*Mol. Cells*, Vol. 20, No. 2, pp. 201-209

**Molecules and Cells**



# An AFLP-based Linkage Map of Japanese Red Pine (*Pinus densiflora*) Using Haploid DNA Samples of Megagametophytes from a Single Maternal Tree

**Yong-Yul Kim*, Hyung-Soon Choi, and Bum-Yong Kang[1]**

*Divison of Genetic Resources, Department of Forest Genetic Resources, Korea Forest Research Institute, Suwon 441-350, Korea;*
*[1] Department of Biology, Life Sciences Centre, Dalhousie University, Halifax, Nova Scotia B3H 4J1, Canada.*

*(Received March 24, 2005; Accepted June 30, 2005)*

**We have constructed an AFLP-based linkage map of Japanese red pine (*Pinus densiflora* Siebold et Zucc.) using haploid DNA samples of 96 megagametophytes from a single maternal tree, selection clone Kyungbuk 4. Twenty-eight primer pairs generated a total of 5,780 AFLP fragments. Five hundreds and thirteen fragments were verified as genetic markers with two alleles by their Mendelian segregation. At the linkage criteria LOD 4.0 and maximum recombination fraction 0.25($\theta$), a total of 152 markers constituted 25 framework maps for 19 major linkage groups. The maps spanned a total length of 2,341 cM with an average framework marker spacing of 18.4 cM. The estimated genome size was 2,662 cM. With an assumption of equal marker density, 82.2% of the estimated genome would be within 10 cM of one of the 230 linked markers, and 68.1% would be within 10 cM of one of the 152 framework markers. We evaluated map completeness in terms of LOD value, marker density, genome length, and map coverage. The resulting map will provide crucial information for future genomic studies of the Japanese red pine, in particular for QTL mapping of economically important breeding target traits.**

**Keywords:** AFLP; Genome Length; Linkage Map; Map Coverage; Megagametophyte; *Pinus densiflora.*

## Introduction

Japanese red pine (*Pinus densiflora* Siebold et Zucc.) is one of the major conifer species in Korea. Efforts to breed the pine started in 1959, focusing on the establishment of a grafted clonal seed orchard (Yim and Noh, 1979). More than 250 clones have been established in the seed orchard as the basic breeding population. The breeding program is primarily based on family selection through progeny testing of open-pollinated half-sib families (Han *et al*., 1986). Molecular breeding work such as quantitative trait loci (QTL) mapping has not yet begun.

Genetic linkage mapping is a key step toward QTL mapping. Reliable QTL mapping depends on the construction of linkage maps with reliable marker order as well as high levels of genome coverage and marker density. In crops, linkage maps have been usually constructed with mapping populations such as inbred lines (Kim *et al*., 2003). In tree species, owing to their long generation time and inbreeding depression, pseudo-testcross mapping populations from single crosses between selected clones have been used (Grattapaglia *et al.*, 1995). However crossing takes a substantial time and needs prior genetic information on the clones, such as marker polymorphism.

Pine has a special feature that enables one to construct linkage maps easily without any mapping population (Wu *et al*., 1999); segregated haploid genomes can be obtained from seeds because the megagametophytes of pine seeds contain haploid genomes inherited from the mother tree. Although the number of haploid genomes in single megagametophytes is usually very limited, there are enough to generate sufficient PCR-based anonymous markers for mapping purposes. Most of the early linkage maps in pines were haploid-based maps constructed with megagametophytes (Nelson *et al*., 1993; 1994; Yazdani *et al*., 1995).

* To whom correspondence should be addressed.
Tel: 82-31-290-1144; Fax: 82-31-290-1040
E-mail: yykim@foa.go.kr

Abbreviations: AFLP, amplified fragment length polymorphism; cM, centiMorgan; ESTP, expressed sequenced tag polymorphism; LOD, log likelihood; PCR, polymerase chain reaction; QTL, quantitative trait loci; RAPD, random amplified polymorphic DNA; SSR, simple sequence repeat.

AFLP, involving a combination of restriction enzyme digestion and PCR-based amplification, is a promising marker system because it has the highest universal applicability, multiplex ratio and reproducibility of all PCR-based dominant marker systems (Myburg *et al*., 2001). Increasing numbers of reports have described the use of AFLP markers for plant genetic mapping, particularly for pine linkage mapping (Costa *et al*., 2000; Lerceteau *et al*., 2000; Remington *et al*., 1999).

In this study, we present an AFLP-based linkage map of Japanese red pine using haploid genomes from a single maternal clone. We also estimate average framework marker spacing, genome length and map coverage. This detailed genetic linkage map of Japanese red pine can serve as a reference map for future genomic research such as QTL mapping of economically important breeding traits.

## Materials and Methods

**DNA sample preparations** Needles and open-pollinated seeds were collected from a ramet of selection clone Kyungbuk 4 in the Anmyeon pine seed orchard. Seeds were germinated on wet filter papers at 28°C for two weeks. Megagametophytes were isolated from the germinating seeds and immediately stored at −25°C. Total DNA was purified from the megagametophytes and needles using a DNeasy Plant Mini Kit (Qiagene). The DNA preparations were quantified by electrophoresing 5 μl of each DNA on 0.8% agarose gels with 0.2 μg/ml ethidium bromide, and visually comparing band intensities with those of known quantities of lambda DNA (Promega). The average quantity of total DNA was 1.5 μg for megagametophytes and 8 μg for needles.

**AFLP procedures** AFLP analysis was performed essentially as described by Remington *et al.* (1999). About 400 ng DNA was used for digestion with *Eco*RI and *Mse*I, and for ligation of adapters. The restriction-ligation (R/L) mixtures were diluted 1:10 with deionized water. Preamplification was carried out using *Eco*RI and *Mse*I primers with two selective nucleotides (E + AC / M + CC). The preamplification solution was also diluted 1:40 with deionized water. Selective amplification was carried out with 28 pairs of E primers with three selective nucleotides (E + 3), and M primers with four selective nucleotides (M + 4). The E + 3 primers were labeled with the fluorescent dye 6-FAM at their 5′-ends. All PCR conditions were as described by Remington *et al.* (1999).

AFLP fragments were resolved on POP4 capillary gel matrices with an ABI Prism 3100 Genetic Analyzer (Applied Biosystems). GeneScan 500-ROX (Applied Biosystems) was used as internal size standard. GeneScan 3.7 (Applied Biosystems) software was used to characterize peak profiles of AFLP fragments, i.e. peak intensity (height), width and size. AFLP fragments with the highest peak intensity over 50 were regarded as fragment presence (coded as ‘A’), and otherwise as fragment absence (coded as ‘a’). A Genotyper 3.7 (Applied Biosystems) was used to categorize the fragments by size and to generate reports of fragment presence/absence strings for each sample. These reports were converted into the input data formats of the mapping software.

**Linkage analysis and map construction** Chi-squared tests were performed to check if individual fragments segregated into the 96 haploid samples in a Mendelian segregation ratio (1:1 ratio for fragment presence versus absence). Only the non-distorted fragments were confirmed as AFLP marker loci with two alleles (*A* and *a*). Linkage analysis was performed using the MAPMAKER/EXP 3.0 program (Lincoln *et al*., 1992). Based on an F2 backcross model, the marker phenotype *A* was genotyped as ‘Aa’, and *a* as ‘aa’. A new set of marker genotype data reciprocally recoded for each locus was also prepared and merged into the original one for detecting coupling- and repulsion-phase linkages.

Linkage groups were assigned with the default linkage criteria of minimum LOD score 4.0 and maximum recombination fraction ($\theta$) 0.3. For each linkage group, an initial set of informative markers was selected using ‘order’ and ‘compare’ commands. The order of the selected markers was confirmed using the ‘ripple’ command with a LOD score of 3.0 as exclusion threshold. Thereafter each marker-ordered linkage group was defined as a ‘framework map’. Markers with low ordering confidence were placed into each linkage group as accessory markers using the ‘place’ command. The linkage maps were generated with the ‘map’ command using Kosambi’s map function.

**Marker distribution** The marker distribution between linkage groups was evaluated by comparing marker density with expectations from the Poisson distribution (Remington *et al.*, 1999). We assumed that each linkage group had a length $G_i$ (= $M_i$ + 2s), where $M_i$ is the map distance between terminal markers of the $i_{th}$ linkage group, and $s$ is the average framework marker spacing calculated by dividing the summed length of all maps by the number of marker intervals. Under a uniform probability distribution for marker location, $s$ is the expected distance from a terminal marker to the chromosome end. If the underlying marker density were the same for all linkage groups, the number of markers ($m_i$) in the $i_{th}$ linkage group would be a sample from a Poisson distribution with parameter, $\lambda_i = mG_i / \Sigma_i G_i$, where $m$ is the total number of markers. We used a two-tailed test for the null hypothesis, i.e. equal marker density in linkage groups.

**Genome length and map coverage** Genome length was estimated using the method-of-moment estimator, $E(G) = n(n\text{-}1)\, X/K$, where $n$ is the number of loci, $X$ is the maximum map distance between locus pairs above a minimum LOD value for linkage grouping ($Z$), and $K$ is the number of locus pairs above $Z$ (Hulbert *et al*., 1988). For three $Z$ values (3, 4 and 5), both $X$ and $K$ were obtained from the list of values generated by the ‘big lods’ command of the MAPMAKER/EXP program. The confidence interval for the estimated genome length was calculated using

**Table 1.** Number of AFLPs generated with the 28 primer combinations.

| Primer combination | | | | Total scored bands | Non-maternal bands | Maternal bands | | | |
|---|---|---|---|---|---|---|---|---|---|
| E[a] + 3 | M[b] + 4 | Name | CpG[c] | | | Total | Mono[d] | Poly[e] | Marker |
| ACA | CCAA | AAA | 0 | 249 | 70 | 179 | 2 | 177 | 18 |
| | CCAG | AAG [f, g, h] | 0 | 172 | 31 | 141 | - | 141 | 15 |
| | CCCG | ACG [f, g, h] | 1 | 170 | 60 | 110 | 7 | 103 | 12 |
| | CCGC | AGC [f, g, h] | 1 | 243 | 88 | 155 | 4 | 151 | 11 |
| | CCGG | AGG [f, g, h] | 1 | 200 | 86 | 114 | 3 | 111 | 17 |
| ACC | CCAA | CAA | 0 | 286 | 47 | 239 | 1 | 238 | 23 |
| | CCAC | CAC | 0 | 10 | 3 | 7 | 0 | 7 | 0 |
| | CCAG | CAG [f, g] | 0 | 213 | 58 | 155 | 4 | 151 | 17 |
| | CCGA | CGA | 1 | 204 | 102 | 102 | 3 | 99 | 4 |
| | CCTG | CTG | 0 | 195 | 41 | 154 | 3 | 151 | 25 |
| | CCTT | CTT | 0 | 227 | 24 | 203 | 5 | 198 | 31 |
| ACG | CCAA | GAA [f, g, h] | 1 | 203 | 64 | 139 | 15 | 124 | 21 |
| | CCAG | GAG [f, g, h] | 1 | 219 | 126 | 93 | 1 | 92 | 13 |
| | CCCA | GCA [f, g, h] | 1 | 172 | 64 | 108 | 7 | 101 | 16 |
| | CCGA | GGA [f, g, h] | 2 | 208 | 101 | 107 | 2 | 105 | 13 |
| | CCGC | GGC [f, g, h] | 2 | 229 | 81 | 148 | 2 | 146 | 16 |
| | CCTA | GTA [f, g, h] | 1 | 122 | 34 | 88 | 3 | 85 | 13 |
| | CCTC | GTC [f, g, h] | 1 | 174 | 60 | 114 | 3 | 111 | 18 |
| | CCTG | GTG [f, g, h] | 1 | 166 | 87 | 79 | 1 | 78 | 18 |
| | CCTT | GTT [f, g, h] | 1 | 145 | 35 | 110 | 7 | 103 | 20 |
| ACT | CCAC | TAC | 0 | 250 | 74 | 176 | 12 | 164 | 27 |
| | CCAG | TAG [f] | 0 | 191 | 47 | 144 | 3 | 141 | 25 |
| | CCCG | TCG [f, g, h] | 1 | 276 | 105 | 171 | 5 | 166 | 19 |
| | CCGC | TGC [f, g, h] | 1 | 246 | 71 | 175 | 5 | 170 | 14 |
| | CCGG | TGG [f, g, h] | 1 | 252 | 96 | 156 | 4 | 152 | 19 |
| | CCTA | TTA | 0 | 265 | 46 | 219 | 17 | 202 | 38 |
| | CCTC | TTC | 0 | 216 | 33 | 183 | 3 | 180 | 28 |
| | CCTT | TTT | 0 | 277 | 58 | 219 | 0 | 219 | 22 |
| Total | | 28 | | 5,780 | 1,792 | 3,988 | 122 | 3,866 | 513 |

[a] Core sequence of *Eco*RI adaptor primer, 5′-GAC TGC GTA CCA ATT C-3′.

[b] Core sequence of *Mse*I adaptor primer, 5′-GAT GAG TCC TGA GTA A-3′.

[c] Number of CpGs in the selective nucleotides of the E + 3 and M + 4 primers.

[d] Number of bands showing monomorphism in the megagametophytes and also detected in the maternal needle sample.

[e] Number of bands showing polymorphism in the megagametophytes and also detected in the maternal needle sample.

[f, g, h] Primers also used in *P. elliottii* (f; Shepherd *et al.*, 2003), *Pinus taeda* (g; Remington *et al*., 1999), and *P. sylvestris* (h; Komulainen *et al*., 2003).

the equation, $I_\alpha(G) = E(G)/(1 \pm n_\alpha/\sqrt{K})$, where $n_\alpha = 1.96$ for $\alpha = 0.05$ (Gerber and Rodolphe, 1994). We also calculated a modified estimator of genome length ($L_e$) to correct Hulbert *et al*.'s estimate for the upward bias related to chromosome ends (Remington *et al.*, 1999):

$L_e = \frac{n(n-1)X}{2K}\left(1 + \sqrt{1 - \frac{2NK}{n(n-1)}}\right)$, where $N$ is the number of haploid chromosome ($N = 12$).

With the assumption of random marker distribution, the proportion of the genome within $d$ cM of a marker was estimated using the formula, $C = 1 - e^{-2dn/L}$, where $C$ is genome coverage, $L$ is the estimated genome length and $n$ is the number of markers (Lange and Boehnke, 1982; Remington *et al.*, 1999).

## Results

**AFLP marker detection** Twenty-eight primer pairs generated a total of 5,780 fragments, of which 3,988 (69%) were maternal bands detected in both the haploid and diploid samples, and 1,792 (31%) were non-maternal bands detected only in the haploid samples (Table 1). Of the maternal bands, 3,866 (96.9%) were polymorphic, and 122 (3.1%) were monomorphic. A chi-squared test showed that 13.2% (513 bands) of the polymorphic bands gave Mendelian segregation ratios, and they were confirmed as AFLP marker loci with two alleles, *A* and *a*. By including the monomorphic bands as AFLP markers (representing dominant homozygotes in the diploid sample), we identi-

fied a total of 635 markers for selection clone Kyungbuk 4. On average, each AFLP primer pair generated 22.7 markers (18.3 polymorphic markers). The primer pairs, CAA (E + ACC / M + CCAA) and TTA (E + ACT / M + CCTA) gave the largest number of total scored bands and AFLP markers, respectively. Conversely, CAC (E + ACC / M + CCAC) generated the smallest number of both. There was a significant negative correlation between number of scored markers and CpG dinucleotides in the selective regions of the E + 3 and M + 4 primers ($r = -0.613$; $p = 0.0007$).

**Linkage analysis and framework map construction** At the linkage criteria of LOD 4.0 and $\theta = 0.3$, 322 markers were distributed into 40 linkage groups. We classified the 40 linkage groups into two categories according to the number of their linked markers, i.e. *major* (≥ 5 markers) and *minor* groups (Zhang *et al*., 2004). Nineteen groups belonged to the major group, and 21 to the minor group (4 quadruplets, 6 triplets and 11 doublets). A total of 266 markers were assigned to the 19 groups (Table 2). The number of linked markers ranged from 5 (LG 18 and 19) to 38 (LG 1). On average, 14 markers were present in each group. At the more strict linkage criteria of LOD 4.0 and $\theta = 0.25$, eight of the 19 groups (LG 1 to LG 8) were split into 2–5 subgroups; a total of 34 subgroups contained 254 markers, *i.e.* 25 subgroups (≥ 4 markers), 6 triplets and 3 doublets. None of the 25 subgroups, except LG 1-2, LG 3-1 and 3-2, were further subdivided at LOD 5.0 and $\theta = 0.25$. A total of 230 markers made up the 25 subgroups, and only a small fraction of the 254 markers (9.4%) were excluded. These results imply that the linkage criteria of LOD 4.0 and $\theta = 0.25$ provide an optimal threshold for generating informative linkage units, i.e. the 25 subgroups.

We constructed framework maps from the 25 subgroups. LOD 3.0 was used as confidence value to verify the order of three adjacent makers randomly chosen from any given subgroup, and also to exclude markers with low confidence from the framework maps. A total of 152 markers eventually made up the 25 framework maps (Table 2, Fig. 1), and the remaining 78 markers were attached to each map as accessory markers. The map length varied from 46.8 cM (LG 19) to 203.8 cM (LG 5-1). Total map distance was 2,341 cM, and the average framework marker spacing was 18.4 cM. Each of the 6 major linkage groups (LG 1, 2, 3, 4, 6 and 8) included 2 framework maps (Table 2). Using the "try" command of the MAPMAKER/EXP software, we checked if the two maps in each of these groups were joined together with an interval support of LOD 3.0. The two maps of LG 6 (LG 6-1 and 6-2) were indeed successfully joined in this way with a 28.9 cM marker interval (Fig. 1). Those of four other linkage groups (LG 1, 2, 3 and 8) were also combined, but with larger gaps (40 to 70 cM). The combined 20 framework maps spanned a total length of 2,593 cM with an average framework marker spacing of 19.8 cM.

**Genome length and map coverage** The genome lengths estimated according to Hulbert *et al*. (1988) were 2,113, 2,870 and 3,386 cM for Z values of 3, 4 and 5, respectively (Table 3). The average genome length was 2,790 cM with a confidence interval of 2,626– 2,975 cM, which was close to the estimate under $Z = 4$ for declaring the 25 framework maps. However, the modified genome estimate obtained by Remington *et al.*'s method was 2,662 cM with a confidence interval of 2,506–2,839 cM. These estimates were smaller than those derived by the method of Hulbert *et al*. where the difference between the two estimates was only about 130 cM.

A two-tailed test showed that there was no significant difference in marker density between the 25 linkage groups at $\alpha = 0.05$ (Table 4). Using a 20 cM map scale (an approximation of the average framework marker spacing of 18.4 cM) and assuming that 24 of the 50 ends of the 25 linkage groups are located within 10 cM (= 20 cM/2) of the 24 telomeres ($N = 12$), the 230 linked markers would cover 82.2% of the genome ($L = 2{,}662$ cM) at an even spacing of 10 cM between adjacent markers. However, using a 30 cM map scale (an approximation to the maximum map distance at $Z = 4$, 31.8 cM), 92.5% of the genome would be within 15 cM of one of the 230 linked markers. On the other hand, using only the 152 framework markers, 68.1% of the genome would be within 10 cM of a framework marker, and 82.0% within 15 cM.

## Discussion

**The AFLP primers and markers generated in Japanese red pine** Screening the primers for many reproducible markers is a prerequisite for any successful AFLP analysis. Remington *et al.* (1999) reported that CpGs in the selective regions of the E and M primers tended to result in fewer markers. To ascertain the effect of the CpGs, it is necessary to obtain adequate AFLP data on CpGs and markers generated by primers. We performed a correlation analysis using the data for *P. sylvestris* (Komulaiden *et al*., 2003) and *P. elliottii* (Shepherd *et al.*, 2003), in which the PCR conditions and pre-amplification primers were the same as those of both Remington *et al.* (1999) and our study. There was a significant negative correlation between CpGs and markers generated in *P. densiflora* and *P. taeda* (Table 5). In addition, the larger number of CpGs in the M + 4 primer tended to generate fewer markers across all species. These results indicate that the base composition of the M + 4 primer selective extensions has a significant effect on the number of segregating AFLP fragments.

PCR-based dominant markers have been criticized for their lack of transferability to different genetic back-

**Table 2.** Summary for the 19 major linkage groups and 25 framework maps.

| Linkage group LOD 4, θ = 0.3 | | Linkage group LOD 4, θ = 0.25 | | Linkage group LOD 5, θ = 0.25 | | Framework map (with the internal support of LOD 4.0) | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | 25 framework maps | | | Combined 20 maps | | |
| Group | Linked marker | Sub-group | Linked marker | Sub-group | Linked marker | Map | Marker | cM | Map | Marker | cM |
| LG 1 | 38 | LG 1-1 | 17 | LG 1-1 | 17 | LG 1-1 | 8 | 137.7 | LG 1C | 14 | 282.4 |
| | | LG 1-2 | 13 | LG 1-2-1 | 10 | LG 1-2 | 7 | 89.3 | | | *(78)** |
| | | | | LG 1-2-2 | 3 | | | | | | |
| | | LG 1-3 | 3 | LG 1-3 | 2 | | | | | | |
| | | LG 1-4 | 3 | LG 1-4 | 3 | | | | | | |
| | | LG 1-5 | 2 | LG 1-5 | 2 | | | | | | |
| LG 2 | 30 | LG 2-1 | 20 | LG 2-1 | 19 | LG 2-1 | 10 | 145.2 | LG 2C | 14 | 269.1 |
| | | LG 2-2 | 10 | LG 2-2 | 9 | LG 2-2 | 4 | 80.9 | | | *(43)** |
| LG 3 | 24 | LG 3-1 | 8 | LG 3-1-1 | 3 | LG 3-1 | 4 | 66.2 | LG 3C | 8 | 213.2 |
| | | | | LG 3-1-2 | 2 | | | | | | *(71)** |
| | | | | LG 3-1-3 | 2 | | | | | | |
| | | LG 3-2 | 5 | LG 3-2 | 5 | LG 3-2 | 4 | 76.0 | | | |
| | | LG 3-3 | 3 | LG 3-3 | 2 | | | | | | |
| | | LG 3-4 | 3 | LG 3-4 | 2 | | | | | | |
| | | LG 3-5 | 2 | LG 3-5 | 2 | | | | | | |
| LG 4 | 21 | LG 4-1 | 11 | LG 4-1 | 11 | LG 4-1 | 9 | 173.7 | LG 4-1 | 9 | 173.7 |
| | | LG 4-2 | 5 | LG 4-2 | 5 | LG 4-2 | 4 | 65.7 | LG 4-2 | 4 | 65.7 |
| | | LG 4-3 | 3 | LG 4-3 | 3 | | | | | | |
| LG 5 | 21 | LG 5-1 | 17 | LG 5-1 | 17 | LG 5-1 | 12 | 203.8 | LG 5-1 | 12 | 203.8 |
| | | LG 5-2 | 2 | LG 5-2 | 2 | | | | | | |
| LG 6 | 17 | LG 6-1 | 11 | LG 6-1 | 9 | LG 6-1 | 6 | 77.6 | LG 6C | 10 | 160.3 |
| | | LG 6-2 | 6 | LG 6-2 | 6 | LG 6-2 | 4 | 53.7 | | | *(29)** |
| LG 7 | 17 | LG 7-1 | 13 | LG 7-1 | 13 | LG 7-1 | 8 | 129.1 | LG 7-1 | 8 | 129.1 |
| | | LG 7-2 | 3 | LG 7-2 | 3 | | | | | | |
| LG 8 | 11 | LG 8-1 | 5 | LG 8-1 | 4 | LG 8-1 | 5 | 78.8 | LG 8C | 9 | 185.4 |
| | | LG 8-2 | 5 | LG 8-2 | 5 | LG 8-2 | 4 | 52.6 | | | *(54)** |
| LG 9 | 11 | LG 9 | 10 | LG 9 | 10 | LG 9 | 5 | 52.8 | LG 9 | 5 | 52.8 |
| LG 10 | 10 | LG 10 | 10 | LG 10 | 10 | LG 10 | 5 | 83.2 | LG 10 | 5 | 83.2 |
| LG 11 | 10 | LG 11 | 10 | LG 11 | 10 | LG 11 | 10 | 160.5 | LG 11 | 10 | 160.5 |
| LG 12 | 9 | LG 12 | 8 | LG 12 | 8 | LG 12 | 7 | 107.3 | LG 12 | 7 | 107.3 |
| LG 13 | 8 | LG 13 | 8 | LG 13 | 8 | LG 13 | 6 | 85.1 | LG 13 | 6 | 85.1 |
| LG 14 | 8 | LG 14 | 8 | LG 14 | 8 | LG 14 | 6 | 86.8 | LG 14 | 6 | 86.8 |
| LG 15 | 8 | LG 15 | 8 | LG 15 | 8 | LG 15 | 7 | 101.2 | LG 15 | 7 | 101.2 |
| LG 16 | 7 | LG 16 | 6 | LG 16 | 6 | LG 16 | 4 | 52.6 | LG 16 | 4 | 52.6 |
| LG 17 | 6 | LG 17 | 6 | LG 17 | 6 | LG 17 | 4 | 71.9 | LG 17 | 4 | 71.9 |
| LG 18 | 5 | LG 18 | 5 | LG 18 | 5 | LG 18 | 5 | 62.3 | LG 18 | 5 | 62.3 |
| LG 19 | 5 | LG 19 | 5 | LG 19 | 5 | LG 19 | 4 | 46.8 | LG 19 | 4 | 46.8 |
| 19 | 266 | 34 | 254 | 37 | 245 | 25 | 152 | 2,341 | 20 | 151 | 2,593 |
| Average | 14 | | 7.5 | | 6.6 | | 6.1 | 93.6 | | 7.6 | 129.7 |

* Interval (gap) between the markers joining two framework maps.

grounds making it difficult to construct intraspecific consensus linkage maps (Gosselin *et al*., 2002; Jones *et al*., 1997). For example, Kubisiak *et al*. (1996) reported that a large proportion of the markers assigned in haploid-based maps were not detected in the diploid mapping population. We also found that 31% of the total AFLP bands were not inherited from the mother tree, and that about 8% of the non-maternal bands gave Mendelian segregation ratios. Co-utilization of codominant markers such as SSR has been suggested to overcome the limitations of dominant

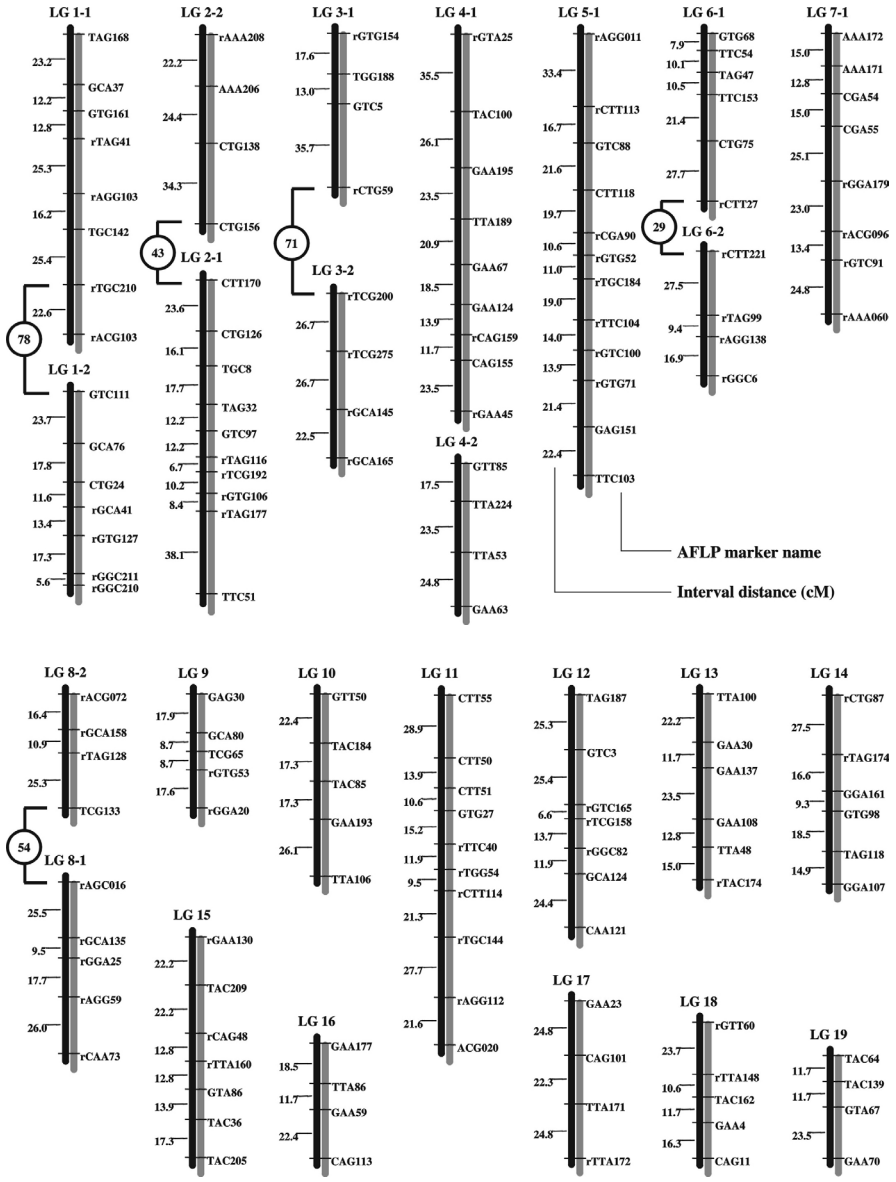


**Fig. 1.** Framework maps of *Pinus densiflora*. The cM function is that of Kosambi. Intervals are shown on the left side of the maps, and marker names on the right. Marker names starting with r are in reverse linkage phase to those not so designated. The numerals circled in bold are the marker intervals (cM) obtained when two maps were joined together using the 'Try' command in MAPMAKER/EXP.

markers in haploid-based mapping (Gosselin *et al*., 2002; Kubisiak *et al*., 1996). However, only a few pines have so far provided such markers (Harry *et al*., 1998; Komulainen *et al*., 2003; Paglia *et al*., 1998). In this study we used an alternative method, called 'crosschecking', to obtain transferable AFLP markers, i.e. maternal markers. By simply comparing all AFLPs in both haploid and diploid samples, we were able to exclude all non-maternal markers from our framework maps. We also suggest using 'reference DNA' for linkage mapping, that is, a set of DNA samples from many clones of the Japanese red pine, and from other pines already used in genomic studies. By using the reference DNA to screen primers, one could establish an AFLP database that would help to identify intra- and inter-specifically transferable AFLP markers in pine species. These seem to be the most practical ways to compensate for the lack of transferability of dominant markers, and so to enhance the utility of AFLP markers in haploid-based mapping.

**Map construction, map coverage and genome length**

There are many ways to evaluate the completeness of a given linkage map. One is to examine the LOD value for declaring linkage group and ordering markers. In general, the LOD values used in conifer mapping studies have ranged from 3.0 to 5.0 (Gosselin *et al*., 2002). We used

**Table 3.** Genome length, *E*(G), estimated by Hulbert *et al*.'s method-of-moment estimator. *n* is the number of linked markers, *K* is the number of locus pairs and *X* is the maximum map distance corresponding to the minimum LOD score (*Z*) for declaring linkage.

| Z value | *n* | *K* | *X* | *E(G)* | Confidence interval |
|---|---|---|---|---|---|
| 3 | 300 | 1,560 | 36.7 | 2,113 | 2,012–2,223 |
| 4 | 299 | 988 | 31.8 | 2,870 | 2,701–3,061 |
| 5 | 289 | 685 | 27.9 | 3,386 | 3,151–3,661 |
| Average | | | | 2,790 | 2,626–2,975 |
| $L_e$* | | | | 2,662 | 2,506–2,839 |

* Genome length estimated by Remington *et al.*'s method (see **Materials and Methods**).

**Table 4.** Tests for random marker distribution among the 25 linkage groups.

| Linkage group | Number of markers ($m_i$) | Map length ($M_i$) | Inferred map length ($G_i$) | Expected number of markers ($\lambda_i$) | Poisson two-tailed *P*-value[a] |
|---|---|---|---|---|---|
| LG 1-1 | 17 | 137.7 | 174.5 | 12.31 | 0.076 |
| LG 1-2 | 13 | 89.3 | 126.1 | 8.89 | 0.069 |
| LG 2-1 | 20 | 145.2 | 182.0 | 12.83 | 0.026 |
| LG 2-2 | 10 | 80.9 | 117.7 | 8.30 | 0.215 |
| LG 3-1 | 8 | 66.2 | 103.0 | 7.27 | 0.306 |
| LG 3-2 | 5 | 76.0 | 112.8 | 7.96 | 0.195 |
| LG 4-1 | 11 | 173.7 | 210.5 | 14.85 | 0.195 |
| LG 4-2 | 5 | 65.7 | 102.5 | 7.23 | 0.272 |
| LG 5-1 | 17 | 203.8 | 240.6 | 16.97 | 0.433 |
| LG 6-1 | 11 | 77.6 | 114.4 | 8.07 | 0.117 |
| LG 6-2 | 6 | 53.7 | 90.5 | 6.38 | 0.545 |
| LG 7-1 | 13 | 129.1 | 165.9 | 11.70 | 0.287 |
| LG 8-1 | 5 | 78.8 | 115.6 | 8.15 | 0.178 |
| LG 8-2 | 5 | 52.6 | 89.4 | 6.31 | 0.398 |
| LG 9 | 10 | 52.8 | 89.6 | 6.32 | 0.057 |
| LG 10 | 10 | 83.2 | 120.0 | 8.46 | 0.233 |
| LG 11 | 10 | 160.5 | 197.3 | 13.92 | 0.181 |
| LG 12 | 8 | 107.3 | 144.1 | 10.16 | 0.315 |
| LG 13 | 8 | 85.1 | 121.9 | 8.60 | 0.510 |
| LG 14 | 8 | 86.8 | 123.6 | 8.72 | 0.493 |
| LG 15 | 8 | 101.2 | 138.0 | 9.73 | 0.364 |
| LG 16 | 6 | 52.6 | 89.4 | 6.31 | 0.557 |
| LG 17 | 6 | 71.9 | 108.7 | 7.67 | 0.356 |
| LG 18 | 5 | 62.3 | 99.1 | 6.99 | 0.302 |
| LG 19 | 5 | 46.8 | 83.6 | 5.90 | 0.462 |

[a] Poisson probability of having as many (for $m_i > \lambda_i$) or as few (for $m_i < \lambda_i$) markers as the observed number, $m_i$ in the $i_{th}$ linkage group under the null hypothesis that the average framework marker spacing (18.4 cM) is the same for all linkage groups. In a two-tailed test, a p-value of 0.025 corresponds to a significance level of 0.05.

LOD 4.0 for linkage grouping because that value provided an adequate number of linkage groups most of which were not further divided at LOD 5.0. Also, using LOD 3.0 for ordering markers, we constructed 25 linkage maps, in which all the framework markers were in the most likely order without any distortion. The number of linked and framework markers generated at various LOD values may be one method of evaluation. We obtained a total of 254 linked markers at LOD 4.0, and assigned 152 framework markers to 25 maps with a LOD interval support of 3.0. Our maps seem to have similar numbers of markers as those for *P. sylvestris* (Yin *et al.*, 2003) and *P. taeda* (Remington *et al.*, 1999).

An ideal genetic linkage map covers a large proportion of the genome, and has the same number of linkage groups as there are haploid chromosomes. Unlike LOD value and number of markers, these seem to be qualitative specifications. We obtained 25 framework maps with a total map length of 2,341 cM, and with 82% coverage of the genome within 10 cM of one of the 230 markers. These 25 maps cover a larger map distance, but a lower proportion of the genome, than the maps of *P. sylvestris* obtained with similar AFLP primers (1,696–1,759 cM, 95% of the genome within 10 cM of one of 300 markers) (Yin *et al*., 2003). This implies that our framework maps have a lower marker density, and therefore that more AFLP markers should be added to the maps in order to obtain nearly saturated maps equivalent to those of *P. sylvestris* (Yin *et al*., 2003).

The haploid chromosomes number of the genus *Pinus* is 12. We obtained 21 unassembled minor linkage groups in addition to the 25 subgroups of the 19 major linkage groups. In fact, all the maps so far constructed have more than 12 linkage groups except those of *P. pinaster* (Costa *et al.*, 2000; Plomion *et al.*, 1995) and *P. taeda* (Remington *et al.*, 1999). More efforts should be made to incorporate the unassembled linkage groups into the 25 maps and to join the framework maps, so as to construct a nearly complete linkage map of the Japanese red pine. We expect that such map assembly can be accomplished with orthologous anchor markers such as the species-transferable ESTPs recently developed in pines (Chagné *et al*., 2004).

In published linkage maps for *Pinus* species, the genome length ranges from 1,300 to 3,000 cM (Costa *et al*., 2000; Nelson *et al*., 1994; Plomion *et al*., 1995). In *P. sylvestris*, which is most closely related to the Japanese red pine (Szmidt and Wang, 1993), genome length varied from 1,600 to 2,500 cM (Komulainen *et al*., 2003; Lerceteau *et al*., 2000; Yin *et al*., 2003). Our estimated genome length based on the two methods was 2,662–2,790 cM, which is apparently larger than that of *P. sylvestris*. This discrepancy may be due to the choice of different linkage criteria and genome length estimators, or to intraspecific differences in recombination rates. According to cytological studies (Rake *et al*., 1980; Wakamiya *et*

**Table 5.** Pearson's simple correlation between CpGs and markers by four pine species.

| CpG source[a] | Total | *P. densiflora* | *P. taeda* | *P. sylvestris* | *P. elliottii* |
|---|---|---|---|---|---|
| Both of E + 3 / M + 4 | −0.317*** | −0.613*** | −0.617*** | $-0.132^{N.S}$ | $0.099^{N.S}$ |
| E + 3 only | $0.035^{N.S}$ | $-0.264^{N.S}$ | $-0.149^{N.S}$ | $0.325^{N.S}$ | $0.278^{N.S}$ |
| M + 4 only | −0.405*** | −0.527*** | −0.558** | $-0.450^{N.S}$ | $-0.169^{N.S}$ |

[a] The data for scored markers and primer pairs were obtained from Remington *et al*. (1999) for *P. taeda*, Komulainen *et al*. (2003) for *P. sylvestris*, and Shepherd *et al.* (2003) for *P. elliottii*.

*, **, *** Significance level for $\alpha$ = 0.05, 0.01 and 0.005, respectively.

$^{N.S}$ Non-significant.

*al*., 1993), the pine genome is composed of 10,000–20,000 mega bases (Mb). Provided our estimate is close to the real genome length of the Japanese red pine, one centiMorgan (cM) should equal 3.5–7.6 Mb. This indicates that the recombination frequency is very low per unit length of DNA in the Japanese red pine, so that many sets of mapping populations and larger sample sizes will be needed to obtain reliable QTL mapping results.

**Map utility** The main purpose of map construction must be its application to QTL mapping for breeding traits. Thus a linkage map is one of the end products of QTL mapping, not itself the final goal. In addition, since most QTL work in tree species has employed pseudo-testcross mapping methods using diploid mapping population from crosses between selected clones (Grattapaglia *et al.*, 1995) haploid-based linkage mapping may not be useful any more in this work. The method relies greatly on the assumption that the parent tree is heterozygous at many marker loci, so that the progeny from a single cross between clones would have enough marker polymorphism and phenotypic variation to identify the QTLs. As it is, most of the pseudo-testcross mapping populations have lower levels of marker polymorphism than one might expect. Thus, considerable efforts should be made to identify the best full-sib family with the highest level of polymorphism.

Grattapaglia *et al*. (1996) suggested the half-sib-based QTL mapping, a modification of the 'daughter design' widely used in domestic animals (Georges *et al*., 1995), may lead to a re-evaluation the utility of haploid-based linkage maps (Wu *et al.*, 1999). The Korea Forest Research Institute has recently established a large set of half-sib mapping populations for the Japanese red pine. This consists of 15 full-sib families with a common female parent, selection clone Kyungbuk 4. We are currently genotyping 15 parental clones using the 513 AFLP markers making up the 25 framework maps. The results should greatly help us to select the full-sib families with the highest marker polymorphism for pseudo-testcross QTL mapping. Since the parental maps will be newly constructed from the selected families, our haploid-based framework maps will be utilized as reference for evaluating them. At present, we are cultivating 96 open-pollinated pine seedlings originating from the seeds that yielded the megagametophytes used in this study. If the QTL analysis uses our framework maps together with ongoing data on the growth of the seedlings, the analysis should provide basic information on the breeding values of the QTL markers for growth traits of Japanese red pine.

## References

Chagné, D., Chaumeil, P., Ramboer, A., Collada, C., Guevara, A., *et al*. (2004) Cross-species transferability and mapping of genomic and cDNA SSRs in pines. *Theor. Appl. Genet.* **109**, 1204–1214.

Costa, P., Pot, D., Dubos, C., Frigerio, J. M., Pionneau, C., *et al*. (2000) A genetic map of maritime pine based on AFLP, RAPD and protein markers. *Theor. Appl. Genet.* **100**, 39–48.

Georges, M., Nielsen, D., Mackinnon, M., Mishra, A., Okimoto, R., *et al*. (1995) Mapping quantitative trait loci controlling milk production in dairy cattle by exploiting progeny testing. *Genetics* **139**, 907–920.

Gerber, S. and Rodolphe, F. (1994) An estimation of the genome length of maritime pine (*Pinus pinaster* Ait.). *Theor. Appl. Genet.* **88**, 289–292.

Gosselin, I., Zhou, Y., Bousquet, J., and Isabel, N. (2002) Megagametophyte-derived linkage maps of white spruce (*Picea glauca*) based on RAPD, SCAR and ESTP markers. *Theor. Appl. Genet.* **104**, 987–997.

Grattapaglia, D., Bertolucci, F. L. G., and Sederoff, R. R. (1995) Genetic mapping of QTLs controlling vegetative propagation in *Eucalyptus grandis* and *E. urophylla* using a pseudo-testcross mapping strategy and RAPD markers. *Theor. Appl. Genet.* **90**, 933–947.

Grattapaglia, D., Bertolucci, F. L. G., Penchel, R., and Sederoff, R. R. (1996) Genetic mapping of quantitative trait loci controlling growth and wood quality traits in *Eucalyptus grandis* using a maternal half-sib family. *Genetics* **144**, 1205–1214.

Han, S. U., Kwon, H. M., Jhun, G. S., Sohn, S. I., and Hyun, J. O. (1986) Heritability of Japanese red pine (*Pinus densiflora* S. et Z.) estimated over age and site. Proceedings of the IUFRO-Conference on Breeding Theory, Progeny Testing and Seed Orchards: Williamsburg, Virginia (USA), pp. 71–84. Published by the North Carolina State University-Industry Cooperative Tree Improvement Program.

Harry, D. E., Temesgen, B., and Neale, D. B. (1998) Codomi-

nant PCR-based markers for *Pinus taeda* developed from mapped cDNA clones. *Theor. Appl. Genet.* **97**, 327–336.

Hulbert, S., Ilott, T., Legg, E. J., Lincoln, S., Lander, E., *et al.* (1988) Genetic analysis of the fungus, *Bremia lactucae*, using restriction length polymorphsim. *Genetics* **120**, 947–958.

Jones, C. J., Edwards, K. J., Cataglione, S., Winfield, M. O., Sala, F., *et al*. (1997) Reproducibility testing of RAPD, AFLP and SSR markers in plants by a network of European laboratories. *Mol. Breeding* **3**, 381–390.

Kim, K. M., Kwon, Y. S., Lee, J. J., Eun, M. Y., and Sohn, J. K. (2003) QTL mapping and molecular marker analysis for the resistance of rice to ozone. *Mol. Cells* **17**, 151–155.

Komulainen, P., Brown, G. R., Mikkonen, M., Karhu, A., Garcia-Gil, M. R., *et al.* (2003) Comparing EST-based genetic maps between *Pinus sylvestris* and *Pinus taeda*. *Theor. Appl. Genet.* **107**, 667–678.

Kubisiak, T. L., Nelson, C. D., Nance, W. L., and Stine, M. (1996) Comparisons of RAPD linkage maps constructed for a single longleaf pine from haploid and diploid mapping populations. *Forest Genet.* **3**, 203–211.

Lange, K. and Boehnke, M. (1982) How many polymorphic genes will it take to span the human genome? *Am. J. Hum. Genet.* **24**, 842–845.

Lerceteau, E., Plomion, C., and Andersson, B. (2000) AFLP mapping and detection of quantitative trait loci (QTLs) for economically important traits in *Pinus sylvestris*: a preliminary study. *Mol. Breeding* **6**, 451–458.

Lincoln, S., Daly, M., and Lander, E. (1992) Construction genetic linkage maps with MAPMAKER/EXP 3.0, Whitehead Institute Technical Report, 3rd ed., Whitehead Institute, Cambridge, Massachusetts.

Myburg, A. A., Remington, D., O'Malley, D. M., Sederoff, R. R., and Whetten, R. W. (2001) High-throughput AFLP analysis using infrared dye-labeled primers and an automated DNA sequencer. *BioTechniques* **30**, 348–357.

Nelson, C. D., Nance, W. L., and Doudrick, R. L. (1993) A partial genetic linkage map of slash pine (*Pinus elliottii* Enghelm. var. *elliottii*) based on random amplified polymorphic DNAs. *Theor. Appl. Genet.* **87**, 145–151.

Nelson, C. D., Kubisiak, T. L., Stine, M., and Nance, W. L. (1994) A genetic linkage map of longleaf pine (*Pinus paulstris* Mill) based on random amplified polymorphic DNAs. *J. Hered.* **85**, 433–439.

Paglia, G. P., Olivieri, A. M., and Morgante, M. (1998) Towards second-generation STS (sequence-tagged sites) linkage maps in conifers: a genetic map of Norway spruce (*Picea abies* K.). *Mol. Gen. Genet.* **258**, 466–478.

Plomion, C., Bahrman, N., Durel, C. E., and O'Malley, D. M. (1995) Genomic mapping in *Pinus pinaster* (maritime pine) using RAPD and protein markers. *Heredity* **74**, 661–668.

Rake, A. V., Miksche, J. P., Hall, R. B., and Hansen, K. M. (1980) DNA reassociation kinetics of four conifers. *Can. J. Genet. Cytol.* **22**, 69–79.

Remington, D. L., Whetten, R. W., Liu, B. H., and O'Malley, D. M. (1999) Construction of an AFLP genetic linkage map with nearly complete genome coverage in *Pinus taeda*. *Theor. Appl. Genet.* **98**, 1279–1292.

Shepherd, M., Cross, M., Dieters, M. J., and Henry, R. (2003) Genetic maps for *Pinus elliottii* and *P. caribaea* var. *hondurensis* using AFLP and microsatellite markers. *Theor. Appl. Genet.* **106**, 1409–1419.

Szmidt, A. E. and Wang, X. R. (1993) Molecular systematics and genetic differentiation of *Pinus sylvestris* (L.) and *P. densiflora* (Sieb. et Zucc.). *Theor. Appl. Genet.* **86**, 159–165.

Wakamiya, I., Newton, R. J., Johnston, J. S., and Price, H. J. (1993) Genome size and environmental factors in the genus *Pinus*. *Am. J. Bot.* **80**, 1235–1241.

Wu, R. L., O'Malley, D. M., and McKeand, S. E. (1999) Understanding the genetic architecture of a quantitative trait in gymnosperms by genotyping haploid megagametophytes. *Theor. Appl. Genet.* **99**, 1031–1038.

Yazdani, R., Yeh, F. C., and Rimsha, J. (1995) Genome mapping of *Pinus sylvestris* (L.) using random amplified polymorphic DNA markers. *Forest Genet.* **2**, 109–116.

Yim, K. B. and Noh, E. R. (1979) Study on the Heritability of *Pinus densiflora* S. et Z. (I). *Korean J. For. Sci.* **42**, 74–82.

Yin, T. M., Wang, X. R., Andersson, B., and Lerceteau-Köhler, 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., and Li, B. (2004) Genetic mapping in (*Populus tomentosa* x *Populus bolleana*) and *P. tomentosa* Carr. using AFLP markers. *Theor. Appl. Genet.* **108**, 657–662.