Theor Appl Genet (2003) 106:751–758
DOI 10.1007/s00122-002-1149-0

Y. Moriguchi · H. Iwata · T. Ujino-Ihara
K. Yoshimura · H. Taira · Y. Tsumura

Development and characterization of microsatellite markers for Cryptomeria japonica D.Don

Received: 27 November 2001 / Accepted: 27 June 2002 / Published online: 15 November 2002

Abstract Thirty four microsatellite markers for *Cryptomeria japonica* D. Don were developed by searching three types of library: a database of *C. japonica* cDNA sequences, a standard non-enriched genomic DNA library and a microsatellite-enriched library using magnetic particles. The enrichment of microsatellite sequences using magnetic particles is very efficient compared to the other two methods both in terms of the numbers of markers generated, and in the polymorphism they detect. The microsatellites developed from the genomic DNA library generally have longer repeat sequences and are more polymorphic than those from cDNA. All the developed microsatellite markers in this study showed polymorphism among 28 plus trees selected from locations scattered throughout Japan. The mean number of alleles per locus (MNA) detected in the 28 plus trees ranged from 2 to 21 with an average of 7.5. The Polymorphism Information Content (PIC) ranged from 0.160 to 0.936 with an average of 0.666. Co-dominant segregation of alleles in a three-generation pedigree of *C. japonica* was demonstrated at 34 microsatellite loci, and the segregation was not distorted from Mendelian expectation for all loci. In 12 out of 34 loci, a null allele was detected. Key relationships between polymorphic parameters, such as MNA and PIC, and the characteristics of microsatellite sequences, such as the longest repeat number, total repeat number and total number of nucleotides, were investigated using rank correlation coefficients, Kendall's τ. A positive correlation was found between repeat lengths and polymorphisms. The markers provide sufficient resolution for investigating gene flow within forests and seed orchards, and for genome mapping.

Keywords Conifers · *Cryptomeria japonica* · Microsatellite · Simple sequence repeat (SSR) · Microsatellite-enriched library

Communicated by D.B. Neale

Y. Moriguchi · H. Taira
Niigata University Graduate School of Science and Technology, Niigata, Niigata 950-2101, Japan

H. Iwata · T. Ujino-Ihara · K. Yoshimura · Y. Tsumura (✉)
Department of Forest Genetics,
Forestry and Forest Products Research Institute, Tsukuba, Ibaraki 305-8687, Japan
e-mail: ytsumu@ffpri.affrc.go.jp
Tel.: +81-298-73-3211 (ext. 444), Fax: +81-298-74-3720

Introduction

Microsatellites, also called simple sequence repeats (SSRs), are tandem repeats of short (1–6 nucleotide) motifs that are interspersed throughout eukaryotic genomes (Stallings et al. 1991; Charlesworth et al. 1994). The first microsatellite markers were developed in humans (Litt and Luty 1989; Weber and May 1989), then in other mammalian and crop species. Microsatellite loci are co-dominant markers and characteristically exhibit high levels of variation in the number of repeat units. Furthermore, once specific primers have been designed, they can be amplified from small amounts of DNA by PCR. Due to these advantages, microsatellite makers are useful and powerful molecular tools for determining parents and for linkage mapping.

Sugi (*Cryptomeria japonica* D. Don) is coniferous species that is naturally distributed in a wide range of Japan and a part of China. The genome size is about 5.0×10^9 bp and the chromosome number is 2n = 22. *C. japonica* is one of the most important commercial afforestation species in Japan, and about 45% of all the man-made forests of Japan are comprised of this species. Because of the importance of *C. japonica* timber, extensive genetic and breeding studies of this species have been performed. Recently, cleaved amplified polymorphic sequences (CAPS), which are PCR-based codominant markers, have been abundantly developed to study its genetic diversity and to design breeding programs (Tsumura et al. 1997; Iwata et al. 2001). Linkage maps of *C. japonica* have also been constructed using several kinds of DNA marker including CAPS (Mukai et al. 1995; Nikaido et al. 2000; Iwata et al. 2001).

Microsatellite markers are more effective for integrating different linkage maps than CAPS markers because of their high degree of polymorphism. Microsatellite markers are also very useful for determining the pollen donors of seeds and/or seedlings within forests. For these reasons, it is very common to develop microsatellite markers. Microsatellite markers have been developed for several coniferous species; from sequence data of a cDNA library such as *Picea abies* (Scotti et al. 2000) and from a standard non-enriched genomic DNA library such as *Pinus radiata* (Smith and Devey 1994), *P. abies* (Peiffer et al. 1997) and *Pinus sylvestris* (Kostia et al. 1995). In a non-enriched method it is difficult to develop microsatellite markers efficiently. Recently, the improvement in methodology using enrichment helped the efficient development of microsatellite markers (Kijas et al. 1994; Lench et al. 1996; Takahashi et al. 1996; Hamilton et al. 1999). There are some reports using a microsatellite-enriched library in coniferous species such as *Pinus taeda* (Elsik and Williams 2001) and *Picea glauca* (Hodgetts et al. 2001). However, they were not well discussed for the characterization of the developed microsatellite markers and the relationship between the repeat length and the polymorphism in each conifer.

We discuss here the efficiency of development using different three methods and the characterization of the developing microsatellite markers. Polymorphism in developed microsatellite markers was investigated using 28 individuals of *C. japonica* in order to evaluate the practical value of the markers in future studies. We also characterized the observed polymorphisms in relation to the repeat length of the microsatellites, to make the future development of microsatellite markers more efficient. Finally, we discuss the relative efficiency of developing microsatellite markers using the three methods exploited in this study.

Materials and methods

Three strategies (cDNA database search, construction standard non-enriched genomic DNA library and construction enrichment DNA library) were used to develop microsatellite markers for *C. japonica*.

cDNA database search

A search was conducted of a *C. japonica* cDNA database including sequence data of approximately 4,500 cDNA clones (Ujino-Ihara et al. 2000) to identify microsatellite sequences. We detected one run of five or more uninterrupted repeats, and ten or more uninterrupted repeats of mononucleotides using the search algorithm developed by K. Yoshimura.

Construction of standard non-enriched genomic DNA library

The F_1 hybrid YA38 between 'Saga3 (female)' and 'Ukiha7 (male)' which are local cultivars was used for the construction of the genomic DNA library. The genomic DNA was extracted from needle tissue of *C. japonica*, YA38, using a modified CTAB method (Tsumura et al. 1995), and was purified using a CsCl density gradient with ultracentrifugation. The genomic DNA of *C. japonica* was digested with *Hin*cII and *Hin*dIII. The digested fragments ranging from 200 to 1,000 bp were ligated into pUC119, and the ligation mixtures were transformed into *Escherchia coli* JM109.

Construction of genomic DNA library enriched for microsatellites

We constructed an enriched library using magnetic particles based on the modified method of Fischer and Bachmann (1998) to enrich fragments including microsatellite sequences. Genomic DNA of YA 38 was digested with *Nde*II, and fragments ranging from 200 to 1,000 bp were ligated into *Sau*3AI cassettes (TaKaRa). After repairing the nicks, the fragments were amplified by PCR with Cassette Primer C1 (TaKaRa). These products were resolved in binding buffer (10 mM Tris-HCl, 1 mM EDTA, 100 mM NaCl, pH 7.5) and hybridized to a 5′ digoxigenin-labelled oligonucleotide probe, 5′DIG$(CT)_{20}$3′, after denaturation. The hybrids were subsequently isolated by binding them to anti-digoxigenin magnetic particles (Boehringer). After rinsing the particles with washing buffer (10 mM Tris-HCl, 1 mM EDTA, 1 M NaCl, pH 7.5), DNA fragments that bound to the probe sequence were recovered by resuspending the particles in elution buffer (6 M Guanidine-HCl). The recovered fragments were then amplified by PCR, and the PCR products were digested with *Nde*II to remove the cassettes. The enriched fragments were ligated into pUC18, and the ligation mixtures were transformed into *E. coli* DH5 α.

Hybridization procedure

Recombinant colonies from the two libraries were transferred to Hybond N^+ membranes (Amersham Pharmacia) separately and screened using two synthetic oligonucleotides, $(CT)_{20}$ and $(CA)_{20}$, labeled by DIG. The hybridization was carried out at 36 °C in hybridization buffer (6 × SSC, 50% formamide) for each probe. Plasmid DNAs from the positive clones were purified using a Quantum Prep Plasmid Miniprep Kit (BIO-RAD) or by an automatic plasmid purification robot (TA 100 Gene Extractor, Takara, Kyoto). The insert DNA of the plasmids was amplified by PCR using primer1 (CCCAGTCACGACGTTGT) and primer3 (GGAAACAGCTATGACCATG). PCR products were purified by an Ultra-Clean PCR purification kit (MO BIO) or Quantum Prep PCR Kleen Spin Columns (BIO-RAD), and were used as template DNA in sequencing reactions. Sequencing was performed using an ABI 377 automatic DNA sequencer with a BigDye Terminator kit (Applied Biosystems) according to the manufacturer's instructions.

Primer design

Based on the sequences containing microsatellites, obtained through the three approaches listed above, PCR primers complementary to the flanking region of the microsatellites were designed, using the OLIGO computer program (National Biosciences). Primers were chosen so that the PCR products would be about 100–300 bp in length (see Table 3).

Polymorphism and inheritance of microsatellite markers in *C. japonica*

To evaluate the polymorphism of the developed microsatellite markers, 28 plus trees (i.e. elite trees selected according to their excellent phenotype), selected from locations scattered throughout Japan, were used as a screening panel. The DNAs of these individuals were extracted from needle tissue using a modified CTAB method (Tsumura et al. 1995). PCR amplifications were carried out using a GeneAmp PCR System Model 9600 (Applied Biosystems) in a total volume of 25 µl containing 20 mM Tris-HCl, pH 8.0, 50 mM KCl, 1.5 mM $MgCl_2$, 0.16 mM of each dNTP,

5 μM of each primer, 5 ng of template DNA, and 0.625 unit of *Taq* polymerase (Promega) or Ex*Taq* polymerase (Takara), with the following temperature profile: 5 min denaturation at 94 °C; followed by 32–40 cycles of 30 s denaturation at 94 °C, 30 s annealing at 52–65 °C, and 30 s extension at 72 °C; with a final extension step of 72 °C for 5 min. Genotypes of 28 individuals for each microsatellite marker were determined by 7.5% polyacrylamide gel electrophoresis as described by Davis (1964) and Ornstein (1964). The PCR products were stained with ethidium bromide or SYBR gold (Molecular probes).

The segregation of alleles at 34 microsatellite loci were compared with expected Mendelian ratios by a χ^2 test. The segregating population of 22 individuals was a sib-crossed pedigree of F_1 hybrids from a cross between 'Saga3 (female)' and 'Ukiha7 (male)' which are local cultivars of *C. japonica*. DNAs of the family and their parents were extracted from needle tissue using a modified CTAB method (Tsumura et al. 1995), and were purified using a CsCl density gradient with ultracentrifugation.

Data analysis

From the genotype data of the 28 individuals comprising the screening panel, the number of alleles per locus (NA), and polymorphism information content, PIC, (Botstein et al. 1980), were obtained for each locus using the program G-DIVERSE developed by H. Iwata (see Table 3). The PIC was calculated as follows:

$$PIC = 2\sum_{i=2}^{l}\sum_{j=1}^{i-1}[p_i p_j(1 - p_i p_j)],$$

where p^i and p^j refer to the frequency of alleles A_i and A_j, respectively, and summation extends over *l*alleles.

The relationships between polymorphic parameters such as the mean number of alleles per locus, MNA and PIC and characteristics of microsatellite sequences such as the longest repeat number (MAX), total repeat number (TOTAL) and the total number of nucleotides (NUC) were examined using the rank correlation coefficient, Kendall's τ (Kendall 1970). The detected microsatellites were classified as follows (Weber 1990). Perfect repeat sequences are defined as tandem repeats without interruption and without adjacent repeats of another sequence. Imperfect repeat sequences are defined as two or more runs of uninterrupted repeats separated by no more than three consecutive non-repeat bases. Terminal runs of uninterrupted repeats (outside of non-repeat bases) must each be at least three full repeats in length. Internal runs of uninterrupted repeats must be at least 1.5 repeats in length. Compound repeat sequences are defined as runs of repeats separated by no more than three consecutive non-repeat bases from a run of at least five uninterrupted repeats of different motif or at least ten uninterrupted mononucleotides. Compound repeats are further divided into perfect or imperfect sub-classes. Kendall's τ values were computed using JMP4 software (SAS Institute).

Result

Microsatellites from cDNA

The database search of approxuimately 4,500 cDNA clones of *C. japonica* identified 303 clones containing five or more repeats. Of these repeats 212 (70.0%) were unique. These included 207 perfect repeats, four imperfect repeats and one compound (perfect) repeat (Table 1a). The 211 non-compound repeats consisted of 79 tri-nucleotides, 87 di-nucleotides and 45 mono-nucleotides (Table 1b). In the microsatellites from the cDNA database search, the most abundant motifs for mono-, di- and tri-nucleotides were A, AT and AAT, respectively (Table 2b).

Table 1 Characteristics of microsatellites detected by searching a cDNA database (cDNA), a standard non-enriched genomic DNA library (standard) and a microsatellite-enriched library (enrichment). The figures show number of clones obtained, and the classification follows Weber (1990)

Item	cDNA	Non-enrichment	Enrichment
(a)			
Perfect repeat	207	3	87
Imperfect repeat	4	–	26
Compound (perfect) repeat	1	–	5
Compound (imperfect) repeat	–	–	3
Unclear	–	–	10
Total	212	3	131
(b)			
Mononucleotide repeat	45	–	–
Di-nucleotide repeat	87	3	122
Tri-nucleotide repeat	79	–	1
Compound	1	–	8
Total	212	3	131

Microsatellites from the genomic library

The approach based on screening a standard non-enriched genomic library generated only three positive clones, which were isolated from about 2,200 genomic *C. japonica* clones. However, from the enriched library, 219 positive clones were isolated from about 2,100 clones. Of the 222 positive clones obtained from these two genomic DNA libraries, 156 (70.3%) had five or more repeats in their inserts. Of these repeats 131 (84.0%) were unique and the longest repeat was $(CT)_{42}$. They included 87 perfect, 26 imperfect, five perfect compound and three imperfect compound repeats. The other ten clones also had microsatellites, but their sequences were unclear and could not be unambiguously resolved (Table 1a). The 123 non-compound repeats consisted of one tri-nucleotide and 122 di-nucleotides (Table 1b). Three kinds of microsatellites were inconsistent with the enriched motif (Table 2a). The microsatellites obtained from the genomic DNA libraries were generally longer than those obtained from cDNA (Table 2a and b).

Primer design

Of the 434 sequences obtained by these three methods, 112 contained either di- or tri-nucleotide repeats with ten or more repeats, excluding mononucleotide repeats. We designed PCR primers to amplify the microsatellite regions with ten or more repeat units, because Weber (1990) reported that the informativeness of human mi-

Table 2 The motifs and numbers of repeats of microsatellites obtained from the genomic DNA library using the enrichment process (a) and cDNA database search (b)

Item	5–9 repeats	10–19 repeats	20–29 repeats	>30 repeats	
(a)					
CT	24	38	34	14	
CA			1	1	
TCC	1				
Unclear					10
Compound					8
Total	25	38	35	15	18

Item	5–9 repeats	10–19 repeats	>20 repeats
(b)			
G		1	
GA	13	4	
GAA	2		
GAC	1		
GAG	1		
GAT	1		
GCA	3		
GCC	3		
GCG	1		
GCT	3		
GGA	3		
GGC	6		
GT	6		
GTA	3		
GTC	3	1	
GTG	4		
A		43	
AAG	6		
AAT	9		
AC	3		
ACC	2		
ACG	1	1	
ACT	1		
AG	14	2	1
AGA	5		
AGC	1		
AGG	6		
AGT	1		
AT	43	1	
ATA	2		
ATC	2		
ATG	2		
ATT	2		
Compound			5
Total	153	53	6

crosatellite sequences with ten or fewer repeats was very low or zero. Of the microsatellites obtained from the cDNA database search, 12 sequences included ten or more repeats. We were able to design PCR primers from three of these sequences. Of the microsatellites obtained from the genomic library, 100 sequences included ten or more repeats. Of these 100 sequences, 36 did not permit primer design because they had insufficient flanking sequences or sequences with low reliability. Finally, we were able to design 67 microsatellite primer pairs from the 112 sequences including ten or more repeats. After PCR optimization, 34 primer pairs each produced a single variable locus (Table 3). The pairs derived from the cDNA and genomic DNA libraries were given designations with the prefixes *Cjcssr* and *Cjgssr*, respectively. Among these *Cjgssr* primers, all but *Cjgssr*01 were developed from the enrichment DNA library.

Characterization of microsatellite markers

The number of alleles per locus, NA, detected in the DNA from the 28 plus trees ranged from two to 21 with an average of 7.5 (Table 3). PIC ranged from 0.160 to 0.936 with an average of 0.666.

Segregation in the sib-crossed pedigree was assessed at 34 microsatellite loci (Table 3). Twenty nine loci were polymorphic but the other five loci were monomorphic in the investigated population. *Cjgssr*140 was excluded from the further test because the pattern was not unambiguously determined in this population. In the result of the χ^2 test, no significant deviations were found, except for *Cjgssr*7. In the segregation analysis, a null allele was detected in 12 out of 34 loci (35.3%).

We examined the correlations between two measures of variation (MNA and PIC) on one hand and three measures of length [(MAX), (TOTAL) and (NUC)] on the other by Kendall's rank order test to characterize the developed microsatellites. *Cjgssr*124 and *Cjgssr*170 were also excluded from the test because its sequence could not be unambiguously determined, and its repeat number was uncertain. We found significant positive correlations between the three measures of length [(MAX), (TOTAL) and (NUC)] and PIC, and between (MAX) and MNA (Table 4).

Discussion

Sequence characterization

Development of microsatellites from the enriched library was more efficient than using standard non-enriched genomic library. In the microsatellites from the enriched library, most motifs were CT except for two CA motifs due to the enrichment using the $(CT)_{20}$ probe (Table 2a). The two CA motifs were presumably detected because we also used a $(CA)_{20}$ probe for hybridization. Except for the microsatellites using the enrichment process, the most abundant motifs for mono-, di- and tri-nucleotides in *C. japonica* were A, AT and AAT, respectively (Table 2b). In rice, the most abundant motifs of detected microsatellites are AT, A, AG and AAT (Wang et al. 1994), in *Arabidopsis* they are A, AT, AAG and AG (Cardle et al. 2000), and in maize, CT, CCT and CCG (Chin et al. 1996). Thus, the *C. japonica* motifs identified in this study are similar to those observed in rice, *Arabidopsis*, and in most common plants (Lagercrantz et al. 1993). In coniferous species such as *Pinus taeda*, *P. strobes* and *P. radiata*, the most abundant motifs were

Table 3 Primer sequences of microsatellite markers in *C. japonica* and the PCR conditions used to amplify them

Locus	Motif	PCR primers	Annealing tempreture (°C)	Fragment size (bp)	NA	PIC	Segregation	Expected	χ^2	Pr
*Cjcssr*01	$(TTA)_{11}$	TGGGAAAAGTAGAGGAAGT AGGGATCTCATTATTCTAACAT	52	195	3	0.160	Invariant			
*Cjcssr*02	$(TA)_{11}$	GACCCCTTCGATTAAAGAAATG TGGCTTTGTGGAGCTTGTT	53	84	2	0.325	Invariant			
*Cjgssr*01	$(CT)_{31}$	TAATGCACCTAGCTAGACTTG AAATGAGGGAGTGATAGTTGT	58	693	5	0.732	9:13	1:1	0.75	0.386
*Cjgssr*5	$(GA)_{12}$	CAAGGTGTGGCAGAGCTGAC GTCCTTTCTCCTCCACCGAATC	65	133	9	0.804	7:11:4	1:2:1	0.88	0.643
*Cjgssr*7	$(TC)_{15}A(TC)_6$	ATCAAATGTGAAGCAAACAGT AGAGGGAAGAGGAAGATAAAGA	60	125	10	0.811	16:6	1:1	5.73	0.017
*Cjgssr*26	$(TC)_3TG(CT)_{10}$	ACTTTTTGCTCTACCTCACAT GCAGAAAGTAGAGAACAAGAAC	60	190	3	0.165	9:13	1:1	0.75	0.386
*Cjgssr*30	$(CT)_{11}$	TCAACATGCTACTATCAACACT TGGTGGGAGAGACCTAGA	60	130	4	0.217	9:13	1:1	0.75	0.386
*Cjgssr*31	$(CT)_{23}$	TCTCTCAATATCCCCTTATTCT GGGGATATAGGGATGTATGG	60	145	10	0.808	7:9:6	1:2:1	0.81	0.668
*Cjgssr*33	$(GT)_{21}$	GGTGGTGTAGAGTTGTCACTT AATTGTCGTTGATTGCTACTTA	62	181	3	0.359	6:16	1:3	0.06	0.811
*Cjgssr*39	$(GA)_{15}$	GGGATAGAGAGCGATAATATAA TCCTTGTTTTTAATTCCTTTC	60	190	4	0.597	10:12	1:1	0.18	0.669
*Cjgssr*47	$(TC)_{19}TT(TC)_4$	GAAAGAGAGGAAGAGACCATAG CCTTGCATCAACCTCACT	60	235	6	0.631	7:5:4:6	1:1:1:1	0.98	0.807
*Cjgssr*77	$(CT)_{10}$	CCTTGTACACTTATTTGTACCT AGGGAGGAGAAATAGACAT	60	106	10	0.785	4:11:7	1:2:1	0.88	0.643
*Cjgssr*78	$(GA)_{21}$	AAGAAGGCATCCAAGAGTGA ACTGCCGATTAACTGATAGCTC	60	157	8	0.787	10:12	1:1	0.18	0.669
*Cjgssr*81	$(TC)_{10}$	TTTCCCTCCTATTCTTTCATTA TTATGAGAGGTTGAGAGTGTTG	60	161	2	0.325	6:10:6	1:2:1	0.18	0.912
*Cjgssr*83	$(CT)_{25}$	GTGGAGTGCTTTCATACTTATT GCTCAGACAAGAGACAACA	60	170	8	0.774	5:13:4	1:2:1	0.92	0.631
*Cjgssr*95	$(AG)_{11}G(GA)_3GG(GA)_3$	ATGGATAGAGATTGAGAGAAAG TCTCCCCCACTCTCTTAC	60	146	3	0.334	Invariant			
*Cjgssr*101	$(GA)_{24}GG(GA)_9$	AGGGATAGAAAATAGAAGAAGG GTTCCTCTTGCATGTTAT	55	167	8	0.766	3:9:3:7	1:1:1:1	5.85	0.119
*Cjgssr*120	$(CT)_{21}CC(CT)_6$	GGGACTTCCCTCCTTTTG AGGACGAACATCATGTAGAAGA	60	230	10	0.842	3:9:10	1:2:1	4.55	0.103
*Cjgssr*121	$(CT)_{25}$	TAGCCATTGTTTTTCATATTTG TGCAACTCTGAAATCCTTTTA	60	205	14	0.910	4:8:4:6	1:1:1:1	1.95	0.583
*Cjgssr*122	$(TC)_{25}$	CCTCATACCCTTTAGTTGACAT GGCACTCATACCCAAAGTAAT	62	198	21	0.936	6:7:3:6	1:1:1:1	2.49	0.477
*Cjgssr*123	$(AG)_3G(AG)_{22}$	TCCATATTACTCTTGCTCAATC AGATGGAGAGGTAGAGATGAA	60	172	8	0.846	8:13	1:1	1.26	0.261
*Cjgssr*125	$(CT)_{18}$	TCTCTGTATATCTCCCACAATC AGGGAAATAGAGGGAAATAGT	60	142	5	0.649	3:13:5	1:2:1	2.18	0.336
*Cjgssr*130	$(CT)_{13}$	TCCCTCCCCATCCATCTC GGGTATCATCAAGCCAAAACAA	60	141	3	0.573	Invariant			
*Cjgssr*140	$(CT)_{21}$	GGGAAGAGGGAAAGAGTTCAT GGAGTGGGAGAGAGGTGAG	60	199	4	0.326	–			
*Cjgssr*142	$(GT)_9(GA)_{21}$	ACCCAATTTTATAGTGACTCCA CGCATCAACTCGACATAATATC	62	261	10	0.860	7:7:7:1	1:1:1:1	21.21	0.000
*Cjgssr*149	$(GA)_{11}$	GGCAACACTAGGGACAACA AGATTGGGAGGAGGAGTATTTA	60	120	11	0.834	8:14	1:1	1.77	0.184
*Cjgssr*159	$(CT)_{23}(CA)_{20}$	CCACCATTTCTCATTTTGTA GGGAGGAGATAGCTTGTAAGTA	60	227	11	0.858	Invariant			
*Cjgssr*175	$(CT)_{16}$	ACCCTTTGAATTATTCCTTGAG AGCGGAAAAAACTACATCAGA	62	186	9	0.820	10:12	1:1	0.18	0.669
*Cjgssr*177	$(TG)_{10}(GA)_{11}A(AG)_3$	ACCGGTATAACTGTAGTAGGTC ATTCTTGGTCTCTCTTCTTTC	60	163	7	0.769	5:12:5	1:2:1	0.18	0.912
*Cjgssr*181	$(AG)_3G(GA)_{23}$	AGAGGGAGGGAGGAATACAT GGCTGAGAGTTTAGGGTTTACA	60	240	8	0.738	5:5:3:8	1:1:1:1	2.66	0.447
*Cjgssr*193	$(CT)_{15}$	CTATCTAGTTCACTGCCAATTC ATGGAGAGTAGCAGAGAAGTT	63	166	8	0.777	10:12	1:1	0.18	0.669

Table 3 (continued)

Locus	Motif	PCR primers	Annealing tempreture (°C)	Fragment size (bp)	NA	PIC	Segregation	Expected	χ^2	Pr
*Cjgssr*124	$(CT)_{36+\alpha}$	AAAATGGGTCATGTCATGT CATTCTCCATCTCACTACCTAT	60	200	11	0.872	4:5:9:2	1:1:1:1	6.53	0.089
*Cjgssr*170	$(CT)_{33+\alpha}$	TGGCGATGTGGTTCTTATGG CAACCGCATCTTTCCCTAATTC	64	219	8	0.833	5:13:4	1:2:1	0.920	0.631
*Cjgssr*13	$(CT)_{13}$	GTCGGGTCCAAGTCAACTAGAA GGGAGACAGGGAGGTATGTAGA	64	270	9	0.824	4:8:3:7	1:1:1:1	3.749	0.290

Table 4 The rank correlation coefficient [Kendall's τ (Kendall 1970)], between two measures of variation (the mean number of alleles (MNA) and polymorphism information content (PIC)) on one hand and three measures of length [the longest repeat number (MAX), the total repeat number (TOTAL) and all number of the nucleotides (NUC)] on the other

	MAX	TOTAL	NUC
MNA	0.2793*	0.2587	0.2486
PIC	0.3141*	0.2821*	0.2721*

*$P < 0.05$

AC or GA (Echt and May-Marquardt 1997; Smith and Devey 1994).

Microsatellite cDNA

Microsatellite markers of tri- and tetra-nucreotide repeat motifs provide greater electrophoretic resolusion than di-nucleotide repeats (Elsik et al. 2000). In this study, tri-nucreotide repeat motifs were detected from cDNA abundantly, though they are of short repeat length. The microsatellites developed from cDNA, *Cjcssr*01 and *Cjcssr*02, generally showed low levels of polymorphism in comparison with those from genomic library. These results are consistent with a study in which the average level of polymorphism was found to be significantly higher for genomic library microsatellites than for cDNA microsatellites in rice (Cho et al. 2000). This corroborates theoretical expectations, since functional restrictions would be expected to maintain low levels of polymorphism among microsatellite sequences in cDNA. The microsatellites contained in cDNAs may be not suitable for developing highly polymorphic microsatellite markers but are important for EST markers.

Polymorphism of microsatellites

The number of alleles per locus detected in the DNA from the 28 plus trees ranged from two to 21 with an average of 7.5 (Table 3). PIC ranged from 0.160 to 0.936 with an average of 0.666. In *P. radiata*, the average PIC value was 0.67 (Smith and Devey 1994), and 0.68 in *P. taeda* (Elsik et al. 2000). Thus, the polymorphism of microsatellites in *C. japonica* is similar to those in other conifers. The level of polymorphism per locus was highest in comparison with other molecular markers (Powell et al. 1996; Russell et al. 1997).

Correlations between variation and length

There were significant positive correlations between the three measures of length [(MAX), (TOTAL) and (NUC)] and PIC, and between (MAX) and MNA (Table 4). This suggests that the length of the longest repeat unit may affect polymorphism in compound or imperfect repeats. This indicates that microsatellites with high repeat numbers have higher variability. The correlations between measures of variation and length in this study are consistent with reports that long microsatellites have high degrees of mutation in *Drosophila melanogaster* (Goldstein and Clark 1995; Schug et al. 1998).

Null allele

In the result of the χ^2 test, no significant deviations were found, except for *Cjgssr*7 and *Cjgssr*142. It might have occurred by chance alone. In the segregation analysis, a null allele was detected in 12 out of 34 loci (35.3%). The null allele has serious problems for population studies due to producing an apparent excess of homozygotes, resulting in incorrect allele frequency estimates and overestimates of inbreeding. The null allele was frequently detected in microsatellite analysis, and in *P. radiata* the frequency of the null allele is approximately 35% (Fisher et al. 1998), and was 38.1% in *Triticum aestivum*, (Stachel et al. 2000). Observation of numerous microsatellite null alleles might be caused by a high mutation rate in this species, since the main cause of null or nonamplifying alleles at microsatellite loci is mutation in the priming sequence (Callen et al. 1993; Jones et al. 1998). Though PCR-null alleles in *C. japonica* might appear to be more than the other organisms, we can use them for genome mapping without any problem after segregation analysis.

Efficient development of microsatellite markers in *C. japonica*

Microsatellite markers have been developed for several coniferous species using various methods (Smith and Devey 1994; Pfeiffer et al. 1997; Khasa et al. 2000; Elsik and Williams 2001). These results suggested that the enrichment process is efficient, and is suitable for large-scale development. Enrichment of target motif or low-copy sequences from a genomic library improves the efficiency of the useful primer pairs identified. There are some methods for the enrichment process, such as using magnetic particle (Kijas et al. 1994), microsatellite arrays (Lunt et al. 1999) and linkers (Hamilton et al. 1999). It is difficult to compare the efficiency of each method directly because the efficiency changes by delicate conditions even in same species. The enrichment process using magnetic particles also promoted the efficiency of the development of microsatellites substantially in *C. japonica*. We could obtain highly informative microsatellite markers in *C. japonica*, and these markers can use efficiency for linkage mapping and for breeding study, such as the understanding of gene flow in seed orchards.

Acknowledgements The authors thank T. Kawahara for information on the enrichment procedure, M. Murai for providing the YA progenies, N. Tani for providing the purified DNAs and helpful comments, and A. Konuma, S. Ueno, T. Sugaya, M. Isshiki and R. Shimamura for their kind advice and help in the laboratory. This work was supported by grants from the Pioneer Special Studies of the Ministry of Agriculture, Fisheries and Forestry, the Program for Promotion of Basic Research Activities for Innovative Biosciences, and Research Fellowships of the Japanese Society for the Promotion of Science for Young Scientists.

References

Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet 32:314–331

Callen DF, Thompson AD, Shen Y, Phillips HA, Richards RI, Mulley JC, Sutherland GR (1993) Incidence and origin of "null" alleles in the $(AC)_n$ microsatellite markers. Am J Hum Genet 52:922–927

Cardle L, Ramsay L, Milbourne D, Macaulay M, Marshall D, Waugh R (2000) Computational and experimental characterization of physically clustered simple sequence repeats in plants. Genetics 156:847–854

Charlesworth B, Sniegowski P, Stephan W (1994) The evolutionary dynamics of repetitive DNA in eukaryotes. Nature 371:215–220

Chin ECL, Senior ML, Shu H, Smith JSC (1996) Maize simple repetitive DNA sequences: abundance and allele variation. Genome 39:866–873

Cho YG, Ishii T, Temnykh S, Chen X, Lipovich L, McCouch SR, Park WD, Ayres N, Cartinhour S (2000) Diversity of microsatellites derived from genomic libraries and GenBank sequences in rice (*Oryza sativa*L.). Theor Appl Genet 100:713–722

Davis BJ (1964) Disc electophoresis. II. Method and application to human serum protein. Ann NY Acad Sci 121:404–427

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

Elsik CG, Williams CG (2001) Low-copy microsatellite recovery from a conifer genome. Theor Appl Genet 103:1189–1195

Elsik CG, Virginia T, Minihan T, Hall SE, Scarpa AM, Williams CG (2000) Low-copy microsatellite markers for *Pinus taeda* L. Genome 43:550–555

Fischer D, Bachmann K (1998) Microsatellite enrichment in organisms with large genomes (*Allium cepa* L.). Biotechniques 24:796–802

Fisher PJ, Richardson TE, Gardner RC (1998) Characteristics of single- and multi-copy microsatellites from *Pinus radiata*. Theor Appl Genet 96:969–979

Goldstein DB, Clark AG (1995) Microsatellite variation in North American populations of *Drosophila melanogaster*. Nucleic Acids Res 23:3882–3886

Hamilton MB, Pincus EL, DiFiore A, Fleischer R (1999) Universal linker and ligation procedures for construction of genomic DNA libraries enriched for microsatellites. BioTechniques 27:500–507

Hodgetts RB, Aleksiuk MA, Brown A, Clarke C, Macdonald E, Nadeem S, Khasa D (2001) Development of microsatellite markers for white spruce (*Picea glauca*) and related species. Theor Appl Genet 102:1252–1258

Iwata H, Ujino-Ihara T, Yoshimura K, Nagasaka K, Mukai Y, Tsumura Y (2001) Cleaved amplified polymorphic sequence markers in sugi, *Cryptomeria japonica* D. Don, and their locations on a linkage map. Theor Appl Genet 163:881–895

Jones AG, Stockwell CA, Walker D, Avise JC (1998) The molecular basis of a microsatellite null allele from the white sands pupfish. J Hered 89:339–342

Kendall MG (1970) Rank Correlation Methods, 3rd edn. Charles Griffin London

Khasa PD, Newton CH, Rahman MH, Jaquish B, Dancik BP (2000) Isolation, characterizaton, and inheritance of microsatellite loci in alpine larch and western larch. Genome 43:439–448

Kijas JMH, Fowler JCS, Garbett CA, Thomas MR (1994) Enrichment of microsatellites from the citrus genome using biotinylated oligonucleotide sequences bound to streptavidin-coated magnetic particles. Bio Techniques 16:657–662

Kostia S, Varvio SL, Vakkari P, Pulkkinen 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

Lench NJ, Norris A, Bailey A, Booth A, Markham AF (1996) Vectorette PCR isolation of microsatellite repeat sequences using anchored dinucleotide repeat primers. Nucleic Acids Res 24:2190–2191

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

Lunt DH, Hutchinson WF, Carvalho GR (1999) An efficient method for PCR-based isolation of microsatellite arrays (PIMA). Mol Ecol 8:891–894

Mukai Y, Suyama Y, Tsumura Y, Kawahara T, Yoshimaru H, Kondo T, Tomaru N, Kuramoto N, Murai M (1995) A linkage map for sugi (*Cryptomeria japonica*) based on RFLP, RAPD, and isozyme loci. Theor Appl Genet 90:835–840

Nikaido AM, Ujino T, Iwata H, Yoshimura K, Yoshimaru H, Suyama Y, Murai M, Nagasaka K, Tsumura Y (2000) AFLP and CAPS linkage maps of *Cryptomeria japonica*. Theor Appl Genet 100:825–831

Ornstein L (1964) Disc electophoresis. I. Background and theory. Ann NY Acad Sci 121:321–349

Pfeiffer A, Olivieri M, Morgante M (1997) Identification and characterization of microsatellites in Norway spruce (*Picea abies* K.). Genome 40:411–419

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

Russel JR, Fuller JD, Macaulay M, Hatz BG, Jahoor A, Powell W, Waugh R (1997) Direct comparison of levels of genetic varia-

tion among barley accessions detected by RFLPs, AFLPs, SSRs and RAPDs. Theor Appl Genet 95:714–722

Schug MD, Hutter CH, Wetterstrand KA, Gaudette MS, Mackay TFC, Aquadro CF (1998) The mutation rates of di-, tri- and tetra-nucleotide repeats in *D. melanogaster*. Mol Biol Evol 15:1751–1760

Scotti I, Magni F, Fink R, Powell W, Binelli G, Hedley PE (2000) Microsatellite repeats are not randomly distributed within Norway spruce (*Picea abies* K.) expressed sequences. Genome 43:41–46

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

Stachel M, Lelley T, Grausgruber H, Vollmann J (2000) Application of microsatellites in wheat (*Triticum aestivum* L.) for studying genetic differentiation caused by selection for adaptation and use. Theor Appl Genet 100:242–248

Stallings RL, Ford AF, Nelson D, Torney DC, Hildebrand CE, Moyzis RK (1991) Evolution and distribution of $(GT)_n$ repetitive sequences in mammalian genomes. Genomics 10:807–815

Takahashi H, Nirasawa K, Furukawa T (1996) An efficient method to clone chicken microsatellite repeat sequences. Jap Poultry Sci 33:292–299

Tsumura Y, Yoshimura K, Tomaru N, Ohba K (1995) Molecular phylogeny of conifers using RFLP analysis of PCR-amplified specific chloroplast genes. Theor Appl Genet 91:1222–1236

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

Ujino-Ihara T, Yoshimura K, Ugawa Y, Yoshimaru H, Nagasaka K, Tsumura Y (2000) Expression analysis of ESTs derived from the inner bark of *Cryptomeria japonica*. Plant Mol Biol 43:451–457

Wang Z, Weber JL, Zhong G, Tanksley SD (1994) Survey of plant short tandem DNA repeats. Theor Appl Genet 88:1–6

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

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