FLUORESCENTLY LABELLED, MULTIPLEXED CHLOROPLAST MICROSATELLITES FOR HIGH-THROUGHPUT PATERNITY ANALYSIS IN *PINUS RADIATA*

J. KENT and T. E. RICHARDSON

New Zealand Forest Research Institute Limited,
Private Bag 3020, Rotorua, New Zealand.

(Received for publication 24 November 1997; revision 1 April 1998)

ABSTRACT

Plants have distinct genomes in their nucleus, chloroplast, and mitochondria. Because the chloroplast genome in *Pinus* species is normally paternally inherited, DNA marker systems capable of detecting polymorphisms in the chloroplast molecule would be useful for studying paternity and pollen movement. Short nucleotide repeats analogous to nuclear microsatellites, or simple sequence repeats (SSRs), have been previously identified from DNA sequence analysis of the *Pinus thunbergii* Parl. chloroplast genome, and polymerase chain reaction (PCR) primers have been designed to amplify 20 of the chloroplast microsatellite loci (cpSSRs). Seven of the 20 cpSSR sites were found to be polymorphic in 51 unrelated *P. radiata* D.Don trees with between two and four alleles per locus. These seven loci have been combined in two fluorescently labelled multiplexed PCRs to generate chloroplast haplotypes resolved in single electrophoretic lanes. This multiplexed protocol largely automates the collection of chloroplast haplotype data for clonal identification, paternity analysis, pollen genotyping, and studying chloroplast diversity in natural populations of *P. radiata* and other conifers.

Keywords: chloroplast haplotypes; chloroplast simple sequence repeats; multiplex polymerase chain reaction; paternity analysis; *Pinus radiata*.

INTRODUCTION

In recent years various DNA analysis methods have been integrated into forest tree improvement programmes worldwide, and especially into New Zealand's *Pinus radiata* improvement programme. DNA marker systems are now used in New Zealand to "fingerprint" and track elite clones in traditional tree-improvement programmes and through various tissue culture and other micropropagation systems, to compare the genetic diversity in unimproved and improved populations, and to verify pedigrees in long-term trials (Richardson, Cato, Corbett, Fisher, Kent, Warr, Yoon 1997; Richardson, Burdon, Carson, Kent, Kumar, Lee, Mackenzie, MacLeod 1997). DNA markers have also been used to identify alleles with important phenotypic effects in *P. radiata* (Kuang *et al.* in press).

In contrast to many tree-improvement programmes which rely mainly on open-pollination and/or supplementary mass-pollination of elite clones serving as maternal parents to achieve genetic gain, New Zealand's *P. radiata* breeding programme uses control-pollination where knowledge of both parents is used to maximise genetic gain (Vincent 1997). This control-pollinated breeding programme places greater emphasis on the paternal parentage in the most advanced genetic material, and DNA markers provide a tool for monitoring this through paternity testing.

Not all DNA marker systems are equally well-suited for paternity analysis. Most profiling marker systems used for "fingerprinting" and linkage mapping are biallelic and dominant (e.g., RAPD and AFLP), which limits the discrimination at individual loci. In addition, AFLP requires extensive DNA modifications prior to analysis, while RAPD requires careful monitoring to ensure reliable assessment across generations, between labs, and over time. Because microsatellite, or simple sequence repeat (SSR), markers in the nuclear genome are multi-allelic, co-dominant, and faithfully transmitted across generations they are widely used for parentage and paternity testing (Powell *et al.* 1996). One disadvantage of nuclear markers for paternity analysis is that the individual's genotype at each locus is composed of both maternal and paternal alleles. In some circumstances—for example, when the parents share an allele(s) which is common in the population—the presence of the maternal allele (and the inability to distinguish it from the paternal allele) reduces the ability to detect pollen contamination. For paternity testing, an ideal marker system should unambiguously assay the pollen parent's contribution to progeny.

In most conifers, including *P. radiata* (Cato & Richardson 1996, and reports cited by them), chloroplasts are typically transmitted to progeny from the pollen parent. The demonstration that regions of the conifer chloroplast genome reminiscent of nuclear SSR are polymorphic within species (Powell *et al.* 1995; Cato & Richardson 1996; Vendramin *et al.* 1996; Echt *et al.* in press) provides an opportunity to use these chloroplast SSR regions (cpSSR) for paternity testing without interference from the maternal genome.

In this paper we report further testing of cpSSR for polymorphism in *P. radiata* and the development of a robust and efficient multiplexed protocol to produce chloroplast haplotypes which can be assayed in single lanes on automated DNA sequencing systems. This protocol provides an avenue for very high-throughput testing of cpSSR haplotypes for paternity and other genetic analyses of chloroplast genome variability.

MATERIALS AND METHODS

Genomic DNA was extracted from needles and buds from 51 genetically diverse *P. radiata* individuals using a modified CTAB method as previously described (Cato & Richardson 1996), and from embryos dissected from mature seeds using Bio 101 Inc. FastDNA™ Kit H (following manufacturer's instructions).

PCR Conditions

Single locus

Twenty chloroplast SSR loci were tested for polymorphism on the 51 trees using primers designed from the published *P. thunbergii* chloroplast genome sequence (Wakasugi *et al.* 1994; Vendramin *et al.* 1996; W. Powell, unpubl. data) (Table 1). Single-locus PCR

TABLE 1–The cpSSR regions evaluated for polymorphism in *Pinus radiata*

Genome position*	Locus name†	Repeat unit	PCR primer sequences Forward sequence (5'–3') Reverse sequence (5'–3')	Observed fragment sizes in base pairs (bp)
1289	cpSSR 1	$(T)_{17}$	TCC TGG TTC CAG AAA TGG AG ‡ TAA TTT GGT TCC AGA ATT GCG	107, 108, 109
9434	cpSSR 2	$(A)_{10}$	AAA CTG ACG TAG ATG CCA TGG ‡ GCG GTA TGA GGG AAG AAG C	134
15211	cpSSR 3	$(C)_8(T)_8A(T)_8$	CTT GGA TGG AAT AGC AGC C § GGA AGG GCA TTA AGG TCA TTA	109, 110, 111
26106	cpSSR 4	$(A)_{14}$	AAT CCG ACA AAA AAG ATT CGG ‡ GCT CCA TTT CAC GTG GTT G	148, 149, 150
30277	cpSSR 5	$(A)_{12}(G)_{10}$	TGT TGA TGT CGT AGC GGA AG ‡ ATG AAA TGA ATC ACT TCC CCC	132, 133
36567	cpSSR 6	$(T)_{11}$	AAA AGA GGA GGA AAA ACA CCT T ‡ AAG AGC AGA CAA GTA AGG GGC	111
41131	cpSSR 7	$(T)_{11}$	TCC CGA AAA TAC TAA AAA AGC A § CTC ATT GTT GAA CTC ATC GAG A	75
45071	cpSSR 8	$(T)_{15}$	AAG TTG GAT TTT ACC CAG GTG § GAA CAA GAG GAT TTT TTC TCA TAC A	162
48256	cpSSR 9	$(T)_{10}$	ACG TTG GAC CAG AGC AGG ‡ CGA ATT TTT CGA AGA ACT AGC G	117, 138 ¶
51928	cpSSR 10	$(T)_{10}$	CTT TCT ACG GAA CGG AAA AGG ‡ GCA CTG CGG GAA AAA AAT AA	144
63771	cpSSR 11	$(T)_{10}$	TGA ACG TGC CAT GAT CAA TT ‡ GGG GCT ATA GTG CAC TTG GAA	138
71987	cpSSR 12	$(T)_{16}$	TCT TTG CAA GAA GGA TGG CT ‡ GGG GAG TAA TCC GTG GAA TT	111, 112
79987	cpSSR 13	$(A)_{12}$	CTT TTG TTT TTC AAC AAT TGC A § ACA TCT ATC TCC CAT ATC GGC	141, 142
87314	cpSSR 14	$(T)_{14}$	TCC AGG ATA GCC CAG CTG ‡ TAT ATC CCC CGT ACT TGG ACC	110
100842	cpSSR 15	$(A)_{10}$	ATA CGT ATG TAT CCC CTA ACT GTC A § TCA ATT TTT GCC ATA TCC TGA	112
102652	cpSSR 16	$(T)_{11}$	TTC CCA GAT CCA TTG AAA TAC A ‡ TAT GTG CGC GAT AAT TTC CA	114
107165	cpSSR 17	$(A)_{10}$	GTT CAT TCG GGA TCC TTA AAA § GTA CTT TCC TTC AGC CAA TCT G	122
107569	cpSSR 18	$(T)_{11}$	AAA GCT TTT ATT GCG GCC § ATG GCA GTT CCA AAA AAG C	95
109612	cpSSR 19	$(A)_{11}$	ATC GAA CAA CGA GAA TAA TCC A ‡ TTG GGG GTG ATA GTG GAA AA	150
110074	cpSSR 20	$(A)_{10}$	TAA GGG GAC TAG AGC AGG CTA § TTC GAT ATT GAA CCT TGG ACA	92, 93, 94, 95

* Genome position of first base in repeat sequence (Wakasugi *et al.* 1994)
† These locus names replace those used by Cato & Richardson (1996)
‡ Primer sequences provided by W. Powell.
§ Primer sequences based on those used by Vendramin *et al.* (1996)
¶ Two monomorphic fragments. See text for details.

amplifications were carried out in a total volume of 10 µl containing 12 ng of template DNA, 0.4 µ*M* of each primer, 2 m*M* $MgCl_2$, 0.1 m*M* each of dNTP (dATP, dGTP, dTTP, dCTP), 0.5 units of *Taq* polymerase (Boehringer Mannheim), 10 m*M* TRIS-HCl (pH 8.3), 50 m*M* KCl, and one of three possible [Fluorescent]dUTPs (PE Applied Biosystems) at varying concentrations: 2 µ*M* [R110]dUTP, 2 µ*M* [R6G]dUTP, 8 µ*M* [TAMRA]dUTP. Samples in 96-well polycarbonate plates were overlaid with paraffin oil and PCR were carried out on Techne PHC-3 thermal cyclers with 1 cycle at 93°C for 3 minutes, 30 cycles of 30 seconds at 94°C, 30 seconds at 58°C, and 30 seconds at 72°C. Individual PCR products with different fluorescent labels ([F]dUTPs) and sizes were then combined to run in a single lane on a PE Applied Biosystems 373A DNA Sequencer. For each lane, 1 µl of the pooled PCR samples was mixed with 0.5 µl size standard (GeneScan-500 [TAMRA] or GeneScan-500 [ROX] (PE Applied Biosystems)), 2 µl deionised formamide, 0.5 µl dextran blue dye loading buffer (4 µl in well), heated to 90°C for 2 minutes and loaded on to a 0.4 mm, 24 well, 6% acrylamide, 12 cm well-to-read gel, made with 7 *M* Urea, 1X TBE. The gel was then run at 2500V/23 watts for 3 hours to allow the 350 bp size standard fragment to be detected.

Multiplex PCR

Dye-labelled primers incorporating a fluorescent label (6-FAM, 5′HEX or TET (ABI)) at the 5′ end of the forward primer were synthesised for the polymorphic loci identified from the single locus trials described above. Where allele size ranges overlapped, the loci were differentiated by using different dye labels. Optimisation experiments were undertaken to determine which primer combinations were most effectively combined in a fluorescent-based multiplexed PCR system. Multiplex PCR amplifications were carried out in a total volume of 10 µl containing 12 ng of template DNA, 2 m*M* $MgCl_2$, 0.1 m*M* of each dNTP (dATP, dGTP, dTTP, dCTP), 0.5 units of *Taq* polymerase (Boehringer Mannheim), 10 m*M* TRIS-HCl (pH 8.3), and 50 m*M* KCl. A range of primer concentrations and combinations were tested. Samples in 96-well polycarbonate plates were overlaid with paraffin oil and PCR were carried out on Techne PHC-3 thermal cyclers with 1 cycle at 93°C for 3 minutes, and 35 cycles of 30 seconds at 94°C, 30 seconds at 58°C, and 30 seconds at 72°C. PCR products from multiplex reactions were then pooled and diluted proportionally to give an absorption signal in the range of 500–2000 fluorescent units. A 0.75 µl aliquot of the pooled sample was combined with 0.5 µl size standard (GeneScan-500 [TAMRA] internal lane size standard (PE Applied Biosystems)), 2 µl deionised formamide, 0.5 µl dextran blue dye loading buffer (3.75 µl in well), heated to 90°C for 2 minutes and loaded on to a 0.4 mm, 36 well, 6% acrylamide, 24 cm well-to-read gel, made with 7 *M* Urea, 1X TBE. Samples were then electrophoresed at 2500V/23 watts for 3.5 or 5 hours to allow the 160bp or 200bp size standard fragment to be detected. Control samples with known allele sizes were also run to facilitate scoring (*see below*).

Data Analysis and Haplotype Assignment

Fluorescently labelled DNA fragments were analysed using GENESCAN™ 672 Software Package Version 2.0.2 (PE Applied Biosystems), utilising the Local Southern method of size calling, in conjunction with GENOTYPER Version 1.1.1. With these packages we sized and quantified DNA fragments by automated fluorescent detection and transformed these data into results files for further analysis. The control samples were used to standardise the allele

assignment at each locus and to correct for possible gel-to-gel variation. The alleles at each locus were assigned "allele numbers" (coded as 1, 2, 3, 4, smallest to largest fragment). Allele numbers for each of the seven informative loci were combined to create a seven-locus cpSSR haplotype. Data obtained from the 51 trees were used to calculate the frequency of each cpSSR haplotype in this population.

RESULTS

Polymorphism Testing

We tested 20 cpSSR previously identified from the chloroplast genome of *P. thunbergii* on 51 *P. radiata* trees to identify polymorphic markers. All primer pairs amplified fragments in *P. radiata,* and seven cpSSR loci were polymorphic. The 14 monomorphic loci included one locus (cpSSR 9) which amplified two fragments for every individual (possibly due to multiple primer binding sites). We found two to four alleles for each of the seven polymorphic cpSSR loci in this 51-tree population. The 1 bp differences observed for each cpSSR were assigned allele numbers, and combined to form a chloroplast haplotype for each tree (Table 2). Altogether 16 haplotypes were detected in the 51 trees tested. Nine of these haplotypes were unique to individual trees, and the remaining seven haplotypes were shared by 3–11 trees.

TABLE 2–Chloroplast SSR haplotype frequencies in the 51 trees tested. Within haplotypes the loci are listed in the following order: cpSSR 20, cpSSR 12, cpSSR 5, cpSSR 13, cpSSR 4, cpSSR 3, and cpSSR 1.

Chloroplast haplotype	No. of trees	Haplotype frequency
1112222	1	0.02
1122122	1	0.02
1122222	1	0.02
2112222	1	0.02
2122122	1	0.02
2122212	1	0.02
2122223	1	0.02
3122322	1	0.02
4222223	1	0.02
2122322	3	0.06
3122222	4	0.08
2121212	5	0.10
2122221	5	0.10
2112232	6	0.12
2122123	8	0.15
2122222	11	0.21

Multiplex Optimisation

A range of primer combinations were trialed to determine which primer sets were suitable to combine in multiplexed PCR reactions (data not shown). Comparable signal intensities for all loci in the multiplex system were obtained by adjusting individual primer concentrations, by diluting the labelled primer with unlabelled primer, and by altering the number of PCR cycles. The optimised protocol consisted of two PCRs, one containing five primer pairs and the other two primer pairs:

Reaction A: 0.1 μM cpSSR 20, 0.15 μM cpSSR 12, 1.8 μM cpSSR 5, 0.2 μM cpSSR 13, and 0.1 μM cpSSR 4

Reaction B: 0.4 μM cpSSR 3 and 0.5 μM cpSSR 1

All forward primers in Reaction A were labelled with 6-FAM, and within Reaction B cpSSR 3 was labelled with TET and cpSSR 1 with 5'HEX. Two labelled primers required diluting with unlabelled primer to achieve appropriate signal intensity—cpSSR 20 (3:5 ratio of labelled primer : unlabelled primer) and cpSSR 3 (1:2 ratio of labelled primer : unlabelled primer). After PCR, 1.5 µl of Reaction A and 2 µl of Reaction B were pooled and a 0.75 µl aliquot of the pooled sample was then combined with an internal lane size standard, deionised formamide, and dextran blue dye loading buffer as described above. The seven loci were then electrophoresed in a single lane (Fig. 1). We found the use of 24 cm well-to-read gels gave accurate and reproducible results with within-gel variations of approximately ± 0.2 bp (data not shown).

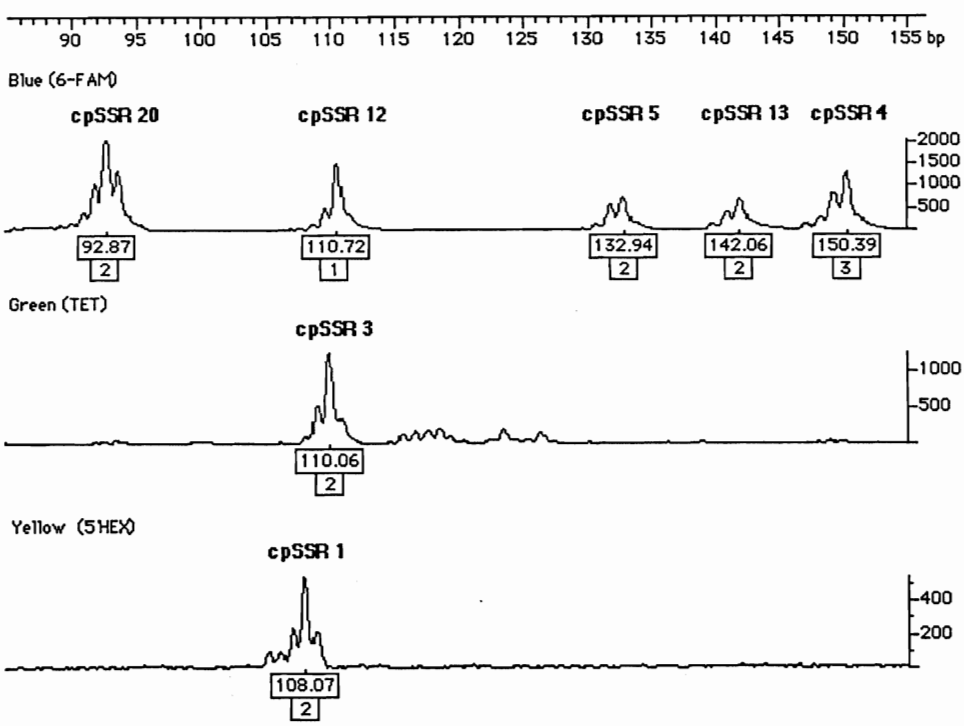

FIG. 1–ABI GENOTYPER electropherograms of three dyes in a single lane showing a complete seven-locus cpSSR haplotype for one tree. Fragment size (in base pairs) and allele number shown under each peak. The haplotype for this individual is 2122322.

DISCUSSION

Polymorphism Testing of cpSSR Loci in *Pinus radiata*

Seven of the 20 cpSSR loci tested revealed fragment-length polymorphism among 51 *P. radiata* trees. Three of these loci (cpSSR 1, 5, 12) were previously reported to be

polymorphic in *P. radiata* (Cato & Richardson 1996) and we found no new alleles at these loci in the 28 additional trees tested here. We identified four new polymorphic cpSSR loci in *P. radiata* (cpSSR 3, 4, 13, 20), yielding a total of 19 cpSSR alleles in the tested trees. The combination of alleles for the seven loci yielded 16 haplotypes in these 51 trees; nine were unique to individual trees and the most common haplotype was found in 21% of the trees.

The development of a fluorescently labelled, multiplexed protocol for these seven cpSSR loci greatly reduced the cost and time required to carry out chloroplast genotyping compared to previously reported methods (Cato & Richardson 1996; Powell *et al.* 1995; Echt *et al.* in press). The ability to resolve the PCR amplicons in a single lane on ABI DNA Sequencers (e.g., Fig. 1) further reduced costs and accelerated the collection of haplotype data. Furthermore, this protocol can be used to assay any source of cpDNA. To date over 1000 *P. radiata* seven-locus cpSSR haplotypes have been generated using this protocol from embryo, needle, or bud tissue with a first-run success rate (all seven loci clearly assignable) of greater than 90%. Full cpSSR haplotypes have also been generated by direct PCR from dried pollen grains (Kent unpubl. data).

In practice, we have found that the genetic discrimination provided by the five cpSSR loci in Reaction A (Fig. 2) is sufficient for many applied genetic testing applications in *P. radiata*. Because a single PCR is used to amplify these five loci, standard ABI four-dye technology permits three genotypes per electrophoretic lane (one dye is reserved for the within-lane size standard). Using the ABI 377 systems with 64 lane, 36 cm well-to-read gels, and assuming

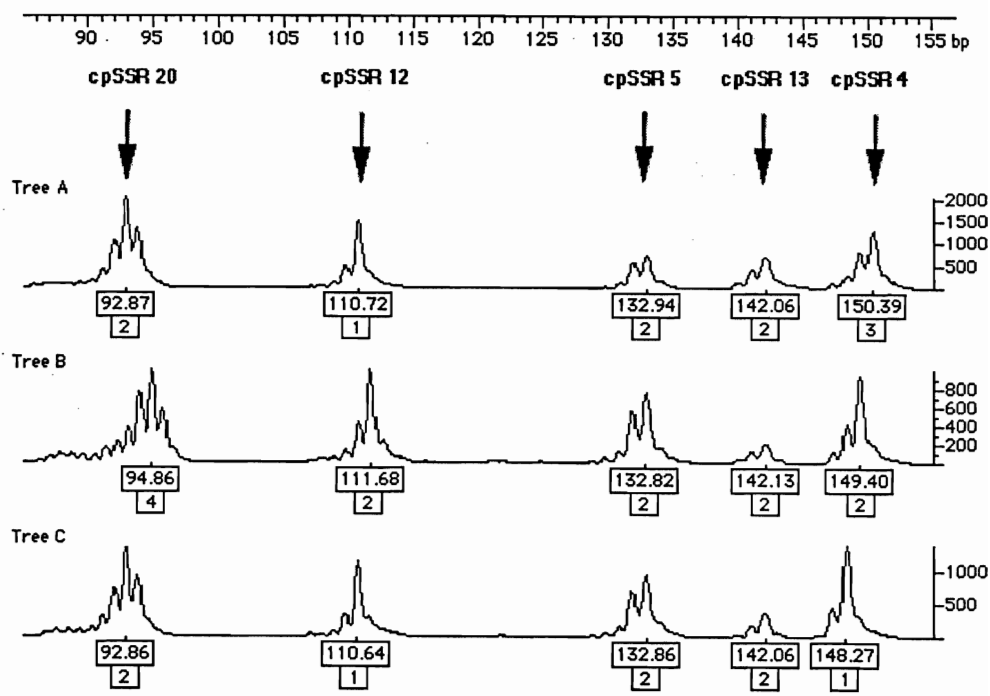

FIG. 2–ABI GENOTYPER electropherograms of Reaction A (five loci) showing different cpSSR haplotypes of three trees. Fragment size (in base pairs) and allele number shown under each peak. The assigned haplotypes for these trees are: Tree A, 21223; Tree B, 42222; Tree C, 21221.

two runs per day, approximately 375 five-locus cpSSR haplotypes can be analysed by one technician per day, compared with approximately 125 haplotypes per day if all seven loci are scored.

This multiplexed cpSSR system has streamlined the collection of cpSSR haplotypes for various genotyping applications including DNA profiling of top clones, paternity testing of embryos from control-pollinated and experimental liquid-pollinated seed, and investigating of chloroplast diversity in natural and selected populations of *P. radiata*. The observations that cpSSR are informative in other species and can be amplified with primers designed from heterologous sources (Powell *et al.* 1995; Cato & Richardson 1996; Vendramin *et al.* 1996; Echt *et al.* in press) suggest that the assay system reported here may be equally useful for chloroplast haplotyping in other species.

ACKNOWLEDGMENTS

Funding for this project was provided by the New Zealand Public Good Science Fund FRST(CO4616) and an NSOF allocation from New Zealand Forest Research Institute Limited. We thank L.Williams (University of Auckland) for his advice on fluorescent electrophoresis, and for running the gels and data analysis on the ABI software, W.Powell (Scottish Crop Research Institute) for providing PCR primers and sequences, and P.Fisher, G.Corbett, J.Lee, S.Cato, and A.Warr (New Zealand Forest Research Institute) for lab or field assistance. R.Gardner and P.Wilcox provided helpful comments on a previous version of this manuscript.

REFERENCES

CATO, S.A.; RICHARDSON, T.E. 1996: Inter- and intraspecific polymorphism at chloroplast SSR loci and the inheritance of plastids in *Pinus radiata* D. Don. *Theoretical and Applied Genetics 93*: 587–592.

ECHT, C.S.; DEVERNO, L.L.; ANZIDEI, M.; VENDRAMIN, G.G.: Chloroplast microsatellites reveal population genetic diversity in red pine, *Pinus resinosa* Ait. *Molecular Ecology* (in press)

KUANG, H.; RICHARDSON, T.E.; CARSON, S.D.; BONGARTEN, B.C.: An allele responsible for seedling death in *Pinus radiata* D.Don. *Theoretical and Applied Genetics* (in press)

POWELL, W.; MACHRAY, G.C.; PROVAN, J. 1996: Polymorphism revealed by simple sequence repeats. *Trends in Plant Science 1(7)*: 215–222.

POWELL, W.; MORGANTE, M.; McDEVITT, R.; VENDRAMIN, G.G.; RAFALSKI, J.A. 1995: Polymorphic simple sequence repeat regions in chloroplast genomes: Applications to the population genetics of pines. *Proceedings of the National Academy of Science USA 92*: 7759–7763.

RICHARDSON, T.; CATO, S.; CORBETT, G.; FISHER, P.; KENT, J.; WARR, A.; YOON, M. 1997: Application of DNA markers in *Pinus radiata.* P.333 *in* Burdon, R.D.; Moore, J.M. (Ed.) "IUFRO '97 Genetics of Radiata Pine", Proceedings of the NZ FRI - IUFRO Conference 1–4 December and Workshop 5 December, Rotorua, New Zealand. *FRI Bulletin No.203.*

RICHARDSON, T.; BURDON, R.; CARSON, S.; KENT, J.; KUMAR, S.; LEE, J.; MACKENZIE, R.; MacLEOD, M. 1997: Genetic variation in natural and selected populations of *Pinus radiata.* P.247 *in* Burdon, R.D.; Moore, J.M. (Ed.) "IUFRO '97 Genetics of Radiata Pine", Proceedings of the NZ FRI - IUFRO Conference 1–4 December and Workshop 5 December, Rotorua, New Zealand. *FRI Bulletin No.203.*

VENDRAMIN, G.G.; LELLI, L.; ROSSI, P.; MORGANTE, M. 1996: A set of primers for the amplification of 20 chloroplast microsatellites in *Pinaceae. Molecular Ecology 5*: 595–598.

VINCENT, T.G. 1997: Radiata pine tree breeding. *New Zealand Tree Grower 18*: 21–24.

WAKASUGI, T.; TSUDZUKI, J.; ITO, S.; SHIBATA, M.; SUGIURA, M. 1994: A physical map and clone of the black pine (*Pinus thunbergii*) chloroplast genome. *Plant Molecular Biology Reporter 12(3)*: 227–241.