Theor Appl Genet (1998) 96: 969–979 © Springer-Verlag 1998

P. J. Fisher · T. E. Richardson · R. C. Gardner

Characteristics of single- and multi-copy microsatellites from *Pinus radiata*

Received: 10 November 1997 / Accepted: 5 January 1998

Abstract Dinucleotide microsatellites were isolated from *Pinus radiata* using both a standard genomic library and libraries enriched for microsatellites. Locus-specific primers were designed to amplify 43 unique microsatellites. Thirty two of these loci had interpretable PCR patterns, 11 of which were polymorphic in a screen of 19 *P. radiata* individuals; all 11 polymorphic loci contained at least 17 repeats in the sequenced plasmid. Six of the eleven primer pairs amplified multiple fragments per individual (3–8), suggesting that these loci were present in multiple copies in the genome. Genotyping a 48-tree *P. radiata* production population with seven of the most polymorphic microsatellites revealed an average of 17 bands per locus (the multi-copy microsatellites were treated as one locus). When tested on known pedigrees, both single and multi-copy microsatellites exhibited co-dominant inheritance and Mendelian segregation. Two loci had null alleles and one locus had a high frequency of non-parental alleles, suggesting a high mutation rate. Eight of these microsatellites, including five multi-copy loci, were placed on a partially constructed *P. radiata* genetic map. Four of the five multi-copy microsatellites had two or more sets of alleles that mapped to the same locus, and the fifth mapped to two unlinked loci. All seven tested primer pairs amplified PCR products from other species of hard pine, three amplified products from soft-pine species, and one amplified bands in other conifers.

Communicated by P. M. A. Tigerstedt

P. J. Fisher[1] (✉) · T. E. Richardson
Forest Research Institute, Private Bag 3020, Rotorua, New Zealand

R. C. Gardner
School of Biological Sciences, University of Auckland, Private Bag 92019, Auckland, New Zealand

Present address:
[1] Neuropsychiatric Genetics Unit, University of Wales, College of Medicine, Heath Park, Cardiff, CF4 4XN, Wales, UK

Key words Molecular markers · *Pinus radiata* · Microsatellites · Multi-copy · Null alleles · Mutation rate

Introduction

Plantation forests of *Pinus radiata* in New Zealand, Chile, Australia and Spain have a combined estate covering more than 2.4 million hectares. Extensive *P. radiata* improvement programmes have been underway for more than 40 years and these programmes have increased genetic gain by as much as 20% for some traits using classical quantitative methods (Eldridge 1982; Gleed 1982; Johnson et al. 1992). In New Zealand, both open- and control-pollinated breeding strategies are currently combined with vegetative propagation to produce advanced genetic material for planting (Vincent 1997).

Over the past decade, molecular markers have been integrated into many plant breeding programmes. Molecular markers allow plant breeders to distinguish between cultivars (Bowers et al. 1996; Charters et al. 1996), to detect mislabelling of high-value genotypes (Adams et al. 1988), and to determine the parentage of control-pollinated crosses (Neale et al. 1992). Molecular markers have also been used to construct genetic maps as well as to locate quantitative trait loci (QTLs) of importance in genetic-improvement programmes (Paterson et al. 1990; Rafalski and Tingey 1993; Binelli and Bucci 1994; Byrne et al. 1995; Gardiner et al. 1996; Xiao et al. 1996; Cadalen et al. 1997) and to speed the integration of new genetic material into elite germplasm (Dale and Chaparro 1996). The use of marker-trait associations in selection programmes has great potential in forest trees because of the long generation times and the late expression of most commercially important traits (Neale et al. 1992; Byrne et al. 1995; Devey et al. 1996).

Many molecular marker systems have been developed for use in plants (reviewed by Powell et al. 1996). Microsatellites, or simple-sequence repeats (SSRs), are 1–6-bp tandemly repeated DNA motifs which may vary in the number of repeats at a given locus. Microsatellite markers have proven to be particularly valuable in plant-breeding programmes because they are multi-allelic, co-dominantly inherited, widely dispersed across the genome, easily scored, and their analysis can be automated (Morgante and Olivieri 1993; Queller et al. 1993; Powell et al. 1995; Kashi et al. 1997).

Microsatellites have been isolated and characterised from several *Pinus* species, including *P. radiata* (Smith and Devey 1994; Fisher et al. 1996), *P. sylvestris* (Kostia et al. 1995) and *P. strobus* (Echt et al. 1996). Smith and Devey (1994) reported on five dinucleotide microsatellites from *P. radiata*, two of which were polymorphic, and we recently reported the isolation of additional dinucleotide repeats from radiata pine using a novel enrichment procedure (Fisher et al. 1996). Here we report the characterisation of 33 *P. radiata* microsatellite loci and assess their level of polymorphism and utility for DNA profiling and linkage mapping in *P. radiata*. We also investigate the transferability of these loci to other coniferous species.

Materials and methods

Plant material and DNA extraction

To construct the genomic libraries and test for polymorphism of microsatellite loci, needles were collected from 48 *P. radiata* trees grown in New Zealand seed orchards. Haploid megagametophyte and diploid embryonic tissue were dissected from control-pollinated seeds of six of these trees to study the inheritance of microsatellites. To test for PCR amplification of microsatellite loci in other conifers, needles were collected from individuals representing 16 species, including nine additional pine species (see Table 3). Genomic DNA was isolated from all tissues using a modified CTAB procedure (Cato and Richardson 1996), or with a FastPrep FP120 machine (Savant) using a BIO 101 kit following the manufacturer's instructions.

Standard library construction and screening

Genomic DNA from one *P. radiata* individual was digested with *Sau*3A, and the restriction fragments (mostly 100 bp–1 kb) were ligated into *Bam*H1-digested, de-phosphorylated plasmid vector pBS-SK + (Stratagene). The ligation mix was cloned, plated, and the resulting colonies lifted onto nylon NEF-978Y membranes (Biotechnology Systems, NEN Research Products) as previously described (Fisher et al. 1996). A $(CT)_8$ or $(GT)_8$ oligonucleotide was end-labelled with [γ^{32}P] -ATP using T4 polynucleotide kinase (GIBCO-BRL), hybridised to membranes according to the manufacturer's instructions (final wash in 0.1 × SSC at 48°C), and then exposed to X-Omat-AR X-ray film (Kodak). Positive colonies were picked onto new membranes and screened a second time as above, to ensure that they contained microsatellites.

Construction and screening of libraries enriched for microsatellites

Genomic libraries were enriched for microsatellites using the method of Fisher et al. (1996). Briefly, degenerate primers designed to anchor at the 5′ ends of microsatellite sequences were used to PCR *P. radiata* genomic DNA. Three primers [PCT4, KKVRVRV$(CT)_6$; PCT7, IIDVRVRV$(CT)_5$; and PCA7, IIDBRYRY$(CA)_5$] were used individually, and in combination, to amplify genomic DNA at multiple loci under the PCR conditions described by Fisher et al. (1996). The PCR products from PCT4 alone, PCT7 alone, and a mixture of PCT7 and PCA7, were cloned to produce three enriched libraries. These libraries were probed as above, with either a $(GA)_{16}$, or a $(CA)_{16}$, oligonucleotide (final wash in 0.1 × SSC at 65°C).

Primer design, PCR and electrophoresis

Plasmid DNA was extracted from clones that were positive in the second screen (Sambrook et al. 1989) and sequenced using M13 universal primers and an ABI 373 automated sequencer, or manually using a T7 DNA sequencing kit (Pharmacia). Primer pairs were designed to amplify specific loci using PRIMER 0.5 (Whitehead Institute for Biochemical Research, Cambridge, Massachusetts, 1991). PCRs were performed either with two locus-specific primers, or one locus-specific primer and one of the degenerate primers (PCT4, PCT7 or PCA7) as described by Fisher et al. (1996). Radiolabelled PCRs containing [α-^{32}P] -dCTP were carried out in 10 µl volumes, as described by Fisher et al. (1996), with appropriate annealing temperatures according to their predicted Tms. Alternatively, PCRs were radiolabelled by incorporating end-labelled primers in the reaction mixes. PCRs with end-labelled primers were carried out as described for other radiolabelled PCRs (Fisher et al. 1996), except that 20 µM of dCTP replaced 0.03 µM of [α-^{32}P]-dCTP and 2 µM of dCTP.

Multiplexed PCRs of primer pairs NZPR4, NZPR5 and NZPR6 (see Table 1) were carried out similarly to the PCRs containing [α-^{32}P]-dCTP, except that 3 pmol of each primer were added for primer pairs NZPR4 AND NZPR5, and 4 pmol of each primer for pair NZPR6.

PCR products were electrophoresed through 6–10% denaturing polyacrylamide gels with an acetate gradient (Sambrook et al. 1989). Fragment sizes were estimated by comparison with either electrophoresed PCR products of known size, or a manually sequenced M13 DNA ladder. Gels were fixed and dried (Sambrook et al. 1989), and exposed to X-ray film (Kodak X-Omat-AR, or Xmat-K) for 1–24 h.

Polymorphism, segregation and transmission

Forty three primer pairs designed from sequence-flanking microsatellites were tested for scoreable amplification of fragments of the expected size from *P. radiata* genomic DNA. Nineteen trees, including the parents of a full-sibling mapping population (see below), were used to determine which microsatellites were polymorphic. Mendelian segregation of putatively polymorphic loci was tested by genotyping parents, progeny, and megagametophytes, from two-generation pedigrees. Gene diversity values for each locus were calculated using data from 42 unrelated trees (Weir 1990).

Testing for null alleles

One locus (NZPR6) was found to have a null allele. To test for non-amplifying "null" alleles at this locus, embryos or megagametophytes from each tree that had a single NZPR6 amplification

product were amplified with two sets of multiplexed primer pairs. One primer pair was specific to NZPR6 (the tested locus) and the second primer pair was specific to a control locus known not to contain a null allele. If no amplification at the tested locus was observed from the DNA of one or more of the progeny, or megagametophytes, but amplification of the expected size product was observed at the control locus, the tree was assumed to have a segregating null allele at the tested locus. If fragments of the same size were observed at the tested locus in all of the progeny, the parental tree was presumed to be homozygous. For the calculation of gene diversity, trees with one null allele were scored as heterozygotes.

Linkage analysis

Linkage relationships among eight polymorphic microsatellites segregating in an existing full-sibling mapping population ($n = 88$ trees) were examined to determine the distribution of these loci in the *P. radiata* genome. Initially, segregation of microsatellite bands were compared in all pair-wise combinations to estimate recombination frequencies (θ) in order to test for departures from the expected ($\theta = 0.5$) for unlinked bands. A pseudo-testcross strategy (Grattapaglia and Sederoff 1994) with MapMaker for Macintosh version 2.0 (Lander et al. 1987) was also used to determine linkage relationships. The criteria for accepting linkage among loci were LOD = 3, $\theta = 0.4$.

Results

Isolation of dinucleotide microsatellites in *P. radiata*

An un-enriched genomic DNA library (average insert size of 550 bp) containing approximately 35 000 clones was screened at low stringency with $(GA)_8$ and $(CA)_8$ oligonucleotides, resulting in 26 GA-positive, and five CA-positive clones. Sequencing of all 31 positive clones yielded 12 unique $(CT)_n$ and four unique $(CA)_n$ repeats, where $n > 8$. Four almost identical sequences were found among the 26 $(CT)_n$-positive clones; one contained a $(CT)_{10}$ repeat, the other three a $(CT)_9$. The longest number of repeats observed in the 16 unique microsatellites was 18.

A total of about 16 000 clones were plated from the three microsatellite-enriched genomic libraries. Approximately 250 GA-positive signals were found when (4000) clones were screened at high stringency with a $(GA)_{16}$ oligonucleotide, and 150 positives were found when 2000 clones were screened with $(CA)_{16}$. Sequencing of 42 clones revealed that every clone contained at least one microsatellite with $n \geq 12$. A total of 33 unique loci were identified in the 42 sequenced clones. From these 33 clones, 47 microsatellites with $n \geq 12$ were identified (some clones contained more than one microsatellite). In the most extreme example, one clone contained seven regions of tandemly repeated DNA and each region consisted of non-microsatellite DNA flanking a microsatellite (locus 7 in Fisher et al. 1996). This clone accounted for 6 of the 47 microsatellites [the seventh microsatellite contained less than 12 (CT) repeats]. Another clone contained three tandem repeats, each with one microsatellite embedded within it. The number of repeats for these loci ranged from 12 to 63 (mean = 25.5).

Polymorphism of the microsatellite loci

Locus-specific PCR primer pairs were designed for 43 of the microsatellite loci described above. The primers were tested to determine if they amplified PCR products in the expected size range and if they were polymorphic across 19 *P. radiata* trees. Primers that amplified scoreable fragments were then tested on progeny and megagametophytes (typically 4–12 of each) to determine if the amplified bands showed Mendelian inheritance (see below and Table 1).

Two features of the results merit comment. Firstly, the overall frequency of polymorphic banding patterns was low. Only 11 of the 32 microsatellites revealed size differences among the fragments amplified from the 19 individual trees (Table 1) with the number of repeats in the sequenced plasmid being a reasonable predictor of polymorphism. The shortest polymorphic microsatellite contained 17 repeats in the sequenced clone. Of the sequenced plasmids with 20 or more repeats, eight out of ten primer pairs revealed polymorphism. In contrast, only 3 of the 23 loci with 9–19 repeats in the sequenced plasmid gave rise to polymorphic banding patterns.

A second feature of the results is that 8 of the 32 primers gave rise to multiple fragments (Table 1). This included situations where there was more than one fragment per individual and each fragment was in every other individual in the population (presumed multi-copy monomorphic, seen in two cases), and when there were more than two bands in any one individual profile (multi-copy, polymorphic, seen in six cases). The latter case is illustrated in Fig. 1, which shows the PCR products from 16 *P. radiata* individuals using primers flanking (A) a single-copy microsatellite (NZPR6), and (B) a multi-copy microsatellite (NZPR1). NZPR6 has a maximum of two fragments per individual, whereas at locus NZPR1, four to seven bands (mean = 4.8) are evident per individual. It is presumed that the multiple bands in loci like NZPR1 arise because there are multiple copies of the amplified region in the genome, each containing a different number of nucleotide repeats (see Discussion). The multiple bands are unlikely to be PCR artefacts because their relative sizes vary in different individuals, they show Mendelian inheritance and, in one case (see below), the bands reliably map to different genetic loci.

The seven most informative primer pairs (three single-copy and four multi-copy) were used to genotype the remaining 29 trees in our test population. All seven primer pairs gave highly polymorphic patterns. In the 48 tested trees, the single-copy microsatellites averaged 14 different size products, and the multi-copy microsatellites averaged 20 different size products (see discussion). Gene diversity for the single-copy

Table 1 Locus names, repeat sequences, number of bands in the population, number of bands per tree, haplotypes and gene-diversity values for the tested SSR loci

Source	SSR	Plasmid repeat	Polymorphism				
			No. of bands in population[a]	No. of bands per tree[b]		No. of haplotypes[b]	Gene diversity
				Average	Range		
Standard library	NZPR21	$(TC)_8(CT)_8(TC)_5$	1	1	1–1	1	0
	NZPR16	$(CT)_9$	2	2	2–2	1	N/A
	NZPR19	$(TC)_9$	1	1	1–1	1	0
	NZPR27	$(TC)_9$	1	1	1–1	1	0
	NZPR28	$(TC)_9$	1	1	1–1	1	0
	NZPR20	(CT) 9 & $(CT)_9C_{16}$	1	1	1–1	1	0
	NZPR24	$(TC)_{10}$	3	3	3–3	1	N/A
	NZPR17	$(CT)_{11}$	1	1	1–1	1	0
	NZPR23	$(TC)_{11}$	1	1	1–1	1	0
	NZPR25	$(CT)_{14}C_{11}$	1	1	1–1	1	0
	NZPR22	$(TC)_{18}$	1	1	1–1	1	0
	NZPR29	$(CA)_9$	1	1	1–1	1	0
	NZPR30	$(CA)_{11}$	1	1	1–1	1	0
Enriched libraries	NZPR9	$(CT)_{12}$	1	1	1–1	1	0
	NZPR10	$(CT)_{14}$	1	1	1–1	1	0
	NZPR11	$(CT)_{15}(AT)_{16}$	1	1	1–1	1	0
	NZPR12	$(CT)_{17}$	1	1	1–1	1	0
	NZPR13	$(CT)_{17}$	1	1	1–1	1	0
	NZPR14	$(CT)_{19}$	1	1	1–1	1	0
	NZPR6	$(CT)_{25}$	12	1.63	1–2	23	0.85
	NZPR7	$(CT)_{27}(AT)_{22}$	25	1.81	1–2	37	0.94[c]
	NZPR5	$(CT)_{29}$	16	2.9	2–4	37	N/A
	NZPR26	$(CT)_{33}$	N/A	N/A	5–7	N/A	N/A
	NZPR1	$(CT)_{17}$	28	5.06	3–8	42	N/A
	NZPR33	$(CT)_{14}$	N/A	N/A	1–2	N/A	N/A
	NZPR3	$(CCCT)_3(CT)_{14}$	5	1.58	1–2	13	0.73[c]
	NZPR4	$(CT)_{20i}$	15	3.26	3–5	28	N/A
	NZPR34	$(TC)_{21}(Tg)_{18}$	1	1	1–1	1	0
	NZPR2	$(CT)_{23}$	14	3.35	2–4	37	N/A
	NZPR35	$(CA)_{23}(TA)_3$	N/A	N/A	1–2	N/A	N/A
	NZPR36	$(CT)_{34}(AT)_{10}$	1	1	1–1	1	0
	NZPR37	$(gA)_{21i}$	N/A	N/A	3–6	N/A	N/A

[a] The number of different-size bands in the population ($n = 19$ for monomorphic loci and $n = 48$ for polymorphic loci)
[b] $n = 19$ unrelated trees for monomorphic loci and $n = 42$ unrelated trees for polymorphic loci
[c] These gene diversity (D) scores were calculated assuming that trees with only one band were homozygous.
N/A, not applicable

microsatellites ranged from 0.73 to 0.94 (Table 1). The relative informativeness for all seven primer pairs was compared by determining the number of distinct haplotypes that each pair revealed among the 42 unrelated trees. Primer pairs NZPR3, NZPR6 and NZPR7 produced 13, 23, and 37 haplotypes, respectively, and multi-copy primers NZPR1, NZPR2, NZPR4 and NZPR5 revealed 42, 37, 28 and 37 halotypes respectively (Table 1).

Mendelian inheritance of microsatellite alleles

Eleven primer pairs were tested for Mendelian segregation of microsatellite alleles. This step was particularly important for the multi-copy microsatellites, to ensure that the multiple bands were not artefacts of the PCR amplification. Both megagametophytes and progeny of known crosses were analysed.

Figure 2 shows an example of the inheritance analysis using F_1 progeny. Three primer pairs (NZPR4, NZPR5 and NZPR6) were multiplexed and tested on two parental trees and 20 of their progeny. Primer pair NZPR6 (top section of the gel) produced bands of 184 bp and 200 bp for parent 1, and 186 bp and 202 bp for parent 2. These four fragments segregated in a 1 : 1 co-dominant pattern in the 20 progeny, and were therefore considered to be allelic. Primer pair NZPR4 (middle section) has bands at 136 bp and 140 bp for parent 1, and at 136 bp and 138 bp for parent 2. Again the bands segregated 1 : 1 and were considered to be allelic. In addition, this primer pair produced a fragment at 128 bp, which was observed in all 48 trees of the production population. Primer pair NZPR5

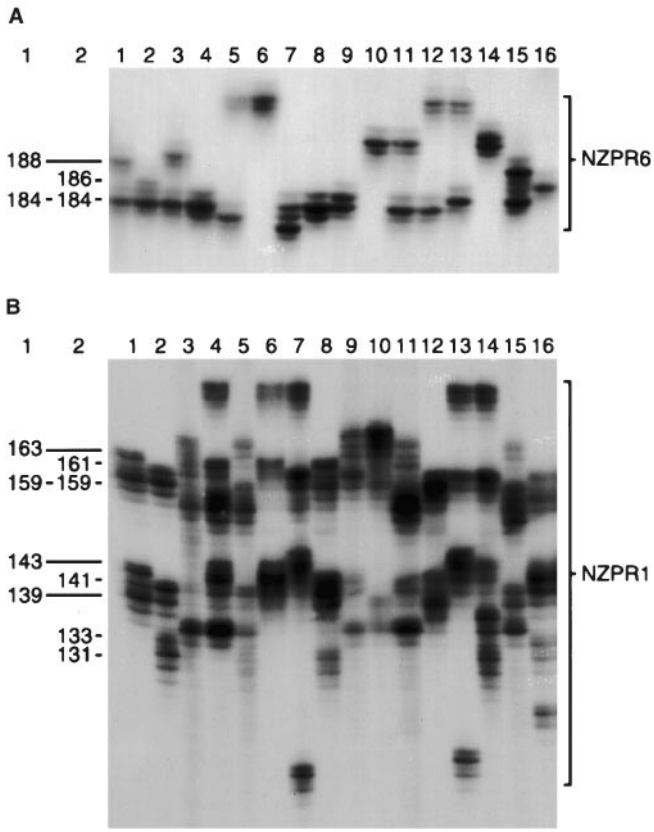

Fig. 1A, B Polymorphism among 16 *P. radiata* trees genotyped with the single-copy microsatellite NZPR6 (**A**), and the multi-copy microsatellite NZPR1 (**B**). The *numbers* to the left of the gels correspond to the size, in base pairs, of the alleles represented by the first two trees in each figure

(lowest section of the gel) amplified bands in two size ranges, marked as "large" and "small". The "large" bands were 114 bp and 112 bp from parent 1, and 116 and 104 bp from parent 2. These pairs of bands segregated 1 : 1 in the progeny as co-dominant alleles. The second "small" pair of bands amplified by this primer pair (84 and 80 bp in parent 1, 82 and 76 bp in parent 2) also segregated in Mendelian fashion. However, the "large" and "small" pairs of bands co-segregated suggesting that they are linked. For example, all progeny that inherited the 104-bp "large" band from parent 2 also inherited the 76-bp "small" band from this parent. There is also a 70-bp band common to both parents and their progeny. This band was only observed when the three primer pairs were multiplexed, and never when the primer pairs were used individually.

The multi-copy microsatellites varied with respect to the number of bands amplified in different progeny. For example, progeny and megagametophytes of each of the parents in lanes 1 and 2 of Fig. 1B were genotyped with microsatellite NZPR1. Parent 1 (see lane 1, Fig. 1B) has four alleles (139, 143, 159 and 163 bp) which segregated in megagametophytes as if there were two linked copies of the microsatellite. The second parent (lane 2, Fig. 1B) has six alleles (131, 133, 141, 141, 159 and 161 bp). The progeny always contained five alleles, two from parent 1, and three from parent 2. Megagametophytes from parent 2 either had alleles 131, 141 and 161, or alleles 133, 141 and 159, indicating three linked microsatellites in parent 2.

Linkage analysis

Linkage analysis was used to verify the inheritance and to investigate the distribution of eight microsatellites in the genome of *P. radiata* (three single-copy and five multi-copy). For linkage analysis we utilised 87 progeny of a cross currently being used to develop a molecular-marker genetic map in *P. radiata*. The results of the initial linkage analysis between pairs of microsatellite bands are shown in Table 2. The putative alleles at multi-copy microsatellite locus NZPR26 mapped to two unlinked loci, whereas the bands in the other multi-copy microsatellites segregated as two completely linked sets of alleles. Each of the single-copy microsatellites mapped to one locus each. NZPR1 and NZPR4 were linked within 22.5 cM (Kosambi function) and two other microsatellites (NZPR4 and one of the NZPR26 loci) were also linked (P. Wilcox, personal communication). The remaining loci were unlinked.

Amplification of *P. radiata* microsatellites in other conifers

Seven primer pairs flanking microsatellites in *P. radiata* were tested for amplification in 16 other conifer species. Template DNA quality was checked by performing PCRs using both PCT4 (Fisher et al. 1996) and primers flanking chloroplast locus PTCP2 (Powell et al. 1996). The amplification products from two primer pairs are shown in Fig. 3, and the results are summarised in Table 3. Four primer pairs, NZPR1, NZPR5 and NZPR7 and NZPR26, amplified bands from all six hard-pine species (subgenus *Pinus*), while the remaining three primer pairs amplified products from at least some of the species tested. Results from more distantly related species were less encouraging. Only three primer pairs (NZPR6, NZPR7 and NZPR26) amplified products from soft-pine species (subgenus *Strobus*), and in most cases the product was a single band. The most "promiscuous" primer pair (NZPR7) amplified a single band from several other conifers, but the bands were the same size in all of these species (Fig. 3B).

Null alleles

The progeny of 13 trees putatively homozygous at NZPR6 were tested to determine if the parents were truly homozygous or whether they contained a non-amplifying "null allele". Six of these thirteen trees

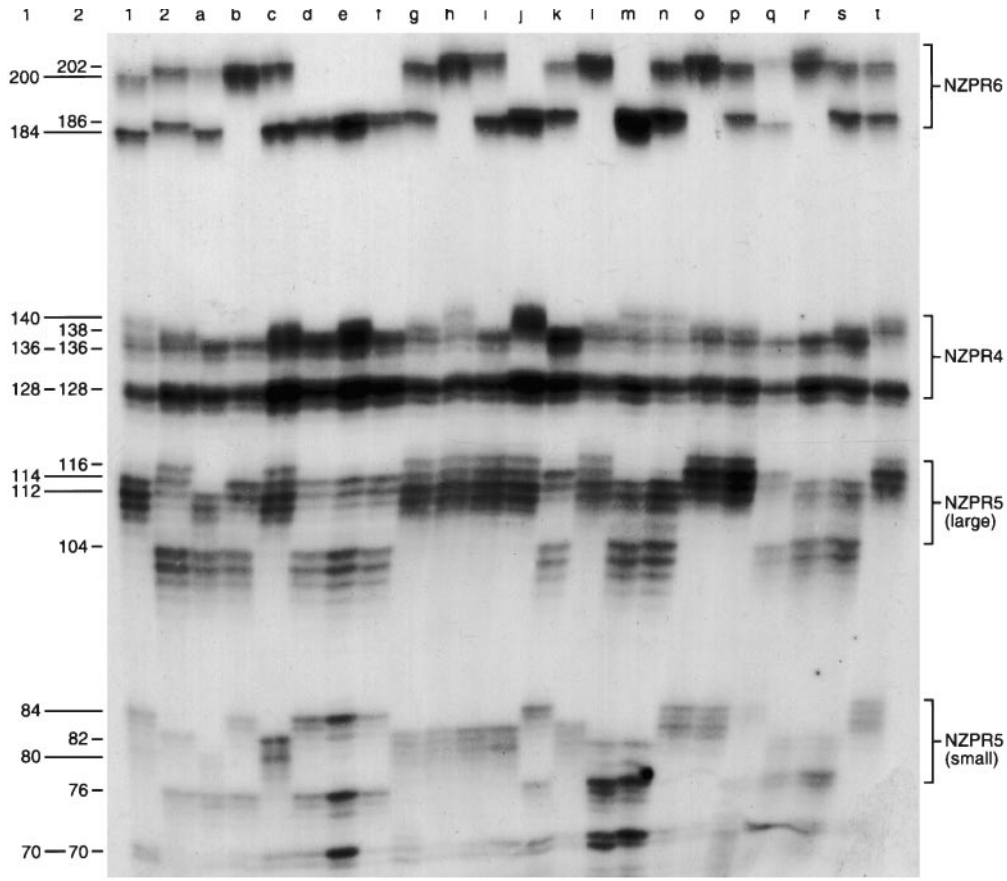

Fig. 2 Mendelian inheritance of microsatellites NZPR4, NZPR5 and NZPR6. The figure shows the results of a multiplex reaction with these three primer pairs using two parents (*lanes 1 and 2*) and 20 of their progeny (*lanes a–t*). Multi-copy microsatellite NZPR5 consists of segregating bands in both the "large" and "small" section. The *numbers* to the left of the gels correspond to the size, in base pairs, of the alleles

Table 2 Recombination fractions (with standard errors) of bands produced by eight primer pairs. The recombination fractions (θ) are between pairs of alleles (above diagonal) together with their standard errors (below diagonal)

Locus	Allele[a]	Multi-copy loci									Single-copy loci			
		NZPR1		NZPR2		NZPR5		NZPR26			NZPR4	NZPR6	NZPR7	NZPR35
		163	143	172	160	114	80	109	99	89				
NZPR1	163	–	0*	0.494	0.494	0.458	0.458	0.446	0.469	0.469	0.081	0.435	0.488	0.488
	143	0	–	0.494	0.494	0.458	0.458	0.446	0.469	0.469	0.081	0.435	0.488	0.488
NZPR2	172	0.139	0.139	–	0*	0.447	0.447	0.46	0.438	0.438	0.465	0.465	0.459	0.412
	160	0.139	0.139	0	–	0.447	0.447	0.46	0.438	0.438	0.465	0.465	0.459	0.412
NZPR5	114	0.137	0.137	0.139	0.139	–	0*	0.469	0.446	0.446	0.488	0.453	0.471	0.341
	80	0.137	0.137	0.139	0.139	0	–	0.469	0.446	0.446	0.488	0.453	0.471	0.341
NZPR26	109	0.156	0.156	0.159	0.159	0.159	0.159	–	0.484	0.484	0.385	0.297	0.477	0.4
	99	0.159	0.159	0.156	0.156	0.157	0.157	0.159	–	0*	0.492	0.492	0.477	0.406
	89	0.159	0.159	0.156	0.156	0.157	0.157	0.159	0	–	0.492	0.492	0.477	0.406
NZPR4		0.075	0.075	0.139	0.139	0.140	0.140	0.153	0.159	0.159	–	0.483	0.465	0.477
NZPR6		0.137	0.137	0.139	0.139	0.137	0.137	0.146	0.159	0.159	0.137	–	0.483	0.453
NZPR7		0.139	0.139	0.139	0.139	0.139	0.139	0.157	0.157	0.157	0.137	0.139	–	0.494
NZPR7		0.139	0.139	0.135	0.135	0.131	0.131	0.156	0.156	0.156	0.139	0.137	0.139	–

[a] Allele size in base pairs
* Significant linkages ($P < 0.01$)

were found to contain a null allele at this locus (data not shown). In addition, the mapping experiment revealed a NZPR7 null allele in parent 2 (Fig. 4). Null alleles were not observed at the other seven microsatellite loci mapped in this pedigree.

Novel alleles at locus NZPR7

Primer pair NZPR7 exhibited a high frequency of non-parental bands in the progeny of the mapping pedigree, as well as in the megagametophytes of both mapping

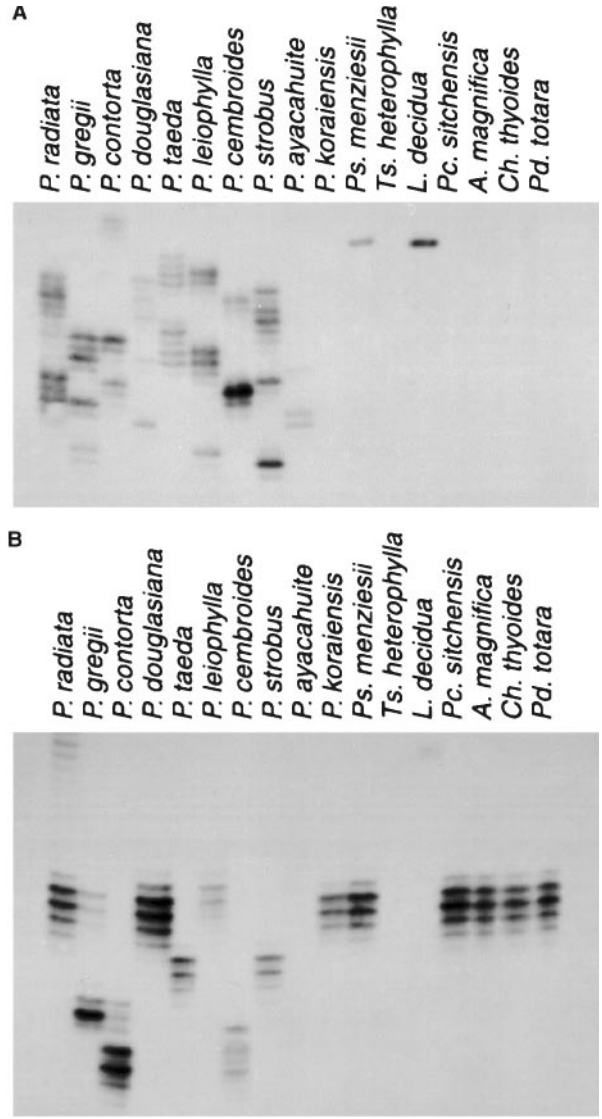

Fig. 3A, B Amplification of primer pairs NZPR26 (**A**) and NZPR7 (**B**) in 17 conifer species. Fragments were amplified in nine conifer species using multi-copy primer pair NZPR26 and in 13 species with single-copy primer pair NZPR7 (*B*). Pines of the subgenus *Pinus* (*lanes 1–6*), subgenus *Strobus* (*lanes 7–10*), and other conifer genera (*lanes 11–17*) are shown

pedigree parents. Figure 5 illustrates some of these novel alleles. Parent 1 (lane 1) has NZPR7 alleles of 145 and 121 bp , and a second multiplexed microsatellite (PR 9.3) giving 97- and 81-bp alleles. Parent 2 (lane 2) has one NZPR7 allele of 113 bp and a null allele (see above), as well as two 99-bp alleles at PR9.3. One progeny (lane a) inherited a single non-parental allele of 153 bp (presumably derived from parent 1), with a null allele derived from parent 2. A second progeny (lane b) has NZPR7 bands at 121 bp and 145 bp . Presumably, one of the NZPR7 bands was derived from parent 1, and the other allele was a novel allele from parent 2. Figure 5 illustrates non-parental NZPR7 alleles of 105, 135 and 121 bp in megagametophytes of parent 1 (lane c) and parent 2 (lanes d and e). All progeny and megagametophytes containing non-parental NZPR7 alleles were genotyped with 3–8 other microsatellites to confirm that they were progeny of the intended parents.

The frequency of novel alleles at NZPR7 was very high. Two out of eighty seven mapping progeny, 2 out of 45 megagametophytes from parent 1, and 6 out of 40 megagametophytes from parent 2, had non-parental NZPR7 alleles. Interestingly, five of the novel alleles from parent 2 were the same size (121 bp). Thus, approximately 6% of both megagametophytes and progeny displayed non-parental NZPR7 alleles. All non-parental alleles were amplified by PCR at least three times, and the size of the fragment was consistent.

Discussion

The frequency of microsatellites in the un-enriched small insert library corresponds to approximately one $(CT)_n$ repeat every 1.3 Mb, and one $(CA)_n$ repeat ($n > 8$) every 4.8 Mb of *P. radiata* genomic DNA. This result confirms the findings of Smith and Devey (1994) which suggested that microsatellites are at least an order of magnitude less abundant in *P. radiata* than has been estimated in other plant species (Condit and Hubbell 1991; Lagercrantz et al. 1993; Morgante and Olivieri 1993). Both $(CT)_n$ and $(CA)_n$ repeats also appear to be less abundant in *P. radiata* than in other pines (Kostia et al. 1995; Echt and May-Marquardt 1997). However, it is difficult to compare microsatellite abundance in pine species because SSR clones derived from *P. strobus* and *P. taeda* contained much longer inserts (14–23 kb, Echt and May-Marquardt 1997), which increases the likelihood that clusters of microsatellites could exist in the same clone (Condit and Hubbell 1991).

The enrichment method described by Fisher et al. (1996) increased the frequency of CT repeats more than 300-fold, and the frequency of CA repeats more than 600-fold, compared with the standard library (for microsatellites with 12 or more repeats). However, there was evidence for increased redundancy following the SSR enrichment, as 33 of the 42 clones had unique sequences, compared with 28 of 31 unique clones in the un-enriched library.

The average size of the microsatellite motifs characterised from the un-enriched library was shorter than those observed in studies involving other plant species (Lagercrantz et al. 1993). Because all of the microsatellites that were characterized from this library were monomorphic, the enriched libraries were screened with longer oligonucleotides to obtain longer microsatellites. This strategy proved successful, and the longer microsatellites resulted in polymorphic markers. A similar correlation between microsatellite length and polymorphism has been reported in other (Weber 1990; Wu and Tanksley 1993; Terauchi and Konuma 1994),

Table 3 The number of bands amplified in various conifers using PCR primers designed from *P. radiata*. Individuals with no PCR products are represented with a dash (–)

	Locus						
	NZPR1	NZPR2	NZPR4	NZPR5	NZPR6	NZPR7	NZPR26
Pinus species							
Hard pines subgenus *Pinus*)							
P. radiata	4	4	3	4	2	2	4
P. gregii	3	4	3	2	–	2	4
P. contorta	4	3	–	2	2	2	3
P. douglasiana	4	3	1	2	1	2	5
P. taeda	3	1	–	1	1	1	5
P. leiophylla	5	–	–	1	1	1	3
Soft pines (subgenus *Strobus*)							
P. cembroides	–	–	–	–	–	2	2
P. strobus	–	–	–	–	–	1	4
P. ayacahuite	–	–	–	–	–	–	2
P. koraiensis	–	–	–	–	1	1	–
Non-Pinus coniferae							
Ps. menziesii	–	–	–	–	–	1	–
Ts. heterophylla	–	–	–	–	–	–	–
L. decidua	–	–	–	–	–	–	–
Pc. sitchensis	–	–	–	–	–	1	–
A. magnifica	–	–	–	–	–	1	–
Ch. thyoides	–	–	–	–	–	1	–
Pd. totara	–	–	–	–	–	1	–

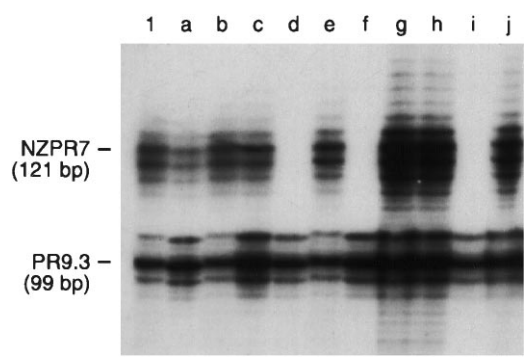

Fig. 4 Inheritance of a null allele. Primer pairs flanking microsatellites NZPR7 and PR9.3 were multiplexed and tested on a parent and ten of its haploid megagametophytes. A null allele at locus NZPR7 segregates 1:1 with the 121-bp allele in the megagametophytes (*lanes a–j*) of the parent (*lane 1*). The parent was homozygous at locus PR9.3, and its 99-bp alleles are present in all ten megagametophytes

but not all (Plaschke et al. 1995; Szewe-McFadden et al. 1996), species. These longer microsatellites from *P. radiata* had high gene-diversity scores (0.73–0.94), suggesting that they will be useful for various genetic analyses.

Our results suggest that the size threshold for polymorphic microsatellites is higher in *P. radiata* than in most other species. In *P. radiata* the majority of microsatellites with less than 20 repeats proved to be monomorphic, and a minimum repeat number of 17–20 was a useful guide for determining whether to design primers for a newly cloned microsatellite locus. In contrast, in other plant and animal species dinucleotide repeats of 10–12 are often successufully targeted and used as molecular markers (Edwards et al. 1991; Plaschke et al. 1995; Bowers et al. 1996; van de Ven and McNicol 1996; Echt and May-Marquardt 1997; Guilford et al. 1997). Nonetheless, we have demonstrated that many long polymorphic dinucleotide repeats can be obtained with a suitable enrichment procedure, and that these microsatellites are typically informative.

Multi-copy microsatellite loci

An unexpected feature of the tested microsatellites was that many of the primer pairs amplified more than two bands per individual. On gels these bands appeared like typical microsatellite amplification products in that they produced stutter patterns, varied in their relative sizes between individual trees, and mapped in a segregating population. We argue below that multiple bands result from the presence of multiple copies of the microsatellite in the genome, and that such multi-copy markers can be advantageous in some instances.

The *P. radiata* genome is very large (approximately 44 pg for nuclear 2C DNA, O'Brien et al. 1996) and reassociation kinetic studies suggest that approximately 75% is moderately to highly repetitive DNA (Dhillon 1987). In our isolation procedure, no attempts were made to eliminate multi-copy DNA, and therefore

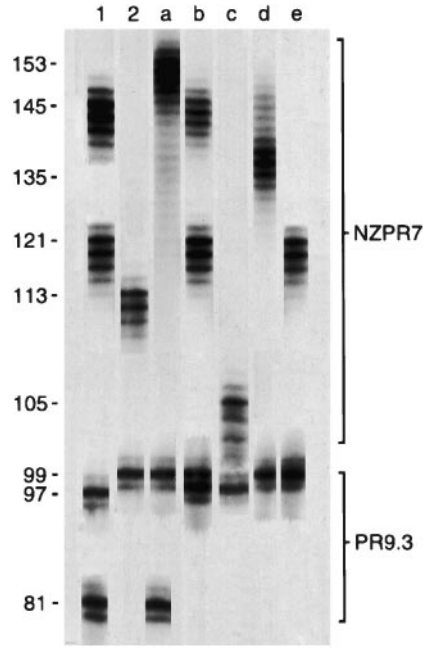

Fig. 5 Non-parallel alleles at locus NZPR7. The seven lanes in this figure were spliced from a 24-lane gel. Results are shown for the parents of the mapping population and for selected megagametophytes and progeny. Parent 1 (*lane 1*) has alleles at 145 bp and 121 bp for NZPR7 and alleles at 97 bp and 81 bp for PR9.3; parent 2 (*lane 2*) has NZPR7 alleles at 113 bp and a null allele, and a pair of 99-bp alleles, at locus PR9.3. Non-parental alleles at locus NZPR7 for each of the progeny are: 153 bp (progeny 1, *lane a*), either 145 bp or 121 bp (progeny 2, *lane b*), 105 bp (megagametophyte, *lane c*), 135 bp (megagametophyte, *lane d*) and 121 bp (megagametophyte, *lane e*). Only parental alleles were observed at locus PR9.3 for all samples

it was not surprising to find that some microsatellites were present in multiple copies.

Both the sequence data and pedigree studies demonstrates that some microsatellites were truly multi-copy, or embedded in repetitive DNA. Firstly, analysis of the relevant regions of the plasmid sequences from these loci revealed no secondary binding sites for the primers adjacent to the microsatellite repeats. Secondly, in two of the sequenced plasmid clones from the enriched library, multi-copy microsatellites were observed embedded within tandemly repeated non-microsatellite DNA. Thirdly, the same repeat-containing clone (NZPR24, Table 1) was obtained four times (in 35 000 clones) from the un-enriched, small-insert library, suggesting thousands of copies of this sequence exist in the *P. radiata* genome. Fourthly, four of the five multi-copy microsatellites had completely linked fragments, which is consistent with a tandemly repeated region, while the fifth primer pair (NZPR26) amplified two unlinked loci, suggesting that very similar sequences are present at two different sites in the genome.

It is possible that the high frequency of multi-copy microsatellites observed here was partly a result of the enrichment procedure employed, rather than a true reflection of the *P. radiata* genome. The enrichment procedure involves amplifying regions of the genome where two adjacent and oppositely oriented microsatellites provide primer-binding sites. The extent to which this enrichment procedure targeted microsatellites within repetitive DNA is difficult to assess, because no polymorphic microsatellites were obtained from the un-enriched library. The only other reported method of microsatellite isolation from *P. radiata* included a screen of a phage library with total genomic DNA. The results of this screen indicated that 25% of the clones contained highly repetitive DNA (Smith and Devey 1994). We were also concerned that this approach might identify a biased subset of microsatellites from particular regions of the genome. This proved not to be the case, as the eight mapped microsatellite loci were reasonably well spread throughout the genome. These loci mapped to seven linkage groups, with 8 cM as the closest linkage (P. Wilcox, personal communication). There was, however, strong linkage of bands within multi-copy primer pairs, as four out of five of these primer pairs produced sets of tightly linked bands. Not enough microsatellites were mapped to exclude the possibility that microsatellite loci are clustered in the *P. radiata* genome.

Although the simultaneous amplification of multi-copy microsatellite loci can sometimes be difficult to analyse when the bands overlap, there are advantages when their inheritance patterns can be discerned. Firstly, many alleles can be scored per individual from one set of specific primers. For example, the complete genotype for all the NZPR1 alleles enabled us to unambiguously differentiate every tree in the 47 tree-production population (data not shown). This level of discrimination was not possible with any of the single-locus markers. A second advantage occurs when the multiple copies are at the same locus. This increases the chance that the locus will show polymorphism and can be used for genetic analysis. Lastly, unlinked multiple loci can be assessed simultaneously (e.g. NZPR26), providing a natural multiplex for mapping, parentage testing, and other analyses.

All seven of the *P. radiata* microsatellites tested amplified fragments in several other hard-pine species. Therefore, it seems likely that a reasonable proportion of microsatellites isolated in one *Pinus* species may be useful in closely related species. Collaborations and exchange of microsatellite primers should save time and money, and this result suggest that some microsatellites should be useful for mapping inter-specific crosses (Paterson et al. 1990; Grandillo and Tanksley 1996). The near total absence of amplification products from the subgenus *Strobus* reaffirms the notion that most microsatellites isolated from hard pines will not be useful in soft pines. Previously the converse had been found. That is, microsatellites isolated from the soft pine, *P. strobus*, were generally unsuccesful in amplifying hard-pine DNA (Echt et al. 1996).

Frequent non-parental alleles were observed at microsatellite NZPR7 (6% of progeny and megagametophytes). To our knowledge, this is the first report of mutation characteristics for a plant microsatellite. This mutation rate is approximately 100-fold greater than the average rate in human (Weber and Wong 1993), and 500-fold greater than the average rate for sheep dimeric repeats (Crawford and Cuthbertson 1996). Each of the ten non-parental NZPR7 alleles was at least four repeat units different from the parental allele. This seems to contradict the stepwise mutation model which postulates single repeat-unit mutations (Valdes et al. 1993).

The presence of segregating null alleles in one tree at NZPR7 and six unrelated trees at NZPR6 serves as a caution when using untested microsatellites as genetic markers. Individuals with a null allele at a given locus would typically be scored as homozygous and this error in genotypic assignment can have important implications for population and parentage studies (Thomas et al. 1994; Pemberton et al. 1995). Because it is usually impractical to determine the frequency of null alleles at each locus by inheritance studies, and because null alleles often exist at low frequencies, they tend to be ignored (Pemberton et al. 1995). However, when correct genotypic assignment is crucial, either homozygosity must be confirmed or only heterozygous loci should be used (Gill et al. 1995; Morton et al. 1995).

In summary, our results suggest that the abundance of microsatellites (low), the repeat-length threshold for polymorphism (high), the frequency of multi-copy loci (high), the frequency of null alleles (high), and perhaps the mutation rate at some loci (high), are fundametally different in *P. radiata*, and perhaps pines generally, compared with other plant species. In spite of these difficulties, the characterisation of several highly polymorphic microsatellites in *P. radiata* has allowed us to demonstrate their usefulness for genotyping individuals, parentage analysis, genetic diversity studies and the construction of genetic linkage maps. Additional microsatellite loci are currently being developed to construct a *P. radiata* framework linkage map composed entirely of microsatellite markers.

Acknowledgements This work was supported by grants from the FRI-GEENZ joint-venture, and New Zealand's Public Good Science Fund (CO4310 and NSOF). We thank G. Corbett and J. Lee (FRI) for *P. radiata* tissue and DNA, P. Wilcox (FRI) for his help with genetic mapping, and K. Weising (Wolfgang Goethe University, Frankfurt) for laboratory assistance.

References

Adams WT, Neale DB, Loopstra CA (1988) Verifying controlled crosses in conifer tree-improvement programs. Silvae Genet 37:147–152

Binelli G, Bucci G (1994) A genetic linkage map of *Picea abies* Karst., based on RAPD markers as a tool in population genetics. Theor Appl Genet 88:283–288

Bowers JE, Dangl GS, Vignani R, Meredith CP (1996) Isolation and characterisation of new polymorphic simple sequence repeat loci in grape (*Vitis vinifera* L.). Genome 39:628–633

Byrne M, Murrell JC, Allen B, Moran GF (1995) An integrated genetic linkage map for eucalypts using RFLP, RAPD and isozyme markers. Theor Appl Genet 91:869–875

Cadalen T, Boeuf C, Barnard S, Bernard M (1997) An intervarietal molecular marker map in *Triticum aestivum* L. Em. Thell. and comparison with a map from a wide cross. Theor Appl Genet 94:367–377

Cato SA, Richardson TE (1996) Inter- and intra-specific polymorphism at chloroplast SSR loci and the inheritance of plastids in *Pinus radiata* D. Don. Theor Appl Genet 93:587–592

Charters YM, Robertson A, Wilkinson MJ, Ramsey G (1996) PCR analysis of oilseed rape cultivars (*Brassica napus* L. spp. oleifera) using 5′-anchored simple sequence repeat (SSR) primers. Theor Appl Genet 92:442–447

Condit R, Hubbell P (1991) Abundance and DNA sequence of two-base repeat regions in tropical genomes. Genome 34:66–71

Crawford AM, Cuthbertson RP (1996) Mutations in sheep microsatellites. Genome Res 6:876–879

Dale G, Chaparro J (1996) Integration of molecular markers into tree-breeding and improvement programs. In: Dieters MJ, Matheson AC, Nikles DG, Harwood CE, Walker SM (eds) Tree improvement for sustainable tropical forestry. Proc QFRI-IUFRO Conf, Caloundra, Australia. pp 472–477

Devey ME, Bell JC, Smith DN, Neale DB, Moran GF (1996) A genetic linkage map for *Pinus radiata* based on RFLP, RAPD, and microsatellite markers. Theor Appl Genet 92:673–679

Dhillon SS (1987) DNA in tree species. In: Bonga JM, Durzan DJ (eds) Cell and tissue culture in forestry. Nijhoff, Dordrecht, Netherlands, pp 298–313

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

Echt CS, May-Marquardt P, Hseih M, Zahorchak R (1996) Characterisation of microsatellite markers in eastern white pine. Genome 39:1102–1108

Edwards A, Civitello A, Hammond HA, Caskey CT (1991) DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. Am J Hum Genet 49:746–756

Eldridge KG (1982) Genetic Improvements from a radiata pine seed orchard. NZ J For Sci 12:404–411

Fisher PJ, Gardner RC, Richardson TE (1996) Single-locus microsatellites isolated using 5′-anchored PCR. Nucleic Acids Res 24:4369–4371

Gardiner SE, Bassett HCM, Noiton DAM, Bus VG, Hofstee ME, White AG, Ball RD, Forster RLS, Rikkerink EHA (1996) A detailed linkage map around an apple scab resistance gene demonstrates that two disease resistance classes both carry the V_f gene. Theor Appl Genet 93:485–493

Gill P, Kimpton CP, Urquhart A, Oldroyd N, Millican ES, Watson SK, Downes TJ (1995) Automated short tandem repeat (STR) analysis in forensic casework – a strategy for the future. Electrophoresis 16:1543–1552

Gleed JA (1982) Tree improvement – first results from a radiata pine production forest. Appita 36:386–390

Grandillo S, Tanksley SD (1996) Genetic analysis of RFLPs, GATA microsatellites and RAPDS in a cross between *L. esculentum* and *L. pimpinellifolium*. Theor Appl Genet 92:957–965

Grattapaglia D, Sederoff R (1994) Genetic linkage maps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudo-testcross: mapping strategy and RAPD markers. Genetics 137:1121–1137

Guilford P, Prakash S, Zhu JM, Rikkerink E, Gardiner S, Bassett H, Forster R (1997) Microsatellites in *Malus* × *domestica* (apple): abundance, polymorphism and cultivar identification. Theor Appl Genet 94:249–254

Johnson GR, Firth A, Brown PC (1992) Value gains from using genetically improved radiata pine stock. NZ For (Feb) 14–18

Kashi Y, King D, Soller M (1997) Simple-sequence repeats as a source of quantitative variation. Trends Genet 13:74–78

Kostia S, Varvio S-L, Vakkari P, Pulkinen P (1995) Microsatellite sequences in a conifer, *Pinus sylvestris*. Genome 38 : 1244–1248

Lagercrantz U, Ellegren H, Andersson L (1993) The abundance of various polymorphic microsatellite motifs differs between plants and vertebrates. Nucleic Acids Res 21 : 1111–1115

Lander E, Green P, Abrahamson J, Barlow A, Daly M, Lincoln S, Newberg L (1987) MAPMAKER: an interactive computer program for constructing primary linkage maps of experimental and natural populations. Genomics 1 : 174–181

Morgante M, Olivieri AM (1993) PCR-amplified microsatellites as markers in plant genetics. Plant Jour 3 : 175–182

Morton NE, Collins A (1995) Statistical and genetic aspects of quality control for DNA identification. Electrophoresis 16 : 1670–1677

Neale DB, Devey ME, Jermstad KD, Ahuja MR, Alosi MC, Marshall KA (1992) Use of DNA markers in forest tree improvement research. New For 6 : 391–407

O'Brien IEW, Smith DR, Gardner RC, Murray BG (1996) Flow cytometric determination of genome size in *Pinus*. Plant Sci 115 : 91–99

Paterson AH, DeVerna JW, Lanini B, Tanksley SD (1990) Fine mapping of quantitative trait loci using selected overlapping recombinant chromosomes, in an interspecies cross of tomato. Genetics 124 : 735–742

Pemberton JM, Slate J, Bancroft DR, Barrett JA (1995) Nonamplifying alleles at microsatellite loci : a caution for parentage and population studies. Mol Ecol 4 : 249–252

Plaschke J, Ganal MW, Röder MS (1995) Detection of genetic diversity in closely related bread wheats using microsatellite markers. Theor Appl Genet 91 : 1001–1007

Powell W, Morgante M, McDevitt R, Vendramin GG, Rafalski JA (1995) Polymorphic simple sequence repeat regions in chloroplast genomes : applications to the population genetics of pines. Proc Natl Acad Sci USA 92 : 7759–7763

Powell W, Morgante M, Andre C, Hanafey M, Vogel J, Tingey S, Rafalski A (1996) The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. Mol Breed 2 : 225–238

Queller DC, Strassmann JE, Hughes CR (1993) Microsatellites and kinship. Trends Ecol Evol 8 : 285–288

Rafalski JA, Tingey SV (1993) Genetic diagnostics in plant breeding : RAPDs, microsatellites and machines. Trends Genet 9 : 275–280

Sambrook J, Fritch EF, Maniatis T (1989) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York

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

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

Terauchi R, Konuma A (1994) Microsatellite polymorphism in *Dioscorea tokoro*, a wild yam species. Genome 37 : 794–780

Thomas MR, Cain P, Scott NS (1994) DNA typing of grapevines: a universal methodology and database for describing cultivars and evaluating genetic relatedness. Plant Mol Biol 25 : 939–949

Valdes AM, Slatkin M, Freimer NB (1993) Allele frequencies at microsatellite loci: the stepwise mutation model revisited. Genetics 133 : 737–749

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

Vincent TG (1997) Radiata pine tree breeding. NZ Tree Grower 18 : 21–24

Weber JL (1990) Informativeness of human $(dC\text{-}dA)_n \cdot (dG\text{-}dT)_n$ polymorphisms. Genomics 7 : 524–530

Weber LJ, Wong C (1993) Mutation of human short tandem repeats. Hum Mol Genet 2 : 1123–1128

Weir BS (1990) Genetic data analysis. Sinauer Associates, Inc. Publishers, Sunderland, Massachusetts

Wu K-S, Tanksley SD (1993) Abundance, polymorphism and genetic mapping of microsatellites in rice. Mol Gen genet 241 : 225–235

Xiao J, Grandillo S, Ahn SN, McCouch SR, Tanksley SD (1996) Genes from wild rice improve yield. Nature 384 : 223–224