Development and characterization of 28 polymorphic EST-SSR markers for *Cunninghamia lanceolata* (Taxodiaceae) based on transcriptome sequences

By Y. WEN[1),2),*)], S. UENO[2)], W. HAN[1)] and Y. TSUMURA[2)]

(Received 9th November 2012)

Abstract

The Chinese fir (*Cunninghamia lanceolata*) is a very important plantation species that is mainly distributed in Southern China. However, genomic analyses of *C. lanceolata* have lagged significantly behind those of other conifer species due to a lack of suitable markers. To address this issue, we examined 27,666,670 reads from *C. lanceolata* transcriptome sequences recorded in the SRA database. Assembly produced 35,633 contigs, 1,822 (5.11%) of which contained one or more SSRs. In total, 2,156 SSRs were identified, giving an average SSR density of 68.4 SSRs/Mb. The most common SSR types were tri-SSRs (41.7%), followed by hexa-(29.8%), penta-(12.7%), di-(11.1%) and tetra-(4.7%) SSRs. EST-SSR markers based on the 1,822 SSR-containing contigs were developed using the CMiB pipeline and primer pairs were designed to target 35 loci. Polymorphism was observed in 28 of these loci, for which the number of alleles per locus ranged from 3 to 12 over 16 plus-tree individuals. The observed heterozygosity (*Ho*), expected heterozygosity (*He*) and fixation index (F_{IS}) values for the targeted loci ranged from 0.125 to 0.938, 0.225 to 0.891 and –0.368 to 0.456, respectively. The corresponding PIC values ranged from 0.210 to 0.881, with an average of 0.573. Some of these markers have since been used in our ongoing genetic diversity analyses of *C. lanceolata*. To our knowledge, this is the first analysis of SSRs and EST-SSR markers in *C. lanceolata*, and the identified EST-SSR markers will be useful in future genetic analyses of *C. lanceolata* and related species.

Key words: Cunninghamia lanceolata; expressed sequence tag; microsatellite markers; de novo assembly; transcriptome sequences.

Introduction

The Chinese fir (*Cunninghamia lanceolata*) is an allogamous, long-lived, evergreen conifer that is mainly distributed in Southern China. It is one of the most important plantation species in this region and is widely used to restore degraded land and as a source of commercial timber in China. However, genetic analyses of this species have lagged behind those of other conifers due to a lack of suitable markers. To date, only a few studies have examined the genetic diversity and phylogeny of *C. lanceolata* and related species, typically by analyzing dominant DNA markers using techniques such as RAPD (YOU and HONG, 1998), ISSR (QI, 2008; YANG et al., 2009) and AFLP (CHUNG et al., 2004). To the authors' knowledge, no microsatellite markers have yet been developed for *C. lanceolata*, which has seriously impeded its genetic analysis and the development of breeding programs.

Genetic studies based on microsatellite markers (also known as simple sequence repeats, or SSRs) have a number of advantages over those based on random fingerprinting approaches (JARNE and LAGODA, 1996; SELKOE and TOONEN, 2006), including the co-dominant inheritance of the markers according to Mendelian laws, their high information content, and their ready detection by PCR amplification. EST-SSR markers offer further advantages over genomic SSRs in that they directly reflect gene expression. However, to develop new species-specific microsatellites, it is necessary to study their flanking sequences using traditional approaches, which is costly and time-consuming (ZANE et al., 2002; SQUIRRELL et al., 2003). With the advent of second generation sequencing technologies, very large amounts of DNA sequence data have been generated and deposited in public databases such as the Sequence Read Archive (SRA), which include high-throughput transcriptome sequences from different taxa. These data are valuable genomic resources that can be used to develop new and specific SSR markers for target species (ZALAPA et al., 2012). In this work, we mined SSRs to develop and characterize a set of EST-SSR markers for *C. lanceolata* from publically-available transcriptome sequences.

Materials and Methods

Plant materials

In order to evaluate the polymorphism of the candidate EST-SSR markers, we selected 16 plus tree individuals from two seed orchards in Southern China. Twelve of these individuals (Y6, Y18, J5, J80, Ht14, Ht16, Jh10, Jh16, Y26, J18, Y110, Y1116) were obtained from the Youxian seed orchard in Hunan province (N27°18', E113°47') and the remaining four were obtained from the Lechang seed orchard in Guangdong province (Lc6, Lc12, Lc18, Lc418) (N25°12', E113°28'). Leaves from the selected individuals were dried over silica gel and their DNA was extracted using a modification of the CTAB method (TSUMURA et al., 1995).

SSR mining and primer design

Transcriptome sequences for *C. lanceolata* (Accession: SRX151872) were downloaded from the NCBI's SRA

1) Central South University of Forestry and Technology, Changsha, Hunan 410004, China.

2) Forestry and Forest Products Research Institute, Tsukuba, Ibaraki 305-8687, Japan.

*) Author for correspondence: YAFENG WEN.
E-Mail: wenyafeng7107@163.com

DOI:10.1515/sg-2013-0018
edited by Thünen Institute of Forest Genetics

database (http://www.ncbi.nlm.nih.gov/sra/SRX151872). CLC Genomics Workbench (http://www.clcbio.com) was used to trim and assemble sequence reads, and the recently-developed CMiB pipeline (UENO et al., 2012) was used to mine SSRs and design PCR amplicon primers for contigs that contained SSRs. The CMiB

Table 1. – Basic properties of the 28 polymorhphic EST-SSR markers developed for *C. lanceolata*.

Locus	Primer sequence (5′- 3′)	Repeat motif	Blastx top hit description	*E* -value	GenBank accession no
contig3078_1424A	F: **GCCTCCCTCGCGCCA**CAATCAGCCAAGTTGTACAGGC R: CATACCTTAGCAAAGCCCTCAGC	$(AT)_8(AG)_{18}$	N		AB757708 AB757709
contig3400_101A	F: **GCCTCCCTCGCGCCA**TGAAATTGCGTTGTACCGAAGG R: TAACGAGACGAGCGACAATCTCC	$(GA)_{13}$	N		AB749554 AB749555
contig5354_691A	F: **GCCTCCCTCGCGCCA**GATCCTCTGGTACTTGGTGCCC R: TGCAAAGTCATGTCATCTCTGGC	$(AT)_9$	N		AB749556 AB749557
contig5410_1886A	F: **GCCTCCCTCGCGCCA**GGCTCGAGTTTGCATCTCACAC R: CACATCCAATCCATACAGGAGGG	$(TC)_9$	N		AB749558 AB749559
contig9724_201A	F: **GCCTCCCTCGCGCCA**GGTCCGGGCATTTAGAGTCATC R: GCTCAGATCCAAGGTGACTCAGG	$(AT)_9$	N		AB749560 AB749561
contig16147-262A	F: **GCCTCCCTCGCGCCA**TGAATGGACTGCCACAAATTCC R: TTCTTTGCAGGAAAGCCAACAAG	$(AG)_{10}$	N		AB749550 AB749551
contig16322_179A	F: **GCCTCCCTCGCGCCA**CTGGCATGTAAAGACCATGTTAGG R: GGCTGAGCCTTTAGTGTATCTTCCC	$(TA)_9$	N		AB749552 AB749553
contig1382_349B	F: **GCCTTGCCAGCCCGC**CTTAAGATAGCAGCGGGAATGG R: CTTGCTCGATTTCTTGCATCTGG	$(CT)_{11}$	N		AB749562 AB749563
contig1997_271B	F: **GCCTTGCCAGCCCGC**AGAAATGTATACGGACCCTGCG R: AAATCAAAGCACACGGTGAGAGC	$(TA)_9$	N		AB749564 AB749565
contig4417_459B	F: **GCCTTGCCAGCCCGC**AAGAGAAGAGGAGGAGGTCCAAG R: CAGGAGCAGGTGCAGTAGCATTC	$(AG)_9$	N		AB749570 AB749571
contig4728_384B	F: **GCCTTGCCAGCCCGC**ATTATCCGAGGCAGATACGCAC R: CTTCTCCGTATTTGATCCATCGC	$(GGA)_{10}$	N		AB749572 AB749573
contig7616_683B	F: **GCCTTGCCAGCCCGC**GAGCCGTGAAGAACGAAGGTCTC R: ACGATCGGATTGTCTCAGAAACG	$(GAA)_{12}$	N		AB749574 AB749575
contig7671_1267B	F: **GCCTTGCCAGCCCGC**TGATCTTGGCATGTCAGTCTGG R: TGTCTGTCTGCCTGCAGTTATGC	$(AT)_9$	N		AB749576 AB749577
contig20158_829B	F: **GCCTTGCCAGCCCGC**TCCACACCTTGCTGCTCTCTTC R: GAGATTAGGGCACTAGCGATGGG	$(AT)_9$	N		AB749566 AB749567
contig25400_116B	F: **GCCTTGCCAGCCCGC**AGCGATTAGAATCCGAGCAGAG R: GCTCGAGATCTGCGTAGGAAGTG	$(ATC)_9$	Hypothetical protein (Flowering promoting factor 1)	3.1E-13	AB749568 AB749569
contig406_1209C	F: **GGACCAGGCTACCGTG**TCATCAGCCTCAGTTTGTACTTGC R: GCAATCATGGGCTCTCTGCAC	$(AT)_9$	N		AB749584 AB749585
contig2573_171C	F: **AGGACCAGGCTACCGTG**AATGCGACTTGCAAATTTCTGG R: CGAATTCCTCAATCACTTGGCTG	$(AGA)_{10}$	N		AB749582 AB749583
contig6064_1563C	F: **GGACCAGGCTACCGTG**CACAGTCAATGTCCGTTATCGTTC R: AATGGCAGCAACATCAGAAATGG	$(TA)_9$	N		AB749586 AB749587
contig6319_250C	F: **CAGGACCAGGCTACCGTG**GCGGCCATTTATATCATCTTC R: CACGCCTGTAATTCATCTCCGTC	$(GAA)_9$	N		AB749588 AB749589
contig12886_2058C	F: **CAGGACCAGGCTACCGTG**GGAGCCCTTAGAGTTACGGAG R: TGGGCTCCATTCATTTGTACTGC	$(ATA)_9$	N		AB749578 AB749579
contig16181_1285C	F: **AGGACCAGGCTACCGTG**GGTACTGCGAATCTTCAAATCC R: TGTTCAAGAAAGGAAGCAAACGG	$(TC)_9$	N		AB749580 AB749581
contig476_526D	F: **GGAGAGCCGAGAGGTG**TTTGGGACCTTATGGAGGTGGAG R: AAACCACCAGGTTGAGAAGCAGC	$(GGA)_9$	N		AB749602 AB749603
contig1560_1789D	F: **GGAGAGCCGAGAGGTG**TTTCGGCTCTCCGACTCCTTAAC R: AGAATCGCGTCCAGAACACAGAG	$(CT)_{11}$	N		AB749594 AB749595
contig4056_974D	F: **GAGAGCCGAGAGGTG**TCCAGGAGTCTGTGAATCCGAAG R: CAGTACCAATTCAACCCAGCAGC	$(CTG)_9$	N		AB749600 AB749601
contig10192_1677D	F: **GAGAGCCGAGAGGTG**TCAAGAAGTTCCGCCATTGAGAG R: CCCATGAGGATTCAGAAACATGC	$(CTT)_{10}$	N		AB749590 AB749591
contig14033_236D	F: **GGAGAGCCGAGAGGTG**TTAATGGTGCAAGGTGGAATTGG R: TTGAATCCCACTGATCACACTGC	$(GAA)_{10}$	N		AB749592 AB749593
contig16781_913D	F: **CGGAGAGCCGAGAGGTG**CTGTTTGTACATTGGCCTCGAC R: TCACAAACCACTGTGCTGGAATG	$(CTC)_9$	N		AB749596 AB749597
contig18815_185D	F: **CGGAGAGCCGAGAGGTG**GTTGGGATGTCATCAAGATTGG R: TGAAAGAGGCGGAAATTGGTAGG	$(TC)_{11}$	N		AB749598 AB749599

Note: N No hits found; 4 universal primers sequence (5'-3') and their PCR fluorophore:
Tail A (FAM), GCCTCCCTCGCGCCA; Tail B (HEX), GCCTTGCCAGCCCGC;
Tail C (NED), CAGGACCAGGCTACCGTG; Tail D (PET), CGGAGAGCCGAGAGGTG.

DOI:10.1515/sg-2013-0018
edited by Thünen Institute of Forest Genetics

pipeline is based on a combination of the CD-HIT-EST, MISA, ipcress and BlastCLUST packages, and produces non-redundant SSR sequences along with corresponding primer pairs. 35 di- and tri-nucleotide microsatellite loci containing SSRs with nine or more repeats each were selected for further evaluation.

PCR amplification and polymorphism testing

Genotyping was performed using fluorescently labelled universal primers and a three-primer PCR approach (BLACKET et al., 2012). Each locus-specific forward primer was modified with a 5' universal primer sequence tail (A, B, C or D; see *Table 1*). The primer pairs were then tested against DNA from two sampled individuals to verify that they afforded successful amplification. PCR was performed according to the protocol supplied with the QIAGEN® Multiplex PCR kit, in a 10.0 µL reaction volume containing 1 × Multiplex PCR master mix (Qiagen), 1 µL of 10 × primer mix, and 5–10 ng of template DNA. The 10 × Primer mix for each marker in the multiplex contained 1.0 µM universal fluorescent primer, 1.0 µM tailed forward primer and 2.0 µM reverse primer. The PCR amplification conditions were: 15 min at 95°C, then 35 cycles of 30 s at 94°C, 90 s at 60°C, and 60 s at 72°C, followed by a 30 min extension step at 60°C. The PCR products were electrophoretically separated on 2% agarose gels stained with ethidium bromide. For loci that were successfully amplified, singleplex PCR amplification was performed to investigate their polymorphism in the 16 sampled individuals. The PCR amplification system and conditions used in this second round of amplifications were the same as those described above. Four fluorescently labelled markers with different universal primers (A, B, C or D) were pooled in a single lane for allele screening and analyzed with an ABI3100 genetic analyzer using the Liz 600 size standards. GENALEX6.3 (PEAKALL and SMOUSE, 2006) was used to estimate a range of genetic parameters for each locus, including the number of alleles per locus (*Na*), observed heterozygosity (*Ho*), expected heterozygosity (*He*), and fixation index (F_{IS}). The polymorphism information content (*PIC*) (BOTSTEIN et al., 1980) was determined for each locus using MS-tools (PARK, 2001). Exact tests of Hardy-Weinberg equilibrium (*HWE*) and Linkage disequilibrium (*LD*) were assessed using version 4.2 of the on-line GENEPOP tool (RAYMOND and ROUSSET, 1995; http://genepop.curtin.edu.au/index.html). Null alleles were detected using Microchecker 2.2.3 (VAN OOSTERHOUT et al., 2004). In order to identify putative functions for SSR-containing contigs, they were blasted against the NCBI NR protein database using Blast2GO (CONESA et al., 2005) with an e-value cutoff of 1e-3.

Results and Discussion

SSRs mining of EST contigs

We examined 27,666,670 reads of *C. lanceolata* derived from transcriptome sequences. Assembly generated 35,633 contigs, of which 1,822 (5.11%) contained one or more SSRs. In total, 2,156 SSRs were identified with threshold repeat numbers of 6, 5, 4, 3 and 3 for di-, tri-, tetra-, penta- and hexa-SSRs, respectively. The average SSR density was 68.4 SSRs/Mb, which is much lower than that observed in *Cryptomeria japonica* (UENO et al., 2012), but higher than the values for loblolly pine and spruce (BÉRUBÉ et al., 2006). Tri-SSRs were the most common SSR type encountered (41.7%), followed by hexa-(29.8%), penta-(12.7%), di-(11.1%) and tetra-(4.7%) SSRs. The most common tri-SSR motif was AAG, which was found in 239 (26.6%) of all tri-SSRs. The most common di-SSR motifs were AG and AT, both of which occurred 109 times (45.6% of all di-SSRs). With the advent of high-throughput sequencing technologies, several computational tools or pipelines have been developed for SSR mining, including TROLL (CASTELO et al., 2002), MISA (THIEL et al., 2003), CMiB (UENO et al., 2012) and ESMP (SARMAH et al., 2012). CMiB is a new method developed in our laboratories that emphasizes the identification of unique primer pairs targeting specific genes and relies on a combination of widely-used programs. It has previously been successfully used to mine SSRs for *Cryptomeria japonica*.

Characterization of selected EST-SSR markers

Of the 35 primer pairs designed, 29 (82.9%) produced clear PCR fragment patterns with one or two bands, one produced multiple (more than two) bands, and 5 (14.3%) did not produce any fragments. Polymorphic genotypes were obtained for 28 loci (*Table 1* and *Table 2*) with the number of alleles per locus ranging from 3 to 12 for the 16 studied individuals. The observed heterozygosity (*Ho*), expected heterozygosity (*He*) and fixation index (F_{IS}) values for each locus ranged from 0.125 to 0.938, 0.225 to 0.891 and –0.368 to 0.456, respectively. Two loci (contig4417_459B and contig16147_262A) showed significant deviation from Hardy-Weinberg equilibrium (*HWE*) due to heterozygote deficiency or null alleles. The *PIC* values for the loci ranged from 0.210 to 0.881, with an average of 0.573. These values indicate that the identified EST-SSR markers exhibit a high degree of polymorphism. Significant linkage disequilibrium was observed for eleven of the 378 possible pairs of the 28 EST-SSR markers ($p < 0.05$). Possible null alleles were identified in two loci (contig4417_459B and contig3078_1424A), with null allele frequencies of 0.22 and 0.11, respectively.

In order to assess the power of these markers, Principal Coordinates Analysis (PCA) was performed using GENALEX6.3. The four individuals from the Lechang seed orchard formed a small cluster based on the PCA results (data not shown), indicating that they were closely genetically related, while the 12 individuals from the Youxian seed orchard formed looser clusters because they were collected over a broader range in Hunan Province. Since their identification, some of the markers developed in this work have been used in our ongoing analysis of genetic diversity in *C. lanceolata*. To our knowledge, the results presented herein represent the first analysis of SSRs and EST-SSR markers in *C. lanceolata*. The newly-established EST-SSR markers will be very useful in future genetic analyses of *C. lanceolata* and closely related species.

DOI:10.1515/sg-2013-0018
edited by Thünen Institute of Forest Genetics

Table 2. – 28 Polymorphic EST-SSR markers of *C. lanceolata* screening in 16 individuals.

Locus	N	Size Rang(bp)	*Na*	*Ho*	*He*	F_{IS}	*PIC*	Null alleles (frequency)
contig3078_1424A	16	343-375	12	0.688	0.883	0.221	0.872	Yes (0.11)
contig3400_101A	16	273-297	10	0.875	0.861	-0.016	0.846	No
contig5354_691A	15	186-190	3	0.400	0.384	-0.040	0.351	No
contig5410_1886A	16	229-233	3	0.625	0.635	0.015	0.557	No
contig9724_201A	16	211-241	10	0.813	0.846	0.039	0.829	No
contig16147-262A	16	288-298	5	0.500	0.668	0.251*	0.623	No
contig16322_179A	16	282-290	4	0.563	0.637	0.117	0.587	No
contig1382_349B	16	258-279	6	0.625	0.648	0.036	0.607	No
contig1997_271B	16	247-255	5	0.813	0.594	-0.368	0.535	No
contig4417_459B	16	269-283	5	0.375	0.689	0.456**	0.646	Yes (0.22)
contig4728_384B	15	336-354	6	0.400	0.444	0.100	0.425	No
contig7616_683B	16	260-281	7	0.688	0.785	0.124	0.752	No
contig7671_1267B	16	144-158	5	0.563	0.664	0.153	0.607	No
contig20158_829B	16	324-344	9	0.813	0.795	-0.022	0.774	No
contig25400_116B	16	337-349	5	0.750	0.713	-0.052	0.660	No
contig406_1209C	16	348-384	11	0.813	0.869	0.065	0.855	No
contig2573_171C	16	240-258	4	0.438	0.416	-0.052	0.392	No
contig6064_1563C	16	357-363	4	0.688	0.664	-0.035	0.616	No
contig6319_250C	16	126-135	4	0.563	0.609	0.077	0.530	No
contig12886_2058C	16	231-240	3	0.563	0.506	-0.112	0.406	No
contig16181_1285C	16	293-297	3	0.375	0.490	0.235	0.397	No
contig476_526D	16	153-159	3	0.313	0.275	-0.135	0.257	No
contig1560_1789D	16	129-147	5	0.438	0.492	0.111	0.458	No
contig4056_974D	16	223-238	3	0.125	0.225	0.443	0.210	No
contig10192_1677D	16	354-375	6	0.750	0.750	0.000	0.717	No
contig14033_236D	16	280-286	3	0.250	0.271	0.079	0.248	No
contig16781_913D	16	136-148	4	0.500	0.416	-0.202	0.392	No
contig18815_185D	16	226-270	12	0.938	0.891	-0.053	0.881	No

Note: N number of individuals genotyped; *Na* number of alleles per locus detected; *Ho* observed heterozygosity; *He* expected heterozygosity; F_{IS} fixation index; *,** significant departure from Hardy-Weinberg equilibrium (*HWE*) ($P<0.05$, $P<0.001$, respectively); *PIC* polymorphism information content.
List of 11 significant linkage disequilibrium pairs markers ($P<0.05$)
contig1997_271B and contig2573_171C, contig1997_271B and contig5354_691A, contig4056_974D and contig5354_691A, contig476_526D and contig5410_1886A, contig2573_171C and contig12886_2058C, contig4417_459B and contig16181_1285C, contig7671_1267B and contig16147-262A, contig18815_185D and contig16147-262A, contig1560_1789D and contig20158_829B, contig5410_1886A and contig20158_829B, contig4417_459B and contig16781_913D.

Acknowledgements

This research was supported by National Natural Science Foundation of China (30972357), the Hunan Provincial Natural Science Foundation of China (10JJ2018), and by a Grant-in-Aid for JSPS Fellows (11F00787).

Reference

BÉRUBÉ, Y., J. ZHUANG, D. RUNGIS, S. RALPH, J. BOHLMANN and K. RITLAND (2007): Characterization of EST-SSRs in loblolly pine and spruce. Tree Genetics & Genomes **3**: 251-259.

DOI:10.1515/sg-2013-0018
edited by Thünen Institute of Forest Genetics

Blacket, M. J., C. Robin, R. T. Good, S. F. Lee and A. D. Miller (2012): Universal primers for fluorescent labelling of PCR fragments – an efficient and cost-effective approach to genotyping by fluorescence. Molecular Ecology Resources **12**: 456–463.

Botstein, D., R. L. White, M. Skolnick and R. W. Davis (1980): Construction of a genetic linkage map in man using restriction fragment length polymorphism. Am J Hum Genet **32**: 314–331.

Castelo, A. T., W. Martins and G. R. Gao (U2002): Trolltandem repeat occurrence locator. Bioinformatics Applications Note **18**: 634–636.

Chung, J. D., T. P. Lin, Y. C. Tan, M. Y. Lin, S. Y. Hwang (2004): Genetic diversity and biogeography of *Cunninghamia konishii* (Cupressaceae), an island species in Taiwan: a comparison with *Cunninghamia lanceolata*, a mainland species in China. Molecular Phylogenetics and Evolution **33**: 792–801.

Conesa, A., S. Gotz, J. M. Garcia-Gomez, J. Terol, M. Talon and M. Robles (2005): Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics **21**: 3674–3676.

Jarne, P and P. J. Lagoda (1996): Microsatellites, from molecules to populations and back. Trends in Ecology and Evolution **11**: 424–429.

Park, S. D. E. (2001): Trypanotolerance in West African cattle and the population genetic effects of election [Ph.D. thesis]. University of Dublin.

Peakall, R. and P. E. Smouse (2006): GENALEX 6: Genetic analysis in Excel. Population genetic software for teaching and research. Molecular Ecology Notes **6**: 288–295.

Qi, M. (2008): Genetic diversity of wide cross population of *Cunninghamia lanceolata* and *platycladu orientalis*. Bulletin of botanical research **28**: 299–303.

Sarmah, R., J. Sahu, B. Dehury, K. Sarma, S. Sahoo, M. Sahu, M. Barooah, P. Sen and M. K. Modi (2012): ESMP: A high-throughput computational pipeline for mining SSR markers from ESTs. Bioinformation **8**: 206–208.

Raymond, M. and F. Rousset (1995): Genepop (Version-1.2): Population genetics software for exact tests and ecumenicism. Journal of Heredity **86**: 248–249.

Selkoe, K. A. and R. J. Toonen (2006): Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. Ecology Letters **9**: 615–629.

Squirrell, J., P. M. Hollongsworth, M. Woodhead, J. Russell, A. J. Lowe, M. Gibby and W. Powell (2003): How much effort is required to isolate nuclear microsatellites from plants? Molecular Ecology **12**: 1339–1348.

Thiel, T., W. Michalek, R. K. Varshney and A. Graner (2003): Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). Theor Appl Genet **106**: 411–422.

Tsumura, Y., K. Yoshimura, N. Tomaru and K. Ohba (1995): Molecular phylogeny of conifers using RFLP analysis of PCR-amplified specific chloroplast genes. Theoretical and Applied Geneticst **91**: 1222–1236.

Ueno, S., Y. Moriguchi, K. Uchiyama, T. U. Ihara, N. Futamura, T. Sakurai, K. Shinohara and Y. Tsumura (2012): A second generation framework for the analysis of microsatellites in expressed sequence tags and the development of EST-SSR markers for a conifer, *Cryptomeria japonica*. BMC Genomics **13**: 136.

Van Oosterhout, C., W. F. Hutchinson, D. P. M. Wills and P. Shipley (2004): MICRO CHECKER: software for identifying and correcting genotyping errors in microsatellite data. Molecular Ecology Notes **4**: 535–538.

Yang, Y. L., X. Q. Ma and M. Q. Z. (2009): Molecular polymorphic analysis for different geographic provenances of Chinese fir. Journal of Tropical and Subtropical Botany **17**: 183–189.

You, Y. and J. S. Hong (1998): Application of RAPD marker of genetic variation of Chinese fir provenances. Scientia Silvae Sinicae **34**: 33–38.

Zalapa, J. E., H. Cuevas, H. Zhu, S. Steffan, D. Senalik, E. Zeldin and B. McCown et al. (2012): Using next-generation sequencing approaches for the isolation of simple sequence repeat (SSR) loci in the plant sciences. American Journal of Botany **99**: 193–208.

Zane, L., L. Bargelloni and T. Patarnello (2002): Strategies for microsatellite isolation: A review. Molecular Ecology **11**: 1–16.

DOI:10.1515/sg-2013-0018
edited by Thünen Institute of Forest Genetics