPs were also attempted using 1-year-old Douglas-fir needles. Using old tissue, we were unable to generate employable AFLP profiles (data not shown). This was most likely due to DNA quality and the presence of extractives, and not to DNA quantity. The presence of contaminants in the DNA can affect the restriction digest and PCR amplifica-

tions necessary to generate AFLP profiles. Clearly, the quality of DNA for AFLP generation is very important, and as such, newly flushing bud material was optimal for isolating clean DNA.

Joining markers into groups

A large number of markers were left unmapped in this study. Of the 531 markers initially identified, only 120 could be placed in linkage groups. As the linkage map was generated using segregation data from eight full-sib families, identified markers segregated in one to six families; thus, the number of segregates varied by locus. When segregation was in just one or two families, there was often not enough statistical information to ascertain linkage, especially if it was weak. The end result is that only more polymorphic loci can be mapped. This can be viewed as an advantage of using multiple small families, as one ends up mapping only those markers of higher frequency, which are also of more use for QTL and association mapping.

As well, only more closely linked loci could be mapped, in principle, as increasing the number of markers may improve map density and join linkage groups. However, Jermstad et al. (1998) attempted to identify linkage among groups by adding more markers and actually found no improvements. In the context of multiple family mapping (this study), alternatively, it could be more advantageous to add more families to coalesce linkage groups. Adding more families should increase the number of loci that segregate in multiple families, and will increase the chance that linked markers will segregate in common families.

Finally, we note that the linkage map presented in this paper was initially grouped at LOD thresholds of 2 or 3, which is not uncommon for trees (e.g., Cervera et al. 2001; Chagne et al. 2002; Jermstad et al. 1998; Scalfi et al. 2004; Travis et al. 1998; Wu et al. 2000). After grouping, all available pairwise data were needed to order markers onto linkage groups. This meant using all pairwise data with LOD values greater than or equal to zero and with a recombination rate of 0.45 or less. Low LOD values indicate that two markers are either far apart on a chromosome or are not linked. These markers may be located on a common linkage group, but at distances that allow significant recombination.

Applications to tree breeding

Linkage maps are a precursor to several genomics applications, including QTL mapping, comparisons of genetic maps among species and transfer of genomic information, and candidate gene association studies. Generally, a specific set of parents is chosen to represent the genome of a particular species or population and can serve as a backbone

for comparing QTL activity in different genetic backgrounds. However, the use of a single family to represent a population is limiting, as the resulting QTL maps, comparative genomics, and candidate gene mapping present a very narrow picture of the genome of a population and eliminate variation that may occur for QTLs within the species. By integrating multiple families, a more comprehensive map can be generated, which represents the more common markers and QTLs within a species, which are of most relevance to immediate breeding programs.

Multiple family analyses have several implications for tree breeding. Tree improvement programs are concerned with selecting characteristics, mainly wood volume and wood quality, to propagate from parents the progeny breeding population. Testing the parents requires the assessment of progeny in large scale, full-sib family field trials, but with rather limited numbers of progeny (10–40) compared to those required for genetic mapping (48–96) or QTL mapping (96–1,584). By adopting multiple family analyses, improvement programs can create “population” linkage maps and ultimately create QTL (Ukrainetz et al. 2007, submitted as accompanying paper <http://dx.doi.org/10.1007/s11295-007-0097-x>) and/or association maps in these populations of selected trees.

Acknowledgments The authors would like to acknowledge funding (to SDM) from Natural Resources Canada (NRCan) Value-to-Wood Program for this project. The authors also gratefully acknowledge Alvin Yanchuk and Michael Stoehr of the BC Ministry of Forests for access to the Douglas-fir breeding trials.

References

- Brown GR, Bassoni DL, Gill GP, Fontana JR, Wheeler NC, Megraw RA, Davis MF, Sewell MM, Tuskan GA, Neale DB (2003) Identification of quantitative trait loci influencing wood property traits in loblolly pine (*Pinus taeda* L.). III. QTL verification and candidate gene mapping. *Genetics* 164:1537–1546
- Cabrita LF, Aksoy U, Hepaksoy S, Leitao JM (2001) Suitability of isozyme, RAPD and AFLP markers to assess genetic differences and relatedness among fig (*Ficus carica* L.) clones. *Sci Hortic* 87:261–273
- Cervera MT, Remington D, Frigerio JM, Storme V, Ivens B, Boerjan W, Plomion C (2000) Improved AFLP analysis of tree species. *Can J For Res* 30:1608–1616
- Cervera MT, Storme V, Ivens B, Gusmao J, Liu BH, Hostyn V, Van Slycken J, Van Montagu M, Boerjan W (2001) Dense genetic linkage maps of three *Populus* species (*Populus deltoides*, *P. nigra* and *P. trichocarpa*) based on AFLP and microsatellite markers. *Genetics* 158:787–809
- Chagne D, Lalanne C, Madur D, Kumar S, Frigerio JM, Krier C, Decroocq S, Savoure A, Bou-Dagher-Kharrat M, Bertocchi E, Brach J, Plomion C (2002) A high density genetic map of maritime pine based on AFLPs. *Ann For Sci* 59:627–636
- Chagne D, Brown G, Lalanne C, Madur D, Pot D, Neale D, Plomion C (2003) Comparative genome and QTL mapping between maritime and loblolly pines. *Mol Breed* 12:185–195
- Chantre G, Rozenberg P, Baonza V, Macchioni N, Le Turcq A, Rueff M, Petit-Conil M, Heois B (2002) Genetic selection within Douglas-fir (*Pseudotsuga menziesii*) in Europe for papermaking uses. *Ann For Sci* 59:583–593
- Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh tissue. *Phytochem Bull* 19:11–15
- Hesterman ND, Gorman TM (1992) Mechanical-properties of laminated veneer lumber made from interior Douglas-fir and lodgepole pine. *For Prod J* 42:69–73
- Hu X-S, Goodwillie C, Ritland K (2004) Joining genetic linkage maps using a joint likelihood function. *Theor Appl Genet* 109:996–1004
- Jermstad KD, Bassoni DL, Wheeler NC, Neale DB (1998) A sex-averaged genetic linkage map in coastal Douglas-fir (*Pseudotsuga menziesii* [Mirb.] Franco var ‘menziesii’) based on RFLP and RAPD markers. *Theor Appl Genet* 97:762–770
- Jermstad KD, Bassoni DL, Jech KS, Wheeler NC, Neale DB (2001a) Mapping of quantitative trait loci controlling adaptive traits in coastal Douglas-fir. I. Timing of vegetative bud flush. *Theor Appl Genet* 102:1142–1151
- Jermstad KD, Bassoni DL, Wheeler NC, Anekonda TS, Aitken SN, Adams WT, Neale DB (2001b) Mapping of quantitative trait loci controlling adaptive traits in coastal Douglas-fir. II. Spring and fall cold-hardiness. *Theor Appl Genet* 102:1152–1158
- Jermstad KD, Bassoni DL, Jech KS, Ritchie GA, Wheeler NC, Neale DB (2003) Mapping of quantitative trait loci controlling adaptive traits in coastal Douglas-fir. III. Quantitative trait loci-by-environment interactions. *Genetics* 165:1489–1506
- Jones CJ, Edwards KJ, Castaglione S, Winfield MO, Sala F, van de Wiel C, Bredemeijer G, Vosman B, Matthes M, Daly A, Brettschneider R, Bettini P, Buiatti M, Maestri E, Malcevski A, Marmioli N, Aert R, Volckaert G, Rueda J, Linacero R, Vazquez A, Karp A (1997) Reproducibility testing of RAPD, AFLP and SSR markers in plants by a network of European laboratories. *Mol Breed* 3:381–390
- Koshy MP, Lester DT (1994) Genetic variation of wood shrinkage in a progeny test of coastal Douglas-fir. *Can J For Res* 24:1734–1740
- Krutovskii KV, Vollmer SS, Sorensen FC, Adams WT, Knapp SJ, Strauss SH (1998) RAPD genome maps of Douglas-fir. *J Heredity* 89:197–205
- Krutovsky KV, Troggio M, Brown GR, Jermstad KD, Neale DB (2004) Comparative mapping in the Pinaceae. *Genetics* 168:447–461
- Loo-Dinkins JA, Gonzalez JS (1991) Genetic control of wood density profile in young Douglas-fir. *Can J For Res* 21:935–939
- Loo-Dinkins JA, Ying CC, Hamm EA (1991) Stem volume and wood relative density of a nonlocal Douglas-fir provenance in British-Columbia. *Silvae Genet* 40:29–35
- Markussen T, Fladung M, Achere V, Favre JM, Faivre-Rampant P, Aragones A, Perez DD, Harvengt L, Espinel S, Ritter E (2003) Identification of QTLs controlling growth, chemical and physical wood property traits in *Pinus pinaster* (Ait.). *Silvae Genet* 52:8–15
- Nigh GD, Ying CC, Qian H (2004) Climate and productivity of major conifer species in the interior of British Columbia, Canada. *For Sci* 50:659–671
- Ott J (1999) Analysis of human genetic linkage, 3rd edn. The Johns Hopkins University Press, Baltimore, p 382
- Paglia G, Morgante M (1998) PCR-based multiplex DNA fingerprinting techniques for the analysis of conifer genomes. *Mol Breed* 4:173–177
- Remington DL, Whetten RW, Liu B-H, O’Malley DM (1999) Construction of an AFLP genetic map with nearly complete genome coverage in *Pinus taeda*. *Theor Appl Genet* 98:1279–1292
- SAS Institute Inc. (2003) SAS system for Windows, Version 9.1. SAS Institute Inc, Cary, NC, USA
- Scalfi M, Troggio M, Piovani P, Leonardi S, Magnaschi G, Vendramin GG, Menozzi P (2004) A RAPD, AFLP and SSR linkage map, and

- QTL analysis in European beech (*Fagus sylvatica* L.). *Theor Appl Genet* 108:433–441
- Sewell MM, Sherman BK, Neale DB (1999) A consensus map for loblolly pine (*Pinus taeda* L.). I. Construction and integration of individual linkage maps from two outbred three-generation pedigrees. *Genetics* 151:321–330
- Sewell MM, Bassoni DL, Megraw RA, Wheeler NC, Neale DB (2000) Identification of QTLs influencing wood property traits in loblolly pine (*Pinus taeda* L.). I. Physical wood properties. *Theor Appl Genet* 101:1273–1281
- Sewell MM, Davis MF, Tuskan GA, Wheeler NC, Elam CC, Bassoni DL, Neale DB (2002) Identification of QTLs influencing wood property traits in loblolly pine (*Pinus taeda* L.). II. Chemical wood properties. *Theor Appl Genet* 104:214–222
- St. Clair JB (1994) Genetic variation in tree structure and its relation to size in Douglas-fir. I. Biomass partitioning, foliage efficiency, stem form, and wood density. *Can J For Res* 24:1226–1235
- Stam P (1993) Construction of integrated genetic linkage maps by means of a new computer package: JoinMap. *Plant J* 3:739–744
- Travis SE, Ritland K, Whitham TG, Keim P (1998) A genetic linkage map of Pinyon pine (*Pinus edulis*) based on amplified fragment length polymorphisms. *Theor Appl Genet* 97:871–880
- Ukrainetz NK, Ritland K, Mansfield SD (2007) Identification of quantitative trait loci for wood quality and growth across eight full-sib coastal Douglas-fir families. *Tree Genet Gen DOI* 10.1007/s11295-007-0097-x
- USDA (2000) Wood handbook: wood as an engineering material. University Press of the Pacific, Madison, Wisconsin, 463 pp
- Van Ooijen JW, Voorrips RE (2001) JoinMap 3.0, software for the calculation of genetic linkage maps. Plant Research International, Wageningen, The Netherlands
- Vargas-Hernandez J, Adams WT (1991) Genetic variation of wood density components in young coastal Douglas-fir—Implications for tree breeding. *Can J For Res* 21:1801–1807
- Vos P, Hogers R, Bleeker M, Reijans M, Vandelee T, Homes M, Frijters A, Pot J, Peleman J, Kuiper M, Zabeau M (1995) AFLP—a new technique for DNA-fingerprinting. *Nucleic Acids Res* 23:4407–4414
- Wagner FG, Gorman TM, Pratt KL, Keegan CE (2002) Warp, moe, and grade of structural lumber curve sawn from small-diameter Douglas-fir logs. *For Prod J* 52:27–31
- Wheeler NC, Jermstad KD, Krutovsky K, Aitken SN, Howe GT, Krakowski J, Neale DB (2005) Mapping of quantitative trait loci controlling adaptive traits in coastal Douglas-fir. IV. Cold-hardiness QTL verification and candidate gene mapping. *Mol Breed* 15(2):145–156
- Wu RL, Han YF, Hu JJ, Fang JJ, Li L, Li ML, Zeng ZB (2000) An integrated genetic map of *Populus deltoides* based on amplified fragment length polymorphisms. *Theor Appl Genet* 100:1249–1256
- Yin TM, Wang XR, Andersson B, Lerceteau-Kohler E (2003) Nearly complete genetic maps of *Pinus sylvestris* L. (Scots pine) constructed by AFLP marker analysis in a full-sib family. *Theor Appl Genet* 106:1075–1083
- Zhang D, Zhang Z, Yang K, Li B (2004) Genetic mapping in (*Populus tomentosa* × *Populus bolleana*) and *P. tomentosa* Carr. using AFLP markers. *Theor Appl Genet* 108:657–662