Theor Appl Genet (1999) 98: 1091–1098

M. Rossetto · A. McLauchlan · F. C. L. Harriss
R. J. Henry · P. R. Baverstock · L. S. Lee
T. L. Maguire · K. J. Edwards

Abundance and polymorphism of microsatellite markers in the tea tree (*Melaleuca alternifolia*, Myrtaceae)

Received: 28 July 1998 / Accepted: 8 October 1998

Abstract The sequencing of 831 clones from an enriched microsatellite library of *Melaleuca alternifolia* (Myrtaceae) yielded 715 inserts containing repeat motifs. The majority of these (98%) were dinucleotide repeats or trinucleotide repeats averaging 22 and 8 repeat motifs respectively. The AG/GA motif was the most common, accounting for 43% of all microsatellites. From a total of 139 primer pairs designed, 102 produced markers within the expected size range. The majority of these (93) were polymorphic. Primer pairs were tested on five selected *M. alternifolia* genotypes. Loci based on dinucleotide repeats detected on average a greater number of alleles (4.2) than those based on trinucleotide repeats (2.9). The loci described will provide a large pool of polymorphisms useful for population studies, genetic mapping, and possibly application in other Myrtaceae.

Key words Simple sequence repeats · Enriched microsatellite library · Molecular marker · Myrtaceae · *Melaleuca*

Communicated by P. Langridge

M. Rossetto (✉) · A. McLauchlan · F. C. L. Harriss · R. J. Henry · L. S. Lee
Centre for Plant Conservation Genetics, Southern Cross University, PO Box 157 Lismore, NSW 2480, Australia
Fax: +61-2-6622 2080
E-mail: mrossett@scu.edu.au

P. R. Baverstock
Centre for Animal Conservation Genetics, Southern Cross University, PO Box 157 Lismore, NSW 2480, Australia

T. L. Maguire
Centre for Coastal Management, Southern Cross University, PO Box 157 Lismore, NSW 2480, Australia

K. J. Edwards
IACR-Long Ashton Research Station, Department of Agricultural Sciences, University of Bristol, Long Ashton, Bristol, BS18 9AF, UK

Introduction

Microsatellites, or simple sequence repeats (SSRs), were initially described in humans (Litt and Luty 1989; Weber and May 1989). Similar findings quickly followed in other mammalian species such as mice (Love et al. 1990), pigs (Johansson et al. 1992) and cattle (Kemp et al. 1993). Their potential as useful markers for plants was promptly recognised, resulting in their successful isolation and application within many species. The majority of early studies involved crop species such as rice (Wu and Tanksley 1993), barley (Saghai Maroof et al. 1994) and wheat (Roder et al. 1995). More recently, there has been an increase in the characterisation of SSRs in woody species such as bur oak (Dow et al. 1995), eucalypts (Byrne et al. 1996) and *Swietenia humilis* (White and Powell 1997).

The principal reason for the increasing success of SSRs as a molecular tool is that they provide a higher incidence of detectable polymorphisms than other techniques such as RFLPs and RAPDs (Powell et al. 1996b). Furthermore, these markers have an abundant and uniform distribution throughout the genome, are co-dominant and segregate in a Mendelian fashion. Their screening relies on simple PCR technology and requires only small amounts of DNA. The reproducibility of SSRs is such that they can be efficiently used by different laboratories to produce consistent data.

The application of SSRs provides a successful approach to marker-assisted selection in plant breeding practice (Gupta et al. 1996; Powell et al. 1996a). The ability of these hypervariable regions to reveal high allelic diversity is particularly useful in distinguishing between close crosses. As a result, they are now considered as the marker of choice for inbreeding crops with little intraspecific polymorphism (Roder et al. 1995).

The wider use of molecular biology techniques in breeding projects is usually a reflection of their efficiency. Until recently developing SSR markers for new

species was a laborious and costly exercise, thus limiting their potential applications. Characterisation relied on searches through published sequences, only possible for a minority of species, or on the production of highly involved genomic DNA libraries. In a comparison between methods used to produce SSR markers in sorghum, Brown et al. (1996) concluded that despite an overall low yield, the screening of genomic-DNA libraries was more efficient than database searching or testing SSRs designed for other species. Recently, the development SSR enrichment techniques, in which selective genomic libraries containing pre-screened inserts are prepared, has increased the efficiency of SSR characterisation in new species (Edwards et al. 1996; and see Powell et al. 1996 a for other examples). The availability of such technology opens new opportunities for large-scale SSR characterisation in species with no previous sequence knowledge.

Melaleuca alternifolia (Maiden and Betche) Cheel, commonly known as the tea tree, is restricted to North Eastern New South Wales and South Eastern Queensland within Australia (Wrigley and Fagg 1993). The species is the principal source for tea tree oil, highly valued in the pharmaceutical and cosmetic industries for its broad-spectrum germicidal properties. Traditionally tea tree oil production relied on leaf material harvested from natural stands but the recent expansion of the industry has resulted in a shift towards plantations. Most of the existing plantations have been established from a limited genetic pool based on anecdotal knowledge. Previous research points to the presence of great genotypic and chemotypic variation throughout the species (Butcher et al. 1992, 1994; Rossetto et al., in preparation, Homer et al., in preparation) as well as the high heritability of oil characters (Butcher et al. 1996). An efficient tree breeding strategy is therefore likely to increase both the quality and the quantity of yield. The development of useful SSR markers linked to specific traits will facilitate the screening of progeny at early growth stages, thus accelerating selective breeding programs.

The present study uses enrichment techniques to efficiently produce a larger number of SSR markers for this economically important species. These markers will be employed in a variety of applications such as population studies and integration in a *M. alternifolia* genetic map. Such a comprehensive study provides an opportunity to investigate the type and distribution of repeat motifs in the tea tree and to compare these results with those of other species.

Materials and methods

Construction of an enriched microsatellite library and the sequencing of clones

The microsatellite enrichment procedure was based on that described by Edwards et al. (1996). The following synthetic oligonucleotides were used in the procedure: $(CT)_{15}$, $(CA)_{20}$, $(ACT)_{14}$, $(AGA)_{14}$, $(CAA)_{14}$, $(CTA)_{14}$, $(CTT)_{14}$, $(CTG)_{10}$, $(CAG)_{10}$, $(GAC)_{14}$, $(AGC)_{14}$, $(CAT)_{14}$, and $(ACA)_{14}$.

Transformed cells were plated onto LB-agar plates containing 100 μg/ml of ampicillin and 50 μg/ml of X-galactosidase. White colonies were grown overnight in 5 ml of LB broth with 100 μg/ml of ampicillin. Plasmid DNA was extracted using the High Pure Plasmid Kit (Boehringer Mannheim). Sequencing was performed by the dideoxynucleotide chain-termination method using an M13 forward 24-mer. DNA sequencing was carried out on an ABI 377 sequencer at commercial sequencing centers.

Primer design, DNA extraction and PCR analysis

Primers were designed from within the regions flanking the repeat motifs for trinucleotide repeats greater than seven and dinucleotide repeats greater than 12 using dedicated software (MacVector 6.0). Primer selection criteria were based on GC content, the melting-temperature curve and the lack of secondary structure.

Primers were screened against five individuals including the original genotype from which the library was constructed (as a positive size control), two parents of a selected cross, and two individuals known to be genetically distinct from a pilot population genetic study (Rossetto et al., in preparation). DNA was extracted using a modified CTAB method (Maguire et al. 1994).

PCR reactions were performed in a 12.5-μl mixture containing 10 mM Tris-HCl pH 8.3 and 50 mM KCl, 2.5 mM $MgCl_2$, 0.5 units of *Taq* DNA polymerase (Boehringer Mannheim), 0.2 mM of each dNTP, 2.0 μM of each forward and reverse primer, 25 ng of template DNA and DNA-free water. PCR reactions were run on a Perkin Elmer 9600 or 9700 thermocycler. The cycling conditions were: 94°C for 3 min, followed by 30 cycles of 94°C 30 s, optimal annealing temperature (ranging from 48°C to 56°C, see Table 4) 30 s, 72°C 1 min, and a final extension step at 72°C 5-min prior to cooling at 4°C.

Primer evaluation

Amplification products were resolved using three electrophoretic methods. Polyacrylamide gels (10%) run at 250 V for 3.5 h in 0.5 TBE and visualised by silver staining (Promega Silver Staining Kit). Agarose gels (3%) run at 80 V for 3.5 h in 1 × TBE and stained with ethidium bromide. Gels were photographed and the size of the amplified fragments was determined by the distance travelled on the gel as compared to a 50-bp ladder. Of these primers 40 were also labelled with ABI fluorescent dyes (TET, HEX or FAM) and visualised following separation by capillary electrophoresis on an ABI 310 sequencer.

Results

Enrichment

Of the 831 clones sequenced a total of 826 (99.4%) contained inserts, of which 715 (86.0%) included microsatellite repeat motifs (Table 1). Of these repeats 541 (75.7%) were unique (i.e. not previously sequenced). Redundant sequences were mainly trinucleotide repeats (91.2%) the majority of which contained combinations of C-G repeat motifs. The rate of redundant sequences was not linear, the first half of the inserts sequenced contained only 2.9% redundant trinucleotide repeats and 0.6% redundant dinucleotide repeats,

whereas the second half contained 13.6% and 1.0% respectively. Insert size varied between 79 and 491 bp, with the average being 216 bp.

Microsatellite description

The majority (98.3%) of repeat motifs were dinucleotide or trinucleotide repeats (Table 1). On average, dinucleotide repeats had a much higher number of repeats (21) than trinucleotide repeats (8), the maximum recorded being 72 and 55 respectively (Fig. 1). The remaining six SSRs were two tetramers, one pentamer, two hexamers and one heptamer, with the number of repeat motifs ranging from three to five.

Table 2 shows that the most common repeat motif was the AG/GA group which represented over 43% of all microsatellites detected in *M. alternifolia*. This was followed in frequency by the GGC/GCG/GCC group which represented over 37% of the detected microsatellites.

Table 1 Microsatellite enrichment success for *M. alternifolia* DNA libraries. The total number of positive colonies successfully sequenced and the frequency of redundant and unique sequences (divided into different categories) is shown

Total sequencing reaction	831
Lacking inserts	0.6%
Lacking microsatellites	14.0%
Redundant sequences:	
Dimers	1.6%
Trimers	16.5%
Type of repeat:	
Dimer	52.3%
Trimer	46.0%
Other:	1.7%
Perfect	76.7%
Imperfect	23.3%
Compound[a]	20.6%

[a] Compound microsatellites include perfect and imperfect sequences

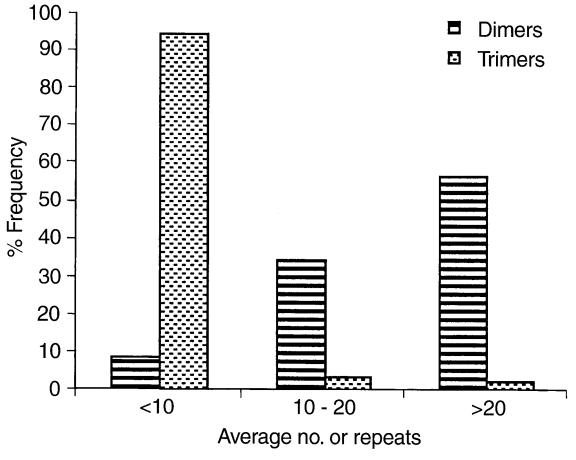

Fig. 1 Frequency of different-sized dinucleotide repeats and trinucleotides repeat motifs in *M. alternifolia*

Table 2 Frequency of the most common microsatellite repeat motifs in *M. alternifolia*

Type of repeat	Frequency
GA/CT	22.6%
AG/TC	20.7%
GGC/CCG	19.1%
CGG/GCC	16.6%
AC/TG	5.1%
CA/GT	3.8%
AGG/TCC	1.6%
GCG/CGC	1.4%
GCT/CGA	1.1%
CTG/GAC	1.1%
Others[a]	6.9%

[a] Includes all repeat motifs with frequency lower than 1.0%

Table 3 Size distribution of inserts sequenced from a microsatellite-enriched *M. alternifolia* DNA library, and the frequency of sequences lacking an adequate flanking region or a microsatellite repeat

Average insert size		216 bp
Inserts smaller than 181 bp:	33.4%	
With an insufficient flanking sequence		19.5%
Without an SSR		23.6%
Inserts between 181 bp and 220 bp:	26.4%	
With an insufficient flanking sequence		18.8%
Without an SSR		14.9%
Inserts between 221 bp and 300 bp:	28.6%	
With an insufficient flanking sequence		24.6%
Without an SSR		15.6%
Inserts greater than 301 bp:	11.6%	
With an insufficient flanking sequence		8.8%
Without an SSR		17.6%

Primer design and polymorphism

Overall, 363 (64.8%) sequences could be considered useful as they contained either dinucleotide repeats with more than 12 repeat motifs, or trinucleotide repeats with more than eight repeat motifs. However, of these sequences 162 (28.9%) did not have sufficient flanking sequences to permit the design of a PCR primer pair. Table 3 shows that there was no correlation between the insert size and the lack of flanking sequence, except for inserts larger than 300 bp.

Of the 201 sequences considered suitable for primer design 139 primer pairs were designed. Of these 102 (73.4%) produced a fragment of the expected size, with 93 (66.9%) showing polymorphism (Table 4). Of the unique sequences 17.0% produced polymorphic primer pairs. However, the overall success figure (i.e. the number of polymorphic microsatellites out of the 831

Table 4 List of the primers developed for *M. alternifolia* using enrichment methods. The repeat motif for each locus and the 5′ to 3′ sequences of the forward and reverse primers are indicated (imperfect repeats are italicised) together with the annealing temperature used (T), the expected size of the original sample (in base pairs), the presence of polymorphism (Y/N represents the successful/unsuccessful detection of polymorphism) and the number of alleles detected in the sample tested

Code	SSR sequence	5′ to 3′	Primer sequences	T	Size	P	A
scu124TT	(GA)32	Fwd	ACA TGC ACA TTG GCG TTA TTG				
		Rvs	CCA GGA CTT GCT TTT ATT CTT CAT	50	135	Y	5
scu125TT	(CT)27	Fwd	CGT CTA CGT CAA TGT CCA CAA C				
		Rvs	CAC TTT CAT TTG AAC GGG TGT TAC A	53	137	Y	5
scu127TT	*(GCC)6* (TC)14	Fwd	CGC CTC TGT TTC TCC CTT T				
		Rvs	GAT CGG AGC CTT CTC GAA T	53	165	Y	5
scu128TT	(AG)26	Fwd	CAA CAA TGC ACA TGG AAG TTC				
		Rvs	GTG ATG GCG GCT GAG TTA G	50	191	Y	5
scu129TT	(GA)21	Fwd	CGT TCG GAT ACG GAG AAG GAA TTG A				
		Rvs	CGT CTC AGG AAG GGC GTT GTT TA	53	201	Y	2
scu130TT	(TG)25	Fwd	CGA GCA ACC CCG TTC GTT T				
		Rvs	CTA AGA ACA CCG ACA ACC CAA GA	53	195	Y	5
scu131TT	(AG)24 (GT)9	Fwd	TCC ATC CTT GAC CAC AAG TT				
		Rvs	GCT AGT CCC ATT TGT TGG AAT	50	151	Y	6
scu132TT	*(GA)25*	Fwd	GAA AAA GCC ATT GCG TGA GGA A				
		Rvs	GCC AAC CTA GCA TTC TCA TGG AAT	53	156	Y	5
scu133TT	(GA)32	Fwd	TGA CGA TGG TGG CAA GGA A				
		Rvs	AAG TAA GTG GAT GTC CTC CCA T	53	203	Y	5
scu134TT	(AG)19	Fwd	ATT TGT GCG TCT TTC TGA TTG TGC T				
		Rvs	GGA ACG GAC CAA GGT GGC AA	53	164	Y	4
scu135TT	(TC)21	Fwd	TGG GTT CGT CGT CGA TTT				
		Rvs	CCT TTC TCT TGA CGG GCA AT	53	140	Y	4
scu136TT	(AC)15	Fwd	GCT TCT AGG ATT ATG GAT TTC TCA G				
		Rvs	GTC ACT TTC ACT TTC AGT TGG AAT	48	169	Y	5
scu137TT	(TC)21	Fwd	CAA GTG ACT CTG ACG ACG TTG G				
		Rvs	CAG TAG GAG AGG GGA TCA ATT TG	53	205	Y	5
scu140TT	(AG)21	Fwd	CAGT TCT CAC AAC TAT TCT ATC TCT T				
		Rvs	GCC TGT GTT TGT CAT CTC ATT	48	166	Y	4
scu141TT	(GT)15	Fwd	GAA GCA CAA TTA CGC CGA TCA A				
		Rvs	TGG TGC AGA TGA TAG ACC CAG AA	50	126	Y	6
scu142TT	(GT)24	Fwd	CCG AGC AGA ATC TGG TCC TCA T				
		Rvs	CGA TGT CGT GGC AGC ATT TCT	53	139	Y	5
scu144TT	*(GA)24*	Fwd	TTC GTG CTC AGT GCG GGC TT				
		Rvs	GCA GGT CGT CGT GTT GTG TTC	56	175	Y	6
scu145TT	*(CT)10 (GCC)9*	Fwd	CCT CCT CCG TTA ATC GCT GC				
		Rvs	ACC CGT CCG CAG AAT CGC AA	56	202	Y	5
scu026TT	*(AG)37*	Fwd	TGC ATC CAC TGT TTA ATC TTT				
		Rvs	CCA CGC CGA GTA ATG TTT C	51	213	N	1
scu029TT	*(CCG)9*	Fwd	CGC CCG TTC ATC GTA AAT				
		Rvs	TCG AAC CAA CCA CTC AGA AA	53	108	N	1
scu030TT	*(CCG)8*	Fwd	TTC TCC CCT TCC AGA AAC AA				
		Rvs	GCC AAG CCA TAC ATT ATC CTA A	53	127	N	1
scu068TT	(TG)25 (AG)13	Fwd	TGA TGC TAA CGA GTG ACC TTG AT				
		Rvs	GCT AAG CAC CAT GCA GAT TCA C	53	158	N	1
scu071TT	*(GGT)11*	Fwd	CGT CCC TTC AAA CCG TCC TCA A				
		Rvs	AGC ATC CTC AGC CAC CAC ACA A	56	359	N	1
scu077TT	(CA)14	Fwd	AAG GCT GAG AAG GGG ATC AGA				
		Rvs	CAC AGC ACT GGA ACA CAC ATG A	50	136	N	1
scu080TT	(GA)27	Fwd	GGG ATT GTT CAA CTT GCC AGC CAT A				
		Rvs	TCA CAC ACA AGG TAG GGA AGA TG	50	150	N	1
scu101TT	(AG)23	Fwd	GGG TGG TCT CTA AAG GCA TCT GA				
		Rvs	TTC TGC ACA CGT AGA ACC TCG ACT	53	143	N	2
scu143TT	(CCG)8 (CT)21	Fwd	GAA GTA ATG ACC AAC GTA GGG AAT				
		Rvs	GAA AAT GCT ACT CGT TTG CAT C	56	381	N	1
scu070TT	*(AG)16*	Fwd	CGA CGC TGC AAT AAT CTC GTT				
		Rvs	TGG GTT CTC CCA TAG AAG TCA TT	50	169	Y	5
scu072TT	(AC)16	Fwd	TGA GCC TAA GCC CAA ACT ACT CT				
		Rvs	AGG TCA TGT AGG TCG CTA T	50	123	Y	4
scu073TT	*(GCA)9*	Fwd	CCT CCC ATC CGC TGT AAA A				
		Rvs	AAC CGC CAC GAA GCC ATG AT	53	94	Y	4
scu074TT	(GA)28	Fwd	CAA GCC ACG CCT GCA CAA AA				
		Rvs	TGA CCA AAC TCT GGG GAG GAG AA	53	229	Y	3

Table 4 Continued

Code	SSR sequence	5′ to 3′	Primer sequences	T	Size	P	A
scu075TT	(GT)18	Fwd	CTG CTA CCC TAA TGT TGT AGG AT				
		Rvs	GAA GAA ATG TTG AGA TTA TGC CG	50	163	Y	2
scu078TT	(TGC)9	Fwd	CCG CCA CGA AGC CAT GAT				
		Rvs	CGT TCC TCC CAT CCG CTG TA	53	93	Y	4
scu079TT	(GA)30	Fwd	GGG GTC TCC TCA AAA TGC TC				
		Rvs	TTG CAC GAA AGA TGA TGC C	53	202	Y	5
scu081TT	(TC)21	Fwd	GGC ACC GAT TGA AGC AAG AA				
		Rvs	GGC GTA GAC ACT GGT TTT GGT	53	115	Y	2
scu082TT	(CT)23 (CA)15	Fwd	TGG TGA TCC ATG TGG GTT TGA				
		Rvs	TTG ACG CTG AGC AGC AGG TAA	53	166	Y	2
scu083TT	(CT)38	Fwd	TTC TCC GAT TCT GAG TGC TCG TCT				
		Rvs	GAA GGT GCC CAA CAT GCT CAA A	53	207	Y	5
scu084TT	*(TC)20*	Fwd	CAG TTG CAG GAG GTA ACC CTA CTT				
		Rvs	CGC CCA TTT ATG TCA CAC GAA T	53	191	Y	6
scu085TT	*(GA)24*	Fwd	GGG TTG GTC TCT GTT TCT GTT A				
		Rvs	CGA GAG TAT TTA GCC TTT CTT CCT T	50	211	Y	6
scu087TT	(CA)18	Fwd	CAT GAG CCT AAG CCC AAA CTA CT				
		Rvs	GCG TGG ACT ACT ATG TGG TCA ATT C	50	172	Y	4
scu089TT	(CT)19	Fwd	CCA GTC AAT GATAGA CCT CAT C				
		Rvs	CAA AAG CCC AAG AGA TCA GGA AA	50	154	Y	4
scu092TT	(GA)23	Fwd	GCG TCA GAA ACA GCC GAG ATT T				
		Rvs	TGC TGC AAA AAA GGA CCG ATA G	56	222	Y	5
scu093TT	(AG)23	Fwd	AAC CGA GGG TGT CCA TCT GAT				
		Rvs	GCC TCC TCC TCT CTT CAT ACT CAA	53	132	Y	6
scu094TT	(CA)14	Fwd	TCC AAA CAG ATC CGC CCA ACT				
		Rvs	TCT TCT CCC TTG CCC CGT TTT	56	181	Y	5
scu096TT	(CT)6 (CT)18	Fwd	GCC ACA TCA GCA TAA CAC AGA ACA				
		Rvs	TGA GCG TCA GTG AAG GTC AAC AT	53	143	Y	6
scu097TT	*(GA)22*	Fwd	GGA TCA GCA AGG GTA AAG CAA A				
		Rvs	CTC GCA GAA GAA GGA AAA CGA T	53	144	Y	5
scu098TT	*(GCT)9*	Fwd	TTG ACG CAT AAG ATG CAA				
		Rvs	TGA GGT GGC TTC ATA TTT GT	50	175	Y	4
scu099TT	(GGC)7 (GAGGAC)3	Fwd	CAA TCC GCT CTT CTT CGT CCA TCT				
		Rvs	CCC ATC CTC ATC CTC CTT CCA CTT	56	138	Y	2
scu104TT	(GA)24	Fwd	GCA ACC ATA GGT GCG ATA AA				
		Rvs	TCA CTT GGC ACG AGC GAA T	50	237	Y	4
scu105TT	*(GA)16*	Fwd	CGA CTG AAG ATC GTC GTG ACA ATG A				
		Rvs	TCT CCT CCC TCA GGC ATT TCT TGA T	53	132	Y	4
scu106TT	(GT)18	Fwd	TCG TTG GCA CAG CAC GAT A				
		Rvs	CTG CAT ATA CAC ATA GAT GCG	48	155	Y	5
scu107TT	(GGC)5 (GA)16	Fwd	CCT CGA CAA AAC AGA GCG TGA				
		Rvs	CAA GGC AAA GAA TCA AAC CCC A	56	241	Y	3
scu109TT	(TC)20	Fwd	TCG AGT TGA GTT TGG TTT CTT C				
		Rvs	CCC ATG TTT CTT GCC CAA	53	182	Y	4
scu110TT	(CA)15	Fwd	GGA TTA TGG ATT TCT CAG AAT GTC				
		Rvs	GTC ACT TTC ACT TTC AGT TGG AAT	48	162	Y	7
scu111TT	(AG)18	Fwd	TGG ACT GCA TCC TCT GTT GC				
		Rvs	GGG GTT TAT GAC TTG TAC TGG TGA	48	258	Y	4
scu112TT	*(GA)22* (GT)13	Fwd	TCT TCC CCA CGG TTG AGC TAA				
		Rvs	ATT GTT TGA CCT CTG CAT GAT TCA	53	283	Y	5
scu113TT	(GA)24	Fwd	CGA GAC TGA CAA GGA GGG AAA GAA A				
		Rvs	CCT GGT GGG TTC ACA GGT CAT CTT	56	226	Y	4
scu115TT	*(TC)24*	Fwd	CCT TCA ACG CCA CCT GCT ACT T				
		Rvs	AAA TAA CTA AAA CCA CGG TCG ATT C	53	180	Y	3
scu116TT	(AG)27	Fwd	GTG ACA ATC CAG TCC AAA TG				
		Rvs	GGT CCA ATT CTC AGT GCT AAC TT	53	235	Y	4
scu118TT	(TG)16	Fwd	GAG TGT GGT TCA AAC AGG GAA GCG A				
		Rvs	CCA GCC GTT CTC GAA CAT TCT	53	256	Y	3
scu119TT	(GA)24	Fwd	CCA AGT TCC AAC AGT TCT GTC CGA A				
		Rvs	GCA GTG CTT GAG AAA TGT CCC GT	53	148	Y	3
scu121TT	*(GCA)9*	Fwd	CCA CTG ATT GTT CGT CAC AAG AAC				
		Rvs	CTC TAA GAT GTG GGG GAC AAA CTT	53	220	Y	2
scu122TT	(AG)23	Fwd	AAC GAG GAC AGG AAG CGT AA				
		Rvs	TCA AGG ACC TCC AGA AGG AT	53	138	Y	6

Table 4 Continued

Code	SSR sequence	5′ to 3′	Primer sequences	T	Size	P	A
scu123TT	(AG)25	Fwd	CCC CTA GCA TCT CGC CAC TGT TAT				
		Rvs	TCG ACG GTT GAC TCT GGT TGT AGT T	53	173	Y	4
scu001TT	(AG)28	Fwd	TAC GCC CAA AAA CCC TTC CCA T				
		Rvs	TTG AGC ACA TGA AAC TGC AAC G	54	228	Y	4
scu003TT	(AG)23	Fwd	CTT CCA CCA GAA ACA TTT TCA C				
		Rvs	GTC GCA AGG ACT TTC AAT CA	50	133	Y	5
scu004TT	(GGC)9	Fwd	CCT TGC TCA ACG AAA AGA CAA				
		Rvs	GAG CCC CAG ACA TCA TTT ACA A	53	156	Y	5
scu007TT	*(CT)23*	Fwd	AAA CCC TAG AAA GAA ACT CCC T				
		Rvs	GGT CTG GAA GCA CCC ATT T	50	129	Y	4
scu008TT	(AG)25	Fwd	GTT GAC TTT CAC TCA CCA A				
		Rvs	CGA ACC CAT AAC CAA AAC A	50	108	Y	4
scu009TT	*(TC)29*	Fwd	AAT GAA AGG GAA ATG CGG C				
		Rvs	TTG CTT GTT GCG TGT CTT GTG	53	241	Y	4
scu010TT	(CT)28	Fwd	GAA GGC ACT CAA GAG CAA A				
		Rvs	GGA CCA TTC GTC TCA TTT TC	53	218	Y	4
scu011TT	(AG)35	Fwd	TGC CTT GAA GCA CAA CTC				
		Rvs	ATG GTC CCC TGA GGA TGT CA	51	129	Y	4
scu013TT	(AG)23	Fwd	CCG AAT CAT CTA AAA GGA GTC T				
		Rvs	CCA AGT TGA CAC ACT TAC TTC AA	51	128	Y	5
scu014TT	(AG)24	Fwd	GCC TGC ATC AGA CAT GAT ATA GA				
		Rvs	CCA TGA GAC TCT TTT GCG ATT	51	132	Y	5
scu015TT	(AG)25	Fwd	GGG AAG ATA CAT TTT CCT TCT C				
		Rvs	ATA ACC TAG TTC GGG AGG T	50	119	Y	6
scu016TT	(AG)25	Fwd	AGA GAT CAA TCA ACC GTC TTG				
		Rvs	CCC ATC TTT TCC CTC ATT TAC	51	144	Y	5
scu019TT	(AG)23	Fwd	CCG AAC AAG GAC GAC CCA TGT T				
		Rvs	CCC CTT TGG TTC AAC ACC AGA G	56	282	Y	4
scu021TT	(GA)24	Fwd	CGC CTC CTT TAC AGC CAA AGA T				
		Rvs	TGA AGC CAA CTT GCT CGG TTT	54	212	Y	4
scu023TT	(GCC)8	Fwd	GCG AGC CCC AAA CAT CAT T				
		Rvs	GCT CAA CGA AAA GAC GAT GGA	54	151	Y	3
scu025TT	*(GCC)8*	Fwd	CAG CAG CCA CCA AGT ATT				
		Rvs	CCT GTA TGT GTT GTA GTT GTG A	53	88	Y	2
scu031TT	*(GCC)8*	Fwd	GCA GTC GAT GCT TTG CAG GTT				
		Rvs	GCA GAA TCG CAA GGG CGT ATT	56	137	Y	4
scu032TT	*(CGG)8*	Fwd	TCG AAC CAA CCA CTC AGA AAG A				
		Rvs	CGC CCG TTC ATC GTA GAT TT	54	110	Y	4
scu036TT	*(AG)24*	Fwd	CGA CTG AAG ATC GTC GTG ACA				
		Rvs	CGC CGC TTT TGA AGA TCC A	53	212	Y	4
scu037TT	(TGG)3 (CGG)7	Fwd	CCC AAC TGG AGT ACT GGT TCA A				
		Rvs	GTC GTT GTG CTC TAC CCG ATG A	56	163	Y	2
scu039TT	*(AG)20*	Fwd	CGG ATT GTC ATG GCA AGA GCA A				
		Rvs	CGA GCA CCA TAC TCA TTC GCA	53	141	Y	4
scu040TT	*(AAG)8 (CGG)7*	Fwd	GTG AAG GAA GGA CGA CGC TT				
		Rvs	CGC CCC GAT AGA AAT CGT T	53	148	Y	4
scu041TT	*(GA)21*	Fwd	GCT TGG CAA ATC GTT TCC CA				
		Rvs	TCG GCT TTC CCT CTT TCA CTG T	53	120	Y	5
scu042TT	*(CCG)9* (TC)4	Fwd	CGC CCG TTC ATC GTA AAT				
		Rvs	TCT GAT GAG ACA GAC ACT CG	53	159	Y	2
scu044TT	(AG)20	Fwd	TCT TGT TTA GTT GCT CAG GCT				
		Rvs	CCT GGA GGC ATT CCT TAA A	50	157	Y	3
scu045TT	(AG)24	Fwd	AGT TTT CGA TTG CCC ACG				
		Rvs	TCC CAT TTT GAG ATG GAC TT	50	188	Y	6
scu048TT	(CT)28	Fwd	TGA GGC AAT GCA CAG TCT CAG T				
		Rvs	TCT CAA CCC ACT AAA GCG GAG T	51	337	Y	3
scu049TT	(GCC)7	Fwd	ACA AGA ATG CCG ACG CCA GT				
		Rvs	TCG TCC ATC TCG CCT GTC AT	56	172	Y	6
scu051TT	(CAG)12	Fwd	CCC CAA AGA CAA AAA GGC AA				
		Rvs	CCG TGA ACG ATA AGA CGT GTT T	53	197	Y	5
scu052TT	*(CTG)32*	Fwd	CTG CCG CAT TAG GGA ATC ATT				
		Rvs	CAG AAC GCC AGT TGA TGA AAG C	56	335	Y	5
scu053TT	*(CTG)12*	Fwd	GTT CTA ATT GCC CGC ACC ATA C				
		Rvs	CCG ATG TCT CCA ATC AAC CAC A	56	215	Y	3

Table 4 Continued

Code	SSR sequence	5′ to 3′	Primer sequences	T	Size	P	A
scu056TT	(CT)21	Fwd	CAC ACA AGC GGG AGT TTT ATC A				
		Rvs	TAC TGT GGG TGG TGG AGT TGA G	53	160	Y	4
scu062TT	(AG)21	Fwd	TGG TTG GAT CGT TGG CAA				
		Rvs	CGC ACA CCC ATT TCT TCT CA	53	128	Y	4
scu063TT	(TC)20	Fwd	CGC ATT GGT CCC TAG CAG TT				
		Rvs	CGT AAC ATC AAC ACA AAC GAA AC	53	178	Y	4
scu064TT	(TC)24	Fwd	CGT GGG CGT GCT CTT TCA AA				
		Rvs	TGA ATG GCT TAG GGA AGG AGG T	56	212	Y	6
scu066TT	(CT)23 (GC)7	Fwd	TCC ACT TTT CCT CGT CTC TGC TA				
		Rvs	CCA GGA CTC GTT GAG ATA ATC CA	56	254	Y	4
scu067TT	*(GCC)8*	Fwd	GCG GGG TTT CCA TTT CAT GG				
		Rvs	AGA CGA ACC ACA AGG TCT GTC GT	56	124	Y	2
scu069TT	(TG)17 (AG)16	Fwd	TTC GCT CTG CAA CAA GCA				
		Rvs	GGA TGG AGG GGT TCT TCA A	50	116	Y	4

inserts sequenced) is lower (11.2%). Limitations in the resolution of agarose and polyacrylamide gels could underestimate the polymorphism.

Out of the 102 primers designed which produced fragments of the expected size, 77 (75.5%) were derived from dinucleotide repeats and 25 (24.5%) from trinucleotide repeats (Table 4). The maximum number of alleles detected from dinucleotide loci from the five tested individuals was seven and with an average of 4.2 with no substantial difference between perfect and imperfect repeats. The maximum number of alleles for trinucleotide repeats was five with an average of 2.9. Perfect trinucleotide repeats produced almost twice as many alleles as imperfect repeats.

The occasional occurrence on non-specific banding patterns did not affect the interpretation of the results as they fell outside the expected size range. Within the screening phase, agarose and polyacrylamide gels produced comparable results and therefore the expected higher resolution of polyacrylamide did not justify its greater labour intensity. In order to obtain a precise marker size, the use of capillary electrophoresis run with an internal size standard proved to be the optimal procedure.

Discussion

The overall SSR enrichment procedure was successful and efficient for *M. alternifolia* and a total of 102 primer pairs was obtained from the total of 541 sequences containing unique SSRs.

As previously reported for other plant species (Liu et al. 1996; Ma et al. 1996; Panaud et al. 1996) the $(AG)_n$ repeat motif was more common than $(AC)_n$ in the tea tree (Table 2). $(AT)_n$ is usually the most common repeat motif in plants (Powell et al. 1996 a) but being a palindromic sequence it is excluded from the enrichment process and therefore its frequency in *M. alternifolia* remains unknown. Previous studies, summarised in Gupta et al. (1996), reported that the most common trinucleotide repeat motifs in plants are $(AAG)_n$ and $(AAT)_n$, with some species-specific variation. In contrast, these repeat motifs were not represented in the tea tree whereas the $(GGC)_n$ group, interestingly not included as a probe during the enrichment phase, was the most abundant. In a large-scale characterisation study on *Lycopersicon* Smulders et al. (1997) found that $(AAG)_n$ and $(AAT)_n$ are particularly common within intron sequences. As a result, these repeat motifs might be more commonly detected during database searches.

The frequency of dinucleotide repeat motifs (52.3%) was only slightly higher than that of trinucleotide repeat motifs (46.0%), but the near absence of large repeat motifs was unexpected particularly since the enrichment procedure was expected to successfully select for larger repeat motifs (Table 1). It is difficult to decide if this result reflects SSR abundance in *M. alternifolia* or is a consequence of the enrichment procedure. Smulders et al. (1997) also found that dinucleotide repeats represented approximately half of the SSRs detected in *Lycopersicon*. However, the frequency of trinucleotide repeats was lower, with a considerable number of large repeat motifs being identified (Smulders et al. 1997).

A total of 102 primer pairs produced fragments of the expected size from the initial 831 clones sequenced. A similar success rate was reported in sorghum by Brown et al. (1996). Two major factors affected the success rate for the identification of useful loci in the tea tree: sequence redundancy and lack of a flanking sequence. The presence of redundant inserts emphasises the probability of diminished returns as a larger number of clones from the same library is sequenced (Table 1). In the tea tree, this appears to be particularly true for C-G trinucleotide repeat motifs and could be connected to their unexpected high frequency. Sequence redundancy could represent a potential inefficiency for large-scale projects. The number of non-redundant clones was still high but proximity to the

origin of the insert further reduced the number of potentially useful sequences. As inserts are size-selected between 100 bp and 50 bp increasing this size bracket might alleviate this problem; however, a larger insert size would require forward and reverse sequencing and so increase costs. Furthermore, despite the fact that the average dinucleotide repeat is 42-bp long, size does not appear to be correlated with the absence of a sufficient flanking region in inserts smaller than the 30-bp insert (Table 3). Even over 300 bp the problem is still encountered with the longest insert in this study (491 bp) not containing a sufficient flanking sequence to design a forward primer.

When primer pairs could be designed, the majority (66.9%) produced polymorphic fragments within the expected size range (Table 4). This success rate was higher than that achieved in other studies (Roder et al. 1995). Heterozygosity was also high, thus increasing the probability of distinguishing between random individuals. With *M. alternifolia*, loci based on dinucleotide repeats generally produced greater polymorphism (4.2 alleles) than those based on trinucleotides (2.9 alleles). The interpretation of binding patterns produced by dinucleotide repeats was not a problem, especially if using fluorescently labelled capillary electrophoresis. The advantage of fluorescently labelled primers has been reported in previous studies (see Diwan and Cregan 1997 for example).

This extensive study supplied a large number of loci which will provide an abundant pool of polymorphisms useful for several applications. The enrichment procedure enhanced the efficiency of detection of useful loci. Over 85% of the clones sequenced contained SSRs and 12.3% of these produced PCR primers (mono- and poly-morphic). The application of multiplexing and automated sequencing should further improve the efficiency of these markers.

Acknowledgements This research was funded by the Australian Tea Tree Oil Research Institute. The authors wish to acknowledge the contribution of L. Homer, P. Hood and B. Williams. The experiments carried out in this study comply with current Australian laws.

References

Brown SM, Hopkins MS, Mitchell SE, Senior ML, Wang TY, Duncan RR, Gonzalez-Candelas F, Kresovich S (1996) Multiple methods for the identification of polymorphic simple sequence repeats (SSRs) in sorghum [*Sorghum bicolor* (L.) Moench]. Theor Appl Genet 93 : 190–198

Butcher PA, Bell CJ, Moran GF (1992) Patterns of genetic diversity and nature of the breeding system in *Melaleuca alternifolia* (Myrtaceae). Aust J Bot 40 : 365–375

Butcher PA, Doran JC, Slee MU (1994) Intraspecific variation in leaf oils of *Melaleuca alternifolia* (Myrtaceae). Biochem Syst Ecol 22 : 419–430

Butcher PA, Matheson AC, Slee MU (1996) Potential for genetic improvement of oil production in *Melaleuca alternifolia* and *M. linariifolia*. New Forests 11 : 31–51

Byrne M, Marquez-Garcia MI, Uren T, Smith DS, Moran GF (1996) Conservation and genetic diversity of microsatellite loci in the genus *Eucalyptus*. Aust J Bot 44 : 331–341

Diwan N, Cregan PB (1997) Automated sizing of fluorescent-labelled simple sequence repeat (SSR) markers to assay genetic variation in soybean. Theor Appl Genet 95 : 723–733

Dow BD, Ashley MV, Howe HF (1995) Characterisation of highly variable $(GA/CT)_n$ microsatellites in the bur oak, *Quercus macrocarpa*. Theor Appl Genet 91 : 137–141

Edwards KJ, Barker JHA, Daly A, Jones C, Karp A (1996) Microsatellite libraries enriched for several microsatellite sequences in plants. Bio Techniques 20 : 759–760

Gupta PK, Balyan PC, Sharma PC, Ramesh B (1996) Microsatellites in plants: a new class of molecular markers. Curr Sci 70 : 45–54

Johansson M, Ellegren H, Andersson L (1992) Cloning and characterisation of highly polymorphic porcine microsatellites. J Hered 83 : 196–198

Kemp SJ, Brezinsky L, Teale AJ (1993) A panel of bovine, ovine and caprine polymorphic microsatellites. Animal Genet 24 : 363–365

Litt M, Luty JA (1989) A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. Am J Hum Genet 44 : 397–401

Liu ZW, Biyashev RM, Saghai Maroof MA (1996) Development of simple sequence repeat DNA markers and their integration into a barley linkage map. Theor Appl Genet 93 : 869–876

Love JM, Knight AM, McAleer MA, Todd JA (1990) Towards construction of a high-resolution map of the mouse genome using PCR-analysed microsatellites. Nucleic Acids Res 21 : 1111–1115

Ma ZQ, Roder M, Sorrels ME (1996) Frequencies and sequence characteristics of di-, tri- and tetra-nucleotide microsatellites in wheat. Genome 39 : 123–130

Maguire TL, Collins GG, Sedgley M (1994) A modified CTAB DNA extraction procedure for plants belonging to the family Proteaceae. Plant Mol Biol Reporter 12 : 106–109

Panaud O, Chen, X, McCouch SR (1996) Development of microsatellite markers and characterisation of simple sequence length polymorphism (SSLP) in rice (*Oryza sativa* L.). Mol Gen Genet 252 : 597–607

Powell W, Machray GC, Provan J (1996 a) Polymorphism revealed by simple sequence repeats. Trends Plant Sci 7 : 215–222

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

Roder MS, Plaschke J, Konig SU, Borner A, Sorrells ME, Tanksley SD, Ganal MW (1995) Abundance, variability and chromosomal location of microsatellites in wheat. Mol Gen Genet 246 : 327–333

Sagai Maroof MA, Biyashev RM, Yang GP, Zhang Q, Allard RW (1994) Extraordinarily polymorphic microsatellite DNA in barley: species diversity, chromosomal locations, and population dynamics. Proc Natl Acad Sci USA 91 : 5466–5470

Smulders MJM, Bredemeijer G, Rus-Kortekaas W, Arens P, Vosman B (1997) Use of short microsatellites from database sequences to generate polymorphism among *Lycopersicon esculentum* cultivars and accessions of other *Lycopersicon* species. Theor Appl Genet 97 : 264–272

Weber JL, May PE (1989) Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. Am J Hum Genet 44 : 388–396

White G, Powell W (1997) Isolation and characterisation of microsatellite loci in *Swietenia humilis* (Meliaceae); an endangered tropical hardwood species. Mol Ecol 6 : 851–860

Wrigley JW, Fagg M (1993) Bottlebrushes, paperbacks and tea trees, and all other species in the *Leptospermum* alliance. Angus and Robertson, Sydney

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