Theor Appl Genet (1998) 97: 327–336

D. E. Harry · B. Temesgen · D. B. Neale

Codominant PCR-based markers for *Pinus taeda* developed from mapped cDNA clones

Received: 10 February 1998 / Accepted: 25 February 1998

Abstract We report a strategy for developing codominant PCR-based genetic markers by using sequenced cDNA clones from loblolly pine (*Pinus taeda* L.). These clones were previously used as probes for detecting restriction fragment length polymorphisms (RFLPs) to generate linkage maps. After assessing the complexity of banding patterns from Southern blots, we selected clones representing relatively simple gene families, and then determined nucleotide sequences for about 200 bp at each end of the cDNA inserts. Specific PCR primers were designed to amplify samples of genomic DNA derived from two loblolly pine mapping populations. Polymorphisms were detected after digesting the amplified DNA fragments with a battery of restriction endonucleases, and most polymorphisms were inherited in a Mendelian fashion. These newly identified genetic markers are codominant and relatively simple to use. By assaying DNA from individuals used to construct RFLP maps, we show that most of these markers map to the same position as the RFLP loci detected using their corresponding cDNAs as probes, implying that these markers have been converted from RFLP to PCR-based methods. These PCR-based markers will be useful for genome mapping and population genetics.

Key words STS · Codominant PCR marker · RFLP · Loblolly pine (*Pinus taeda* L.) · Comparative map

Communicated by G. Wenzel

D. E. Harry[1] · B. Temesgen · D. B. Neale (✉)
Institute of Forest Genetics, USDA Forest Service,
Department of Environmental Horticulture,
University of California,
Davis, CA 95616 USA
Fax: +1 530-754-9366
E-mail: dbneale@vcdavis.edu

Present address:
[1] Nicholas Turkey Breeding Farms, 19449 Riverside Drive,
PO Box Y, Sonoma, CA 95476, USA

Introduction

We are engaged in research to identify and understand genes controlling important traits in loblolly pine (*Pinus taeda* L.), the most widely planted and commercially important forest species in the United States. Much of our previous work has used restriction fragment length polymorphism (RFLP) markers detected using probes from cloned cDNAs (Devey et al. 1991, 1994; Groover et al. 1994). These RFLP markers are reliable and informative, and they facilitate comparative studies across pedigrees and species (Ahuja et al. 1994). However, procedures for RFLP analysis are technically difficult, thereby restricting their use to relatively few laboratories.

Genetic markers based on DNA amplification are relatively simple and easy to use, and various types of polymerase chain reaction (PCR)-based markers have been developed. These PCR-based systems differ in the nature of the primers used to target amplified DNA, as well as in the nature of the amplified fragments. For example, random amplified polymorphic DNA (RAPD) markers are typically obtained by using short (10–14 nucleotide) primers (Williams et al. 1993). Polymorphisms are usually revealed by the presence or absence of DNA fragments among the amplification products from several individuals. RAPDs have the obvious advantage of requiring no a priori knowledge of DNA sequences, but they are typically inherited in a dominant manner that precludes unambiguous genotypic classifications. Moreover, RAPDs suffer from a lack of reproducibility because they are sensitive to subtle differences in sample storage or processing and in reaction conditions (Neale and Harry 1994).

Amplified fragment length polymorphism (AFLP; Vos et al. 1995) offers an alternative PCR-based strategy for producing large numbers of markers without prior knowledge of DNA sequences. AFLP is proving to be an effective technique for generating DNA markers in many plant and animal species (Mackill et al. 1996; Osten et al. 1996); however, it is not yet understood how broadly applicable such AFLP markers will be among unrelated individuals or in different populations. Although many RAPD and AFLP markers can be developed with relatively modest initial costs, neither marker system is well suited for detecting multiple alleles that are common in many outbreeding species.

Codominant markers are advantageous in that different alleles can be distinguished from one another. Simple sequence repeats (SSR, also called microsatellites or short tandem repeats, STR) are codominant markers that reveal high levels of genetic variation in many eukaryotes. SSRs consist of widely dispersed clusters of tandemly repeated motifs of two to four nucleotides (e.g., $[CA]_n$ or $[AAT]_n$). Different alleles contain different numbers of repeated motifs that can be detected as length variants among amplified DNA fragments generated using primers directed to unique sequences flanking the repeats (Beckman and Soller 1990). The utility and overall heterozygosity of SSR markers is now being explored for a variety of plant species (Akkaya et al. 1992; Morgante and Oliveri 1992; Senior and Heun 1993; Szewc-McFadden et al. 1996). In conifers, SSR markers have been developed for *Pinus radiata* (Smith and Devey 1994), *P. strobus* (Echt and May-Marquardt 1997), and for *Picea sitchensis* (Van De Ven and McNicol 1996). In addition to SSRs, codominant PCR markers can also be developed using DNA sequences from known genes (Konieczny and Ausubel 1993; Palumbi and Baker 1994), from anonymous genomic clones (Bradshaw et al. 1994), or from the ends of selected RAPD fragments (Paran and Michelmore 1993). Once DNA fragments have been amplified, then allelic polymorphism can be detected using any of several methods (Lessa and Applebaum 1993).

This paper describes codominant PCR-based markers developed from sequences of mapped cDNA clones from *P. taeda*. A primary goal was to determine whether sufficient polymorphisms could be found to enable using the amplified fragments as genetic markers. We reasoned that genetic markers based upon conserved coding sequences might also be useful in other pines and conifers, and also serve as anchors for comparative mapping in conifers.

Materials and methods

cDNA clones, sequence analysis, and primer design

Twenty-one cDNA clones were selected from among those previously used as RFLP probes to generate linkage maps (Devey et al. 1994; Groover et al. 1994). The cDNA library had been constructed using random priming, and inserts were non-directionally cloned into a phagemid vector (Devey et al. 1991). Most of the clones were selected because they revealed well-resolved and relatively simple RFLP banding patterns on Southern blots of loblolly pine genomic DNA; however, a few clones revealed more complex banding patterns (Kinlaw and Harry, unpublished data; Kinlaw and Gerttula 1993). About 200 bp of nucleotide sequence was determined at each end of the cloned cDNAs using manual methods (Sequenase kit, US Biochemicals, with [^{35}S]-dATP) and standard sequencing primers (e.g., T3, T7). Nucleotide sequences were compared against those contained in GenBank using BLASTN and FASTA. If genomic sequences from similar plant genes were identified, intron locations and sizes were considered in selecting PCR primers.

In addition to anonymous cDNA clones, a previously sequenced genomic clone of alcohol dehydrogenase (ADH; Harry et al., 1989, and unpublished data) was also included. In this case, primers were selected for conserved sequences in the fifth and tenth exons.

Potential PCR primers were evaluated by using the computer program PRIMER version 0.5, Lincoln et al., unpublished, Whitehead Institute, Massachusetts Institute of Technology (available at www.genome.wi.mit.edu). Primers were selected to have similar properties to facilitate developing standardized conditions for PCR reactions. Primers were 18–21 nucleotides long, with a GC content of 50–60%, and melting temperatures between 59°C and 61°C. Potential primer pairs were also evaluated to minimize self-complementarity (Roux 1995).

Plant materials

Individuals were drawn from two distinct mapping populations (Devey et al.1994; Groover et al.1994). Each population included four unrelated grandparents, two parents, and a large number of full-sib progeny. Most of our analyses were based on 16–25 randomly chosen progeny, but some analyses included as many as 96 progeny. We also analyzed megagametophytes (haploid), which were dissected from wind-pollinated seeds from parents in the mapping populations.

Genomic DNA was isolated from loblolly pine needles (Devey et al. 1991), modified to include an additional organic extraction (phenol : chloroform : isoamyl alcohol, 25 : 24 : 1) followed by an ethanol precipitation. Purified DNA was dissolved in 0.1 × TE (1 m*M* TRIS-HCl, 0.1 m*M* EDTA) at a concentration of 5 ng/μl and stored at 4°C.

DNA was obtained from megagametophytes dissected from germinated seeds. Seeds were surface-sterilized (1% sodium hypochlorite), stored at 4°C for 2–4 weeks, and then placed on moistened filter paper in petri dishes. Petri dishes were maintained at 25°C for 5–14 days, until the radicle had elongated 1–2 cm. Seeds were then dissected to remove seed coats and to separate megagametophytes from embryos. Megagametophytes were frozen on dry ice and stored at −70°C in individual microfuge tubes. Frozen tissues were ground to a fine powder, and DNA was isolated as described by Dellaporta et al. (1985).

PCR amplification and product analysis

PCR reaction mixes included 10 m*M* TRIS-HCl pH 8.3, 50 m*M* KCl, 1.5 m*M* $MgCl_2$, 200 μ*M* of each dNTP, 1% DMSO, and 1 μ*M* of each primer. Typical reaction volumes were 25 μl, and for these we used 15 ng of DNA template and 0.65 units of *Taq* DNA Polymerase (AmpliTaq®; Perkin-Elmer, Foster City, Calif.). All reactions were performed in a Perkin-Elmer 480 DNA Thermal Cycler (Norwalk, Conn.) using 0.5 ml thin-walled reaction tubes with an overlay of about 50 μl mineral oil. We used a combination of "hot-start" (Chou et al. 1992) and "touchdown" (Don et al. 1991) temperature profiles.

DNA amplification was carried out by denaturing at 94°C for 1 min, annealing for 1 min, and extending at 72°C for 2 min. In initial cycles, the annealing temperature was progressively lowered from 65°C to 58°C by 1°C every second cycle. Samples were subjected to an additional 30 cycles of amplification after reaching the final annealing temperature of 58°C. After completing the last cycle, a final extension at 72°C was done for 5 min.

Detection and inheritance of DNA polymorphisms

Amplified DNA fragments were screened for polymorphisms revealed by presence or absence of bands, differences in length, and for sequence variation as revealed by changes in restriction sites. In preliminary experiments using PCR products amplified from unrelated genomic clones, restriction enzymes were screened for their ability to effect complete digests of DNA in a 1:1 (v/v) mix of restriction buffer and PCR buffer. Twelve enzymes were selected: *Ava*II, *Bfa*I, *Bas*JI, *Bst*NI, *Dra*I, *Hae*III, *Hin*fI, *Hin*P1I, *Mse*I, *Msp*I, *Rsa*I, and *Scr*FI. Restriction digests consisted of amplified DNA in 5 µl of PCR reaction mix, 1 µl of 10 × restriction buffer, 0.5 µl (2–10 units) enzyme, and 3.5 µl of water. Restriction buffers were used as supplied with the enzymes, except they were supplemented with dithiothreitol (1 m*M*) and bovine serum albumin (100 µg/ml). Samples were digested for 2–3 h at the supplier's recommended temperature.

DNA samples were subjected to electrophoresis using 2% agarose gels, run for 2–3 h at 5 V/cm in 1 × TBE buffer (TRIS-borate-EDTA; Maniatis et al. 1982). Ethidium bromide (0.5 µg/ml) was included in both the gel and the buffer.

Results

Primer design and DNA amplification

We determined putative identities for 10 of the 21 cDNA clones (Table 1) by comparing their sequences against those contained in GenBank (as of August 1995). For clones pPtIFG1635 and pPtIFG2025, a similar genomic sequence had also been reported. These clones contain coding sequences for the small subunit of ribulose bisphosphate carboxylase (rbcS), for which a cDNA sequence had been characterized from *Pinus thunbergiana* (i.e., *P. thunbergii*; Yamamoto et al. 1988) and a genomic clone had been characterized from *Larix laricina* (Hutchison et al. 1990). Coincidentally, clones pPtIFG2025 and pPtIFG1635 had previously been thought to share similar sequences based on results from Southern blots (A. Groover and M. Sewell, personal communication).

After aligning the sequences from *Pinus taeda*, *P. thunbergiana*, and *Larix laricina*, we selected PCR primers corresponding to different portions of the gene (Fig. 1A). A generic forward primer was chosen for a conserved region near the 5′ end of the *rbc*S gene (Fig. 1B). Similarly, a generic reverse primer was chosen in a conserved region near the 3′ end of pPtIFG2025's coding sequence (Fig. 1B). In addition, two reverse primers were selected to correspond to the 3′ untranslated regions of pPtIFG1635 and pPtIFG2025 (Fig. 1B). We reasoned that such 3′ specific primers might preferentially amplify a subset of the *rbc*S gene family. The generic primer set derived from the coding sequences of pPtIFG2025 was designated stsPtIFG2025c, whereas the primer sets specific for the individual cDNAs were designated stsPtIFG2025-3′ and stsPtIFG1635-3′ (Table 1).

Primers were also designed for each of 19 additional cDNAs (Table 1). All of the primers were initially screened using their corresponding cDNAs as templates. Only 1 primer set (stsPtIFG2707) failed to amplify its cDNA counterpart, and it was dropped from all subsequent analyses. PCR products, which ranged in size from 170 to 950 bp, were readily obtained from the remaining clones (Table 1).

Fifteen pairs of PCR primers, representing 13 cDNA clones and 1 genomic clone, reliably amplified genomic DNA templates isolated from loblolly pine (Table 1). Despite trying different DNA samples, and also making minor adjustments in buffer components (e.g., Mg^{++}), 7 primer sets (Table 1) failed to reliably amplify DNA from genomic templates.

In most instances, PCR products from genomic DNA templates were visualized as a single fragment on agarose gels stained with ethidium bromide. Fragments amplified from genomic DNAs ranged in size from 340 to 1200 bp (Table 1). Small-molecular-weight fragments, if present at all, were relatively minor components. For 8 of 13 cDNA-based primer sets, PCR fragments from genomic templates were measurably larger (100–330 bp; Table 1) than fragments amplified from corresponding cDNAs. This additional DNA indicates the presence of one or more introns in the genomic fragments. No introns were indicated in PCR fragments from the remaining 5 primer sets.

Genetic polymorphisms

Most polymorphisms appeared to involve base substitutions resulting in the gain or loss of a restriction site rather than length variation in the amplified DNA fragments. For example, PCR products from different individuals generated using primer sets stsPtIFG624 and stsPtIFG2253 could not be distinguished until they were digested with restriction enzymes (Fig. 2). After being digested with *Hin*P1I, the single 940-bp PCR fragment from stsPtIFG624 (Fig. 2A) is cleaved into several fragments of 490, 450, and 370 bp (Fig. 2B). A fourth fragment of 120 bp is not shown here (Fig. 2B) but is readily apparent in other gels (e.g., Fig. 3). Phenotypes of individual alleles are clearly distinguished using DNA amplified from megagametophytes and then digested with *Hin*P1I. Megagametophytes from the seed parent of the Base Map population (P_1, Fig. 3, and samples 1–9) show only one phenotype, whereas megagametophytes from the pollen parent (P_2, Fig. 3, and samples 10–18) reveal two phenotypes. One phenotype is identical to the megagametophytes from

Table 1 PCR primers derived from DNA clones of loblolly pine[a]

Clone name	Primer sequence (forward and reverse, 5′ → 3′)	PCR fragment length (bp)		
		cDNA	Genomic	Best identity[b]
pPtIFG602	GGA GAG AAT ATC TGT ACG CCG GCT TTG CAT TGA AGA AAG CA	430	–	–
pPtIFG616	CAG CAT TAC AGA CAT CCA GCA CGC TGT CTT CTT TCT CAC CC	170	–	–
pPtIFG624	CAC AAT TGC CAG ATG GGT C CTT CTC TAG CAA CGA TCC GG	620	940	Protein kinase
pPtIFG653	TTA CAA ATG ACG ACC AGA ATC G TAT GAG CCT CGA CCA CAG C	430	–	–
pPtIFG893	GGA CTG AAG GGA TCT AGC TGG CAG CCC AAA TTC CAT CGT C	450	620 450 420[f]	Nonspecific lipid transfer protein
pPtIFG1454	ACA TCA ATC AAG TTG GCC TTG ACG ACC ATC TCC AAC CAC TC	350	350	–
pPtIFG1457	AGA TGG CGT CCC TTC ATT C ACG GCA AGA GGT GCT TTG	520	–	–
pPtIFG1584	CGA AGC AAA GGA TGT CAC G TGT TGA GGT GGG GAT TGG	340	340	Deoxychalcone synthase
pPtIFG1588	GCT TGT GCT AGT TGC TGC TG GAC GAC CTC TGC TCA AAA CC	700	–	–
pPtIFG1623	CAC CTT TGG GAT AAT TGA CCA TGG CTG TTA GTA CCC TAC CTC C	370	–	–
pPtIFG1635-3′[c]	AAG AAG ACA ACG AGC AAC GG GCC CAC TCG AAT CAC AAA A	650	870	RbcS gene, small subunit
pPtIFG1917	ATA TCC GTC GCC TGG TTA AG GAT TCT CAA AGC AGC CCA AG	320	450	–
pPtIFG1934	GAC GAA GTT GGT GGC GTA G TTC TGT TTG TGC GCC TAC TG	850	850	Light-harvesting complex chlorophyll a/b binding protein
pPtIFG1950	AAA CCA GCA GCC ACA TGA G TAT TAA GAA GGC GGC GGT AC	350	450	–
pPtIFG2009	CAC AGT TCC CCA CAG CAA C ACA AGC GGT TCA GTG GCT C	400	600	Cyclosporin A-regulated Csa-19 gene product
pPtIFG2022	GCC CAC TCT AAC TGA AGC TCC TGA GCG ACG TGA CAG AGA AG	950	–	Glutamine synthetase
pPtIFG2025c[c]	AAG AAG ACA ACG AGC AAC GG AGG AGA TGC ACT GCA CTT GG	400	620	RbcS gene, small subunit
pPtIFG2025-3′[c]	AAG AAG ACA ACG AGC AAC GG GGG GTT GAA AGA ATG GCC	560	780	RbcS gene, small subunit
pPtIFG2166	CTG CTG TTG AGC TTG TGT ACG TGC CCG TGT AAA GAT GAC AG	400	400	Pyruvate dehydrogenase
pPtIFG2253	CCA ATT TGC ACT TTG CCC CCA AAG CCC AAA TCC ATG	370	370	Fructose-bisphosphate aldolase, chloroplast associated
pPtIFG2707[d]	AGA GCT GAA TCG TCG TCT GG AAA GAC TCG GTG CCC TCA C	–	–	–
pPtIFG2723	GCA CAC CAA AGC AGC ATG AAT GGG ACT TGC ACC TTC AG	550	850	–
pPt3001	TGT GTC ATA CAG ACC TTT AT AGG CCA AAT ATT GCT ACT G	700	1200	Alcohol dehydrogenase

[a] All clones are cDNAs, except for pPt3001, a genomic clone of ADH
[b] Based on comparisons with sequences in GenBank
[c] PtIFG1635 and PtIFG2025 contain sequences corresponding to *rbc*S. The same forward primer is used with each of three reverse primers
[d] Failed to amplify cDNA template
[e] Size was deduced from coding sequences
[f] Observed in the Base Map population only

P_1 (Fig. 3, samples 10, 13, 15, and 18), while in the second phenotype, a 490-bp restriction fragment appears to replace the 370- and 120-bp restriction fragments (Fig. 3, samples 11–12, 14, and 16–17).

Our genetic interpretation is that genomic DNA amplified from allele A_1 contains two *Hin*P1I restriction sites, yielding three fragments of 450, 370, and 120 bp. DNA amplified from allele A_2 contains only

one *Hin*P1I site, yielding two restriction fragments of 490 and 450 bp. Hence, genomic DNA amplified from an A_1A_1 homozygote (e.g., P_1, Fig. 3) is characterized by the presence of a 370-bp *Hin*P1I restriction fragment (and a fainter 120-bp fragment) but lacks the 490-bp fragment. Genomic DNA amplified from an A_1A_2 heterozygote (e.g., P_2, Fig. 3) is characterized by the presence of both the 490-bp and 370-bp *Hin*P1I restriction fragments. Genomic DNA amplified from an A_2A_2 homozygote (not shown) would be identical to that of the A_2 megagametophytes depicted in Fig. 3 (lanes 11–12, 14, and 16–17). This genetic interpretation is supported by 1:1 Mendelian segregation ratios that were observed using both haploid and diploid samples (Figs. 3, 4 and Table 2).

Polymorphisms detected using stsPtIFG2253 are also amenable to a simple genetic interpretation. In addition to the 370-bp undigested PCR fragment (Fig. 2C), a second 350-bp fragment is observed after digestion with *Dra*I (Fig. 2D). We surmise that DNA amplified from allele A_1 contains no *Dra*I sites, while DNA amplified from allele A_2 contains a *Dra*I site about 20 bp from one end of the 370-bp fragment. Hence, for the Base Map population, the seed parent and both maternal grandparents are heterozygous A_1A_2 (Fig. 2D, lanes 1–3), whereas the pollen parent and the paternal grandparents are homozygous A_1A_1 (Fig. 2D, lanes 4–6). For the QTL population, the maternal grandfather, paternal grandfather, and pollen parent are all heterozygous A_1A_2 (2-banded phenotypes, Fig. 2D, lanes 8, 10, and 12), whereas the seed parent and the maternal grandparents are homozygous A_1A_1 (1-banded phenotypes, Fig. 2D, lanes 7, 9, and 11). Analyses based on megagametophytes and diploid progeny confirm this interpretation (not shown), which is corroborated by segregation ratios (Table 2).

In addition to polymorphisms revealed after digesting with restriction enzymes, other polymorphisms were detected as the presence or absence (+/−) of a PCR product. When stsPtIFG2723 was used, no PCR products were observed from the seed parent of the Base Map population nor from the pollen parent of the QTL population. In both mapping populations, PCR products were observed from only half of the diploid progeny. Among megagametophytes, about half of those sampled from the + parents gave PCR products. None of the megagametophytes from the − parents gave a PCR product. We infer that stsPtIFG2723 reveals two alleles: the A_1 allele gives a PCR product, whereas the A_n allele does not. In both populations, one parent is heterozygous A_1A_n, whereas the other parent is a homozygous null, A_nA_n (Table 2). This interpretation is consistent with the grandparent phenotypes in both populations, and we presume that homozygous A_1A_1 and heterozygous A_1A_n individuals are phenotypically indistinguishable. Digestion of these products with restriction enzymes did not reveal additional polymorphisms.

A second +/− polymorphism was also detected using stsPtIFG893. A 420-bp fragment (Table 1) was observed only in the Base Map population and it segregated, as expected, for an allele inherited as a Mendelian dominant (Table 2). Interestingly, this 420-bp fragment is smaller than its cDNA counterpart.

Segregation ratios for most markers closely conformed to Mendelian expectations (Table 2). One exception was stsPtIFG2025-3′, for which there is an apparent excess ($\chi_2 = 3.90$, $P \leq 0.048$) of A_1A_1 genotypes among individuals from the Base Map population. RFLPs detected using pPtIFG2025 for these same individuals behaved in an identical manner, and no significant segregation distortion was observed for this locus among a larger sample (Devey et al. 1994).

Two primer sets revealed polymorphisms for which simple Mendelian interpretations could not be made. Primers selected from an ADH genomic clone (Table 1) amplified a single fragment of 1200 bp, but subsequent restriction digests revealed many fragments (not shown), perhaps reflecting polymorphisms in several genes. Amplified DNA from stsPtIFG2166 revealed a repeatable polymorphism that was observed in both mapping populations, yet we have been unable to provide a satisfactory genetic interpretation even using megagametophytes. Perhaps this polymorphism involves multiple genes, or is confounded with one or more null alleles.

Coincidence of RFLP- and PCR-based polymorphisms

Because the PCR-based markers were derived from cDNA clones that had been mapped using RFLPs, we could determine whether the RFLP- and PCR-based markers co-segregate, as expected for closely linked loci. To illustrate, we show an autoradiograph aligned with an agarose gel (Fig. 4) using DNA from the same individuals in the Base Map population. The autoradiograph (Fig. 4A) was made using a probe from pPtIFG624, and the agarose gel (Fig. 4B) shows DNA amplified using primers stsPtIFG624 and then digested with *Hin*P1I. Heterozygous individuals with two restriction fragments on the Southern blot (e.g., samples 2, 4 and 5) also show the larger (i.e., 490 bp) PCR fragment. In other words, the RFLP allele represented by the higher molecular-weight restriction fragment is coincident with the PCR-based allele characterized by the presence of a 490-bp *Hin*P1I fragment. Complete coincidence was observed among 74 progeny in the Base Map population (Table 2). RFLP data were not available from the QTL population.

Likewise for stsPtIFG2253, complete coincidence among 125 individuals (96 from the Base Map, 29 from the QTL population, Table 2) was observed between the RFLP- and PCR-based markers. Among the remaining markers (Table 2), there was generally good coincidence between the PCR markers and their RFLP

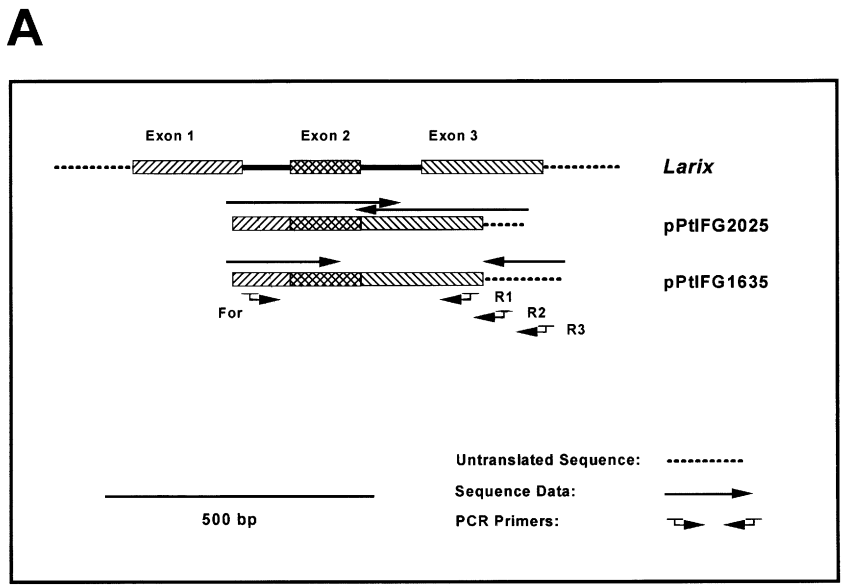

Fig. 1A, B Alignment of *rbc*S sequences from *Pinus taeda*, *P. thunbergiana*, and *Larix laricina*. **A** Loblolly pine cDNAs pPtIFG1635 and pPtIFG2025 relative to the *rbc*S gene from *L. laricina* (Hutchison et al. 1990) and relative to locations of the forward and three reverse (R1, R2, and R3) PCR primers. **B** Nucleotide sequences from selected regions of *rbc*S sequences (shown 5′ to 3′ for the sense-strand), aligned with corresponding sequences from *P. taeda*, *P. thunbergiana*, and *L. laricina*. Primer locations are indicated by |------|. *Uppercase letters* depict conserved bases, whereas *lowercase letters* depict bases that differ among sequences

B

Upstream, translated

```
                          |------------------------|
PtIFG2025:        ...GgG-aAg-AAG-ACA-ACG-AGC-AAC-GGt-gg...
PtIFG1635:        ...GaG-cAg-AAG-ACA-ACG-AGC-AAC-GGt-gg...
P. thunbergiana:  ...GaG-cAg-AAG-ACA-ACG-AGC-AAC-GGt-gg...
Larix laricina:   ...GaG-cAa-AAG-ACA-ACG-AGC-AAC-GGc-tc...
```

Downstream, translated

```
                                        |------------------------|
PtIFG2025:        ...................CCG-CCA-AGT-GCA-GTG-CAT-CTC-CTT-CA...
P. thunbergiana:  ...................CCG-CCA-AGT-GCA-GTG-CAT-CTC-CTT-CA...
Larix laricina:   ...................CCG-CCA-AGT-GCA-GTG-CAT-CTC-CTT-CA...
```

Downstream, untranslated

```
                                                   |----------------------|
PtIFG2025: .....................................GCC-GGC-CAT-TCT-TTC-AAC-CCC-GTT...
```

Downstream, untranslated

```
                                                 |-------------------------|
PtIFG1635:        ................................TtG-TTT-TGT-GAT-TCG-AGT-GGG-CAAT...
P. thunbergiana:  ................................TcG-TTT-TGT-GAT-TCT-AGT-GGG-CAAT...
```

counterparts. For stsPtIFG2009, we found 1 discrepancy among 35 individuals (Table 2), but whether this is a bona fide recombinant or a misclassified genotype is not clear.

However, we were not able to establish coincidence between all PCR-based and RFLP-based markers (Table 2). Segregation for stsPtIFG1950 does not coincide with any of the mapped RFLP loci. We were also unable to establish a correspondence between an RFLP marker and either of the +/− polymorphisms revealed by stsPtIFG893 and stsPtIFG2723.

Discussion

In this report we show that nucleotide sequences from mapped cDNA clones can be used to design PCR primers for amplifying samples of genomic DNA and reveal polymorphisms inherited as codominant alleles. More than 60% of the tested primer pairs successfully directed the amplification of genomic DNA, and more than 60% of these revealed a Mendelian polymorphism in at least one of two full-sib families. In most instances,

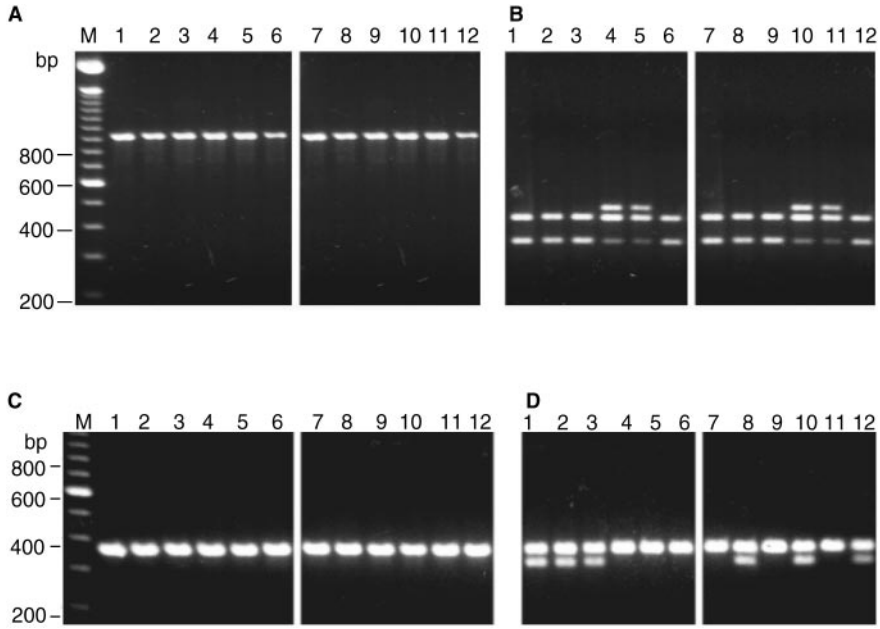

Fig. 2A–D PCR products amplified from loblolly pine genomic DNA using primers derived from partially sequenced cDNA clones. *Lanes 1–6* are from the Base Map population (Devey et al. 1994), while *lanes 7–12* are from the QTL population (Groover et al. 1994), arranged as follows: *lanes 1, 7* maternal grandmother, *2, 8* maternal grandfather, *3, 9* seed parent, *4, 10* pollen parent, *5, 11* paternal grandmother, *6, 12* paternal grandfather. Lane *M* is a 100-bp DNA ladder. Panels **A** and **C** show undigested PCR products from stsPtIFG624 and stsPtIFG2253, respectively. These same PCR products are also shown after being digested with restriction enzymes: Panel **B** stsPtIFG624 with *Hin*P1I; Panel **D** stsPtIFG2253 with *Dra*I

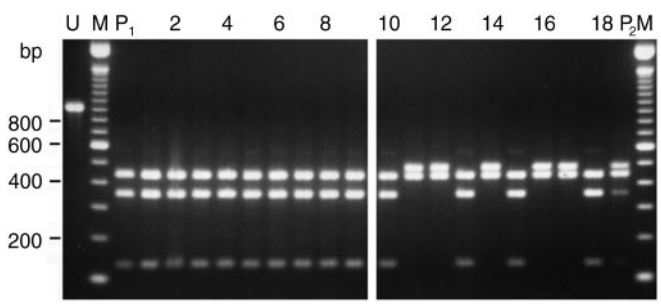

Fig. 3 Mendelian segregation of PCR products amplified using DNA from loblolly pine megagametophytes. Primers from stsPtIFG624 were used with haploid and diploid DNA samples from the Base Map population. *Samples 1–9* are megagametophytes from the seed parent (P_1), and *samples 10–18* are megagametophytes from the pollen parent (P_2). *Sample U* is an undigested amplification product of P_1, whereas all the other PCR products were digested with *Hin*P1I. *Sample M* is a 100-bp DNA ladder

these newly derived PCR-based markers are closely linked to the RFLP markers detected by the cDNA probes from which the PCR primers were selected. Because the cDNA clones used in this study have been shown to hybridize to DNA from a variety of other pines and conifers (Ahuja et al. 1994), we anticipate that at least some of the primers described here will also be useful for amplifying DNA from related species (Harry et al., in preparation).

There are several reasons why selecting PCR primers from cDNAs might fail to reveal Mendelian polymorphisms a priori. First, the location or size of introns could prevent successful amplification of genomic templates, particularly if the locations of primers with respect to introns cannot be assessed. This is particularly troublesome for primers derived from unidentified cDNAs or from cDNAs for which no corresponding genomic sequence is available. Another potential problem is that PCR fragments from expressed genes may not contain sufficient sequence variation to allow for the detection of polymorphisms. Finally, the genomes of pines and perhaps other conifers include many gene families that might be co-amplified using PCR primers derived from a cDNA sequence. Given such reasons for pessimism, our overall rate of success is encouraging. What steps did we take to help ensure success?

First, we gleaned substantial information by examining RFLP banding patterns from Southern blots probed with the cDNA clones. In selecting cDNAs to include in our study, we specifically emphasized clones representing smaller gene families. We anticipated difficulties in establishing Mendelian interpretations for those PCR products arising from different genes. ADH genes in pine, for example, belong to a complex gene family (Kinlaw et al. 1990; Perry and Furnier 1996). RFLP banding patterns using ADH probes are complex, and only two of many ADH genes have been recently mapped (Groover et al. 1994). In the present study, PCR products generated using the primers from an ADH genomic clone (Table 1) appeared as a single 1200-bp band, but restriction digests revealed that multiple classes of fragments had been amplified.

Table 2 Genetic polymorphisms and Mendelian segregation in two mapping populations of loblolly pine detected using PCR fragments amplified with primers derived from cDNA clones. Also shown are restriction enzymes used to reveal the polymorphisms

Locus	Base Map population				QTL population			
	Enzyme	Parental genotypes	Segregation ratios[a]	Linked RFLP locus[b]	Enzyme	Parental genotypes	Segregation ratios[a]	Linked RFLP locus[b]
stsPtIFG624	*Hin*P1I	$A_1A_1 \times A_1A_2$	39:35	PtIFG624.a	*Hin*P1I	$A_2A_2 \times A_1A_2$	37:39	No RFLP data
stsPtIFG893[c]	Uncut	$A_1A_n \times A_1A_1$	12:10	Not linked	No polymorphism detected			
stsPtIFG1635-3′	No polymorphism detected				*Hin*P1I	$A_2A_2 \times A_1A_2$	27:29	PtIFG1635.a
stsPtIFG1934	*Bsa*JI	$A_1A_1 \times A_1A_2$	38:38	PtIFG1934.a	*Bsa*JI	$A_1A_2 \times A_2A_2$	6:4	PtIFG1934.a
stsPtIFG1950	*Hin*fI	$A_1A_1 \times A_1A_2$	45:48	Not linked	*Hin*fI	$A_1A_2 \times A_1A_2$	19:44:29[d]	Not linked
stsPtIFG2009	No polymorphism detected				*Bst*NI	$A_2A_2 \times A_1A_2$	17:18	PtIFG2009.a
stsPtIFG2025-3′	*Bst*NI	$A_1A_1 \times A_1A_2$	21:10	PtIFG2025.a	*Mse*I	$A_1A_1 \times A_1A_2$	25:30	Not linked
stsPtIFG2253	*Dra*I	$A_1A_1 \times A_1A_2$	48:48	PtIFG2253.a	*Hin*fI	$A_1A_1 \times A_1A_2$	19:12	PtIFG2253.a
stsPtIFG2723[c]	Uncut	$A_nA_n \times A_1A_n$	20:16	Not linked	Uncut	$A_1A_n \times A_nA_n$	9:10	Not linked

[a] Includes diploid progeny, and as available, megagametophytes from a heterozygous parent
[b] RFLP data for diploid progeny (Devey et al. 1994; Groover et al. 1994)
[c] Polymorphisms are characterized by the presence or absence of PCR products
[d] $A_1A_1 : A_1A_2 : A_2A_2$

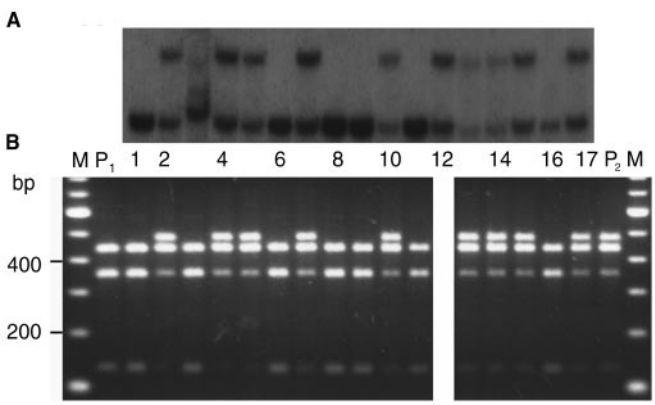

Fig. 4A, B Mendelian segregation for RFLP- and PCR-based markers among progeny in the Base Map population. **A** Southern blot of DNA samples digested with *Hin*dIII and probed with PtIFG624. **B** PCR products from the same individuals (sample 12 was not available) amplified using primers from stsPtIFG624 and digested with *Hin*P1I. Also shown are the seed parent (P_1), pollen parent (P_2), and 100-bp DNA ladder (M)

Sequence comparisons identified similarities with previously characterized sequences which in some instances helped to target primers for specific regions in known genes (Fig. 1). The estimated length of intron sequences (220 bp, Table 1) contained in the amplified products from stsPtIFG1635-3′, stsPtIFG2025c, and stsPtIFG2025-3′ closely corresponds to the combined length of the two introns in *Larix* (Hutchison et al. 1990). Furthermore, amplification products generated using the 3′-specific primers are a subset of those amplified using the coding sequence primers. Polymorphisms observed using stsPtIFG2025-3′ were not observed using stsPtIFG1635-3′ (and vice versa), whereas both classes of polymorphisms could be detected using stsPtIFG2025c. However, the relevant restriction fragments were less abundant using stsPtIFG2025c, hence the polymorphisms were more difficult to detect.

For cDNAs in which sequence comparisons failed to provide additional information for targeting PCR primers, we proceeded empirically with reasonable success. Of the 21 sets of PCR primers that amplified their cDNA templates, two-thirds also amplified genomic templates. Large intron sequences may account for some failures. Other failures could be caused by selecting a primer from a cDNA sequence that is coincident with the junction of two exons in a gene. The relative degree to which these factors contributed to failure remains unknown. Because DNA sequences will become more plentiful in the future, sequence comparisons will become increasingly useful in selecting PCR primers from cDNA sequences.

Careful consideration of PCR reaction conditions also proved worthwhile. First, primer specificity was ensured by using high-stringency conditions such as 1.5 mM magnesium and annealing temperatures of 2–3°C below the calculated T_m (Roux 1995). In addition, we found both hot-start (Chou et al. 1992) and touchdown procedures (Don et al. 1991) helpful to reliably obtain large amounts of amplified DNA that was largely free of smaller molecular-weight products. In contrast to Tsumura et al. (1997), we found little advantage in modifying PCR reaction conditions for specific pairs of primers. We decided instead to use the same conditions for all primer sets.

Most polymorphisms described in this study appear to involve base substitutions, or perhaps insertions or deletions of only a few bases. The extent of sequence variation among conifer alleles is not known, but it must be relatively high given that these polymorphisms

were detected using only 12 restriction enzymes. We were unable to detect more than two alleles segregating within either mapping population, even by digesting with multiple enzymes. Because the corresponding RFLP markers detected three or four alleles at several loci, our results suggest that additional polymorphisms might yet be detected. Additional polymorphisms could be revealed by surveying additional restriction enzymes or by using a more sensitive detection method (Lessa and Applebaum 1993).

Polymorphisms characterized by the presence or absence of an amplified product tend to be more problematic than those detected using restriction enzymes. Null alleles are sometimes difficult to score with confidence because the relative abundance of amplified product can vary among samples and failed reactions can be scored as nulls. Null alleles are also problematic because they tend to be inherited in a recessive manner, so genotypic classifications are ambiguous. Analysis of megagametophytes from putative heterozygotes can be helpful in verifying null alleles, as demonstrated by stsPtIFG2723 (Table 2). A likely cause of null alleles is sequence variation coincident with a primer's annealing site. Redesigning primers to slightly shift their location upstream or downstream could help alleviate this problem. Some PCR products may result from chance amplifications, in which case the amplified fragment might share little sequence similarity with the cDNA from which primers were designed. The 420-bp fragment from stsPtIFG893 (Table 1) may represent such an example. Not only is this fragment smaller than its counterpart amplified from a cDNA template, it also fails to hybridize to a probe from pPtIFG893 (C. Kinlaw, personal communication).

In most instances, polymorphisms detected among the amplified DNA fragments were closely linked with RFLPs detected using a corresponding cDNA probe (Fig. 4, Table 2). These results illustrate the ability to convert RFLPs to PCR-based markers, even though the polymorphisms per se are not identical. Furthermore, RFLPs may detect alleles not seen in amplified DNA fragments, and vice versa.

Gene families complicate the process of associating loci and alleles among related markers. RFLP probes from pPtIFG2025 and pPtIFG1635 detect *rbc*S genes at distinct loci (Devey et al. 1994; Groover et al. 1994), and there are additional polymorphisms that have not yet been mapped because they do not segregate within the existing mapping populations (M. Sewell, personal communication). In the Base Map population, stsPtIFG2025-3′ shows complete cosegregation with the RFLP locus PtIFG2025.a (Table 2). In the QTL population, stsPtIFG1635-3′ shows complete cosegregation with RFLP locus PtIFG1635.a (Table 2) and yet it is not linked to stsPtIFG2025-3′, nor is stsPtIFG2025-3′ linked to any other mapped gene belonging to the *rbc*S family. It seems that stsPtIFG2025-3′ detects a polymorphism in the QTL population that has not yet been mapped using RFLPs. Complications from gene families may also account for our results using stsPtIFG1950. As judged from Southern blots, pPtIFG1950 detects several genes, but we could not establish linkage between the PCR-based polymorphism and RFLPs. It is less clear why we were unable to show linkage for the PtIFG2723 polymorphisms. stsPtIFG2723 detects a null allele, which could complicate genotypic classifications for some individuals. Alternatively, Southern blots probed using pPtIFG2723 reveal uneven hybridization intensities (M. Sewell, personal communication), suggesting that genotypic classifications could also be problematic for RFLPs. A more extensive effort is now underway to map these (and other) PCR-based markers (Temesgen et al., in preparation).

In this report, we describe an approach for developing codominant STS markers from cDNA clones. Such PCR-based markers provide an alternative to RFLPs for dissecting genetic variation in functional genes of loblolly pine. Our approach, which is the same as that used by Tsumura et al. (1997) for *Cryptomeria*, differs from other strategies based upon microsatellites, cloned genes of known sequence, or even RAPD fragments. Primers selected from cDNA sequences may not be well situated to efficiently amplify genomic DNA. In addition, expressed genes may not harbor sufficient sequence variation for effective use as genetic markers. But despite these potential drawbacks, our results, together with those of Tsumura et al. (1997), suggest that developing STS markers from cDNA sequences is a reasonable strategy. A potential advantage to this strategy is that primers based on coding sequences may amplify orthologous DNA in several species and thereby facilitate comparative mapping (Lyons et al. 1997). Tsumura et al. (1997) report some success in using their cDNA-based STS markers in several conifer species. We have also begun to investigate the use of these markers in other conifer species (Harry et al., in preparation), for use in comparative mapping and to increase our basic understanding of genome evolution in *Pinus*. If pine genomes share a high degree of collinearity, comparative maps that include conserved anchor loci will greatly facilitate the process of identifying and mapping homologous genes or quantitative trait loci across species.

Acknowledgments We are grateful to Claire Kinlaw for providing cDNA sequences, sharing results from database searches, and for general discussions throughout the project. Chris Baysdorfer oversaw sequencing efforts for some cDNA clones. Christine Scobee and Tan Ho provided technical assistance. Craig Echt and Daniel Perry reviewed an earlier draft of the manuscript. We also thank Mitch Sewell, Andrew Groover, and Bradley Sherman for their assistance with RFLP data. We thank Weyerhaeuser Company for providing plant materials. This research was supported by grants from the USDA National Research Initiative Plant Genome Program (93-37300-8785) and the Dendrome Database project.

References

Ahuja MR, Devey ME, Groover AT, Jermstad KD, Neale DB (1994) Mapped DNA probes from loblolly pine can be used for restriction fragment length polymorphism mapping in other conifers. Theor Appl Genet 88:279–282

Akkaya MS, Bhagwat AA, Cregan PB (1992) Length polymorphisms of simple sequence repeat DNA in soybean. Genetics 132:1131–1139

Beckmann JS, Soller M (1990) Toward a unified approach to genetic mapping of eukaryotes based on sequence tagged microsatellite. Bio/Technology 8:930–932

Bradshaw HD Jr, Villar M, Watson BD, Otto KG, Stewart S, Stettler RF (1994) Molecular genetics of growth and development in *Populus*. III. A genetic linkage map of a hybrid poplar composed of RFLP, STS, and RAPD markers. Theor Appl Genet 89:167–178

Chou Q, Russell M, Birch DE, Raymond J, Bloch W (1992) Prevention of pre-PCR mis–priming and primer dimerization improves low-copy-number amplifications. Nucleic Acids Res 20:1717–1723

Dellaporta SL, Wood J, Hicks JB (1985) Maize DNA minipreparation. In: Malmberg R, Messing J, Sussex R (eds) Molecular biology of plants. Cold Spring Harbor Laboratory Press, New York, pp 36–37

Devey ME, Jermstad KD, Tauer CG, Neale DB (1991) Inheritance of RFLP loci in a loblolly pine three-generation pedigree. Theor Appl Genet 83:238–242

Devey ME, Fiddler TA, Liu B-H, Knapp SJ, Neale DB (1994) An RFLP linkage map for loblolly pine based on a three-generation outbred pedigree. Theor Appl Genet 88:273–278

Don RH, Cox PT, Wainwright BJ, Baker K, Mattick JS (1991) Touchdown PCR to circumvent spurious priming during gene amplification. Nucleic Acids Res 19:4008

Echt CS, May-Marquardt P (1997) Survey of microsatellite DNA in pine. Genome 40:9–17

Groover A, Devey ME, Fiddler T, Lee J, Megraw R, Mitchell-Olds T, Sherman B, Vujcis S, Williams C, Neale DB (1994) Identification of quantitative trait loci influencing wood specific gravity in outbred pedigree of loblolly pine. Genetics 138:1293–1300

Harry DE, Mordecai KS, Kinlaw CS, Loopstra CA, Sederoff RR (1989) DNA sequence diversity in alcohol dehydrogenase genes from pines. In: Proc 20th Southern Forest Tree Improvement Conf. Sponsored Publ No 42 of the Southern Forest Tree Improvement Committee, pp 373–380

Hutchison KW, Harvie PD, Singer PB, Brunner AF, Greenwood MS (1990) Nucleotide sequence of the small subunit of ribulose-1,5-bisphosphate carboxylase from the conifer *Larix laricina*. Plant Mol Biol 14:281–284

Kinlaw CS, Harry DE, Sederoff RR (1990) Isolation and characterization of alcohol dehydrogenase cDNAs from *Pinus radiata*. Can J For Res 20:1343–1350

Kinlaw CS, Gerttula SM (1993) Complex pine gene families of pines. In: Proc 22nd Southern Forest Tree Improvement Conf, pp 14–17

Konieczny A, Ausubel FM (1993) A procedure for mapping Arabidopsis mutations using co-dominant ecotype-specific PCR-based markers. Plant J 4:403–410

Lessa EP, Applebaum G (1993) Screening techniques for detecting allelic variation in DNA sequences. Mol Ecol 2:119–129

Lyons LA, Laughlin TF, Copeland NG, Jenkins NA, Womack JE, Obrien SJ (1997) Comparative anchor tagged sequences (CATS) for integrative mapping of mammalian genomes. Nat Genet 15:47–56

Mackill DJ, Zhang Z, Redona ED, Colowit PM (1996) Level of polymorphism and genetic mapping of AFLP markers in rice. Genome 969–977

Maniatis T, Fritsch EF, Sambrook J (1982) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, New York

Morgante M, Olivieri AM (1992) PCR-Amplified microsatellites as markers in plant genetics. Plant J 3:175–182

Neale DB, Harry DE (1994) Genetic mapping in forest trees: RFLPs, RAPDs, and beyond. AgBiotech News Inf 6:107N–114N

Otsen M, Den Bieman M, Kuiper MTR, Pravenec M, Kren V, Kurtz TW, Jacob HJ, Lankhorst A, Van Zutphen BFM (1996) Use of AFLP markers for gene mapping and QTL detection in the rat. Genomics 37:289–294

Palumbi SR, Baker CS (1994) Contrasting population structure from nuclear intron sequences and mtDNA of humpback whales. Mol Biol Evol 11:426–435

Paran I, Michelmore RW (1993) Development of reliable PCR-based markers linked to downy mildew resistance genes in lettuce. Theor Appl Genet 85:985–993

Perry DJ, Furnier GR (1996) *Pinus banksiana* has at least seven expressed alcohol dehydrogenase genes in two linked groups. Proc Natl Acad Sci USA 93:13020–13023

Roux KH (1995) Optimization and troubleshooting in PCR. PCR Methods Appl 4:S185–S194

Senior ML, Heun M (1993) Mapping maize microsatellites and polymerase chain reaction confirmation of the targeted repeats using a CT primer. Genome 36:884–889

Smith DN, Devey ME (1994) Occurrence and inheritance of microsatellites in *Pinus radiata*. Genome 37:977–983

Szewc-McFadden AK, Kresovich S, Bliek SM, Mitchell SE, McFerson JR (1996) Identification of polymorphic, conserved simple sequence repeats (SSRs) in cultivated *Brassica* species. Theor Appl Genet 93:534–538

Tsumura Y, Suyama Y, Yoshimura K, Shirato N, Mukai Y (1997) Sequence-tagged-sites (STSs) of cDNA clones in *Cryptomeria japonica* and their evolution as molecular markers in conifers. Theor Appl Genet 94:764–772

Van De Ven WTG, McNicol RJ (1996) Microsatellites as DNA markers in Sitka spruce. Theor Appl Genet 93:613–617

Vos P, Hogers R, Bleeker M, Reijans M, Van De Lee T, Hornes 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

Williams JGK, Hanafey MK, Rafalski JA, Tingey SV (1993) Genetic analysis using random amplified polymorphic DNA markers. Recombinant DNA 218:704–740

Yamamoto N, Kano-Murakami Y, Matsuoka M, Ohashi Y, Tanaka Y (1988) Nucleotide sequence of a full length complementary DNA clone of ribulose bisphosphate carboxylase small subunit gene from green dark-grown pine *Pinus thunbergii* seedling. Nucleic Acids Res 16:11830