GMR Genetics and Molecular Research
Online Journal
ISSN:1676-5680

Development and cross-species transferability of unigene-derived microsatellite markers in an edible oil woody plant, *Camellia oleifera* (Theaceae)

B.G. Jia[1,2], Q. Lin[1,2], Y.Z. Feng[3], X.Y. Hu[4], X.F. Tan[1,2], F.G. Shao[1,2] and L. Zhang[1,2]

[1]Key Laboratory of Cultivation and Protection for Non-Wood Forest Trees, Ministry of Education, Central South University of Forestry and Technology, Changsha, Hunan, China
[2]Cooperative Innovation Center of Cultivation and Utilization for Non-Wood Forest Trees of Hunan Province, Central South University of Forestry and Technology, Changsha, Hunan, China
[3]China Paulownia Research Center, Zhengzhou, Henan, China
[4]School of Forestry and Landscape Architecture, Anhui Agricultural University, Hefei, Anhui, China

Corresponding author: L. Zhang
E-mail: triwoodtim918@126.com

Genet. Mol. Res. 14 (2): 6906-6916 (2015)
Received December 18, 2014
Accepted April 8, 2015
Published June 18, 2015
DOI http://dx.doi.org/10.4238/2015.June.18.33

ABSTRACT. *Camellia oleifera* is an important edible oil woody plant in China. Lack of useful molecular markers hinders current genetic research on this tree species. Transcriptome sequencing of developing *C. oleifera* seeds generated 69,798 unigenes. A total of 6949 putative microsatellites were discovered among 6042 SSR-containing unigenes. Then, 150 simple sequence repeats (SSRs) were evaluated in 20 varieties of *C. oleifera*. Of these, 52 SSRs revealed

polymorphism, with the number of alleles per locus ranging from 2 to 15 and expected heterozygosity values from 0.269 to 0.888. The polymorphic information content varied from 0.32 to 0.897. Cross-species transferability rates in *Camellia chekangoleosa* and *Camellia japonica* were 90.4 and 78.8%, respectively. The 52 polymorphic unigene-derived SSR markers serve to enrich existing microsatellite marker resources for *C. oleifera* and offer potential for applications in genetic diversity evaluation, molecular fingerprinting, and genetic mapping in *C. oleifera*, *C. chekangoleosa*, and *C. japonica*.

Key words: *Camellia oleifera*; *C. chekangoleosa*; *C. japonica*; Unigene; Microsatellite; Cross-species transferability

INTRODUCTION

The woody plant *Camellia oleifera*, a member of the Theaceae family, is economically important for the production of tea oil in China. Tea oil, extracted from the seeds, is considered a high-quality edible oil and is also called ‘‘eastern olive oil’’ because its composition is highly similar to that of olive oil (Ma et al., 2011). Oil-tea trees include more than ten species, of which *C. oleifera* is the dominant species in tea oil production, being the most widely distributed and planted in China and producing the highest yields. A number of authorized elite varieties have so far been released for tea oil production. However, more new varieties with novel traits, such as high quality, multipurpose use, and biotic and abiotic stress resistance, are required for commercial plantations. Understanding the genetic diversity and relatedness among germplasm resources is useful for breeding and clonal improvements. Microsatellite markers, also called simple sequence repeat (SSR) markers, are powerful tools for genetic diversity evaluation, molecular fingerprinting, and genetic mapping. In the past few years, transcriptome sequencing technology has developed to offer a fast, cost-effective, and reliable approach to the generation of large expression-data sets, in both model and non-model plants with large, complex genomes (Marioni et al., 2008; Mortazavi et al., 2008; Nagalakshmi et al., 2008). In addition, this technology provides an opportunity to identify and develop unigene-derived microsatellite markers (Gupta and Gopalakrishna, 2010; Dutta et al., 2011; Zhang et al., 2012). These new markers are considered superior to genomic SSR markers because they potentially code for functional proteins and can increase the efficiency of marker-assisted selection.

Recently, microsatellites have been analyzed and developed in *Camellia chekangoleosa*, an allied species of *C. oleifera* also used for tea oil production (Wen et al., 2012; Shi et al., 2013). *C. chekangoleosa* is a diploid species, whereas different ploidy levels occur in *C. oleifera* (Huang, 2013). *C. oleifera*-derived specific microsatellite markers have not yet been reported, hindering genetic research efforts. In our previous study, we performed Illumina platform-based transcriptome sequencing of developing *C. oleifera* seeds in order to understand seed fatty acid metabolism (Shao, 2011). The objective of this study was to characterize microsatellites from the transcriptome sequences and develop polymorphic microsatellite markers in *C. oleifera*. These novel unigene-derived microsatellite markers will provide a useful tool for genetic research and comparative genome analysis in *C. oleifera* and allied species.

MATERIAL AND METHODS

Plant materials

Plant materials used for validation of SSRs comprised 20 varieties of *C. oleifera,* including the 'Huashuo' cultivar, which we utilized for complementary DNA (cDNA) library construction and transcriptome sequencing. In addition, 18 elite germplasm clones of *C. chekangoleosa* and 15 varieties of *Camellia japonica* were used to investigate cross-species transferability. All plants were conserved at the Germplasm Repository of the Jiangxi Academy of Forestry, Nanchang, China. Young leaves were collected in spring and stored at -80°C until use.

Source of transcriptome sequences

Our laboratory, in collaboration with Beijing Genomics Institute in Shenzhen, China, constructed a cDNA library using developing seeds of the 'Huashuo' cultivar at the lipid synthesis initiation phase and the peak lipid synthesis phase, i.e., 180 and 300 days after flowering, respectively. Detailed procedures for cDNA library construction, sequencing, and *de novo* transcriptome assembly were described previously (Shao, 2011).

Detection of microsatellites and designing of SSR primers

Microsatellites were detected using the Microsatellite (MISA) tool, with parameters set for detection of perfect di-, tri-, tetra-, penta-, and hexanucleotide motifs with at least six, five, five, four, and four repeats, respectively. Primer pairs were designed using Primer Premier 5.0. The major parameters for primer design were set as follows: SSR motifs ≥ 20 bp, primer length from 18 to 25 nucleotides, and PCR product size from 100 to 300 bp (Table 1).

Validation of microsatellite markers

Genomic DNA was isolated from young leaves of 20 *C. oleifera* varieties, 18 *C. chekangoleosa* clones, and 15 *C. japonica* varieties using a DNA isolation kit (Tiangen Biotech, China). Of the 20 *C. oleifera* varieties, three were first used to test the primers. Each PCR was carried out in a total volume of 10 μL containing 1X buffer, 2 mM $MgCl_2$, 0.2 mM of each dNTPs, 0.2 μM of each primer, 0.5 U *Taq* DNA polymerase (Tiangen Biotech), and 10 ng DNA template. PCR was performed on a thermal cycler (ABI9700, Applied Biosystems, USA) under the following conditions: 94°C for 5 min; 30 cycles of 30 s at 94°C, 45 s at 56°C, and 45 s at 72°C; 10 cycles of 30 s at 94°C, 45 s at 53°C, and 45 s at 72°C; and a final extension of 5 min at 72°C. The amplified products were analyzed on 2% agarose gels. Any loci generating products of the expected size were then assessed for polymorphisms in the 20 *C. oleifera* varieties, using the M13 (-21) (5'-TGTAAAACGACGGCCAGT-3') sequence-tag method (Schuelke, 2000). Fluorescently labeled PCR products were analyzed concurrently with the GeneScan-500 LIZ Size Standard on an ABI 3730XL sequencer, and sizes were determined with GeneMapper v4.0.

Table 1. Characteristics of 52 polymorphic SSRs developed from *Camellia oleifera*.

Unigene ID	Forward primer		Reverse primer		Repeat	Size (bp)	GenBank reference No.	Putative function
	Sequence (5'-3')	Ta (°C)	Sequence (5'-3')	Ta (°C)				
CoUg65	CAGGTTTGCTCCATTACTCG	55.8	GGTTTTTGGTTGCCTTTCTA	55	$(AG)_{11}$	246	AAT40532.1	Hydrolase, alpha/beta fold family protein
CoUg641	CCTTATCATCAAGCGGTTCT	54.6	GTTTGTCGCAGTGGTAGCAC	56.3	$(CT)_{11}$	232	NP_201297.1	COP1-interacting protein 8
CoUg3680	CAGCAGCCCAGAAATTAAA	54	GGGACGATGACGACGACAA	59.4	$(GA)_{13}$	139	XP_002302561.1	F-box family protein
CoUg4931	CTTCATCTCCGTTTGCTCT	52.3	GACTTTGCCTCCTTTTGTG	52.8	$(AG)_{16}$	142	-	-
CoUg5179	AATGGAGAATGAATGGACAG	51.9	GCAGAAAGTGATATTGGGTG	52.9	$(GA)_{11}$	199	XP_001758569.1	ArfRP1-family small GTPase
CoUg5452	CTCTGCCGTTTTTATTATAG	48.2	AGAGAGAGGGAAACAGGAC	51	$(TC)_{14}$	137	AAZ66745.1	Coronatine-insensitive 1
CoUg5585	TCCTTCGCCATTCTACTCCAT	58.9	AACCATCTCTAACCGCCCCTT	61.8	$(AG)_{14}$	100	-	-
CoUg6123	ATTCTTCTGTCTTTATCTCTCC	49.4	CTTCTTTGATTCCGATTTTG	52.8	$(CT)_{13}$	178	-	-
CoUg7330	ACAACCATCTTCCTCTCCCC	57.6	CGTCGTCCCTGTTCACCTCT	59.6	$(CT)_{11}$	146	NP_187304.1	Agenet domain-containing protein
CoUg7967	TGTGAATTTCTGACGAGGGTTGC	63.9	TTGGGGGTGGCTGTGATGGGGAC	73.4	$(TC)_{15}$	156	NP_196051.2	Pectate lyase family protein
CoUg8529	ATTCGCTTGCCTCCGCCACA	68.2	CCGATTCTCTACACTCCTTT	51.1	$(GA)_{11}$	256	NP_179311.1	Ubiquitin family protein
CoUg9308	TACGAATAGAGGAAAGAGAGACT	52.1	TCGATCCATTGCTACTAAGGT	54.4	$(GA)_{12}$	113	NP_001118683.1	Dual specificity protein phosphatase
CoUg10320	CGTATTAGCAGAAAGCACAC	51.2	ATTTGGGTTTAGAAAGAGGT	50.3	$(CT)_{12}$	208	-	-
CoUg11003	ACATAGAGGAGTCTAAAGGG	47.7	TAAAAAAGCAAGGAGGGAAG	54.6	$(TG)_{10}$	213	-	-
CoUg12783	GCACTTTGTGTGGATTTGAT	52.9	CGGAAGATTGTGGAGGTTTG	57.8	$(AG)_{10}$	206	-	-
CoUg13050	ACACACGAGAGAGAGGGAGA	53.8	CAATGTTGACGATCAATTGG	54.3	$(AG)_{13}$	220	-	-
CoUg13386	TTCGGCAGAGTGTGTATAG	48.6	GTTGAAGATGGAAGATGGTA	49.5	$(CT)_{10}$	193	-	-
CoUg14253	GTTTTGGTGGATCTGTGAAA	52.9	TAACAATTGGGCTCTTATGA	51	$(AG)_{10}$	263	Y3627_ARATH	VHS domain-containing protein
CoUg17053	TGTGGTTTTGGCGATTGGAT	61	GTTTTGGGGCTGTTCTGCTG	60.5	$(CT)_{10}$	191	NP_195531.1	FAR1-related sequence 5
CoUg17210	CAATTGTAGCATCGCATCTG	54.9	TTGGTTAGGCCATCAACGTC	58.3	$(TC)_{11}$	141	NP_001147153.1	TPR domain-containing protein
CoUg66187	ACATAAAAAATGGGGAGAGG	53.7	GCAAGGGTGTAGCAGAATC	51	$(GA)_{10}$	300	BAA95720.1	Uridine kinase-like protein
CoUg2348	CCACTCCGCATAGAAACGAC	58	CGAGCATACAACAAAACCTT	53.1	$(TCT)_{8}$	188	NP_179257.1	Nodulin family protein
CoUg3009	CAGCTCTCGATTGATAAGCG	56.2	AGCAATTCCACCACTGTGTG	56.1	$(CGT)_{10}$	158	XP_002515161.1	Conserved hypothetical protein
CoUg3402	ACTCTTGTGGGTGAATGTTG	52.6	GCTGGTAGGTTGGTTATGTT	51.9	$(GAT)_{8}$	205	CAE30325.1	NIN-like protein 2
CoUg4364	GTGGTCCTGGAGATCTGTCC	56	TTTCGCTCTATCCGTTGTTC	55.6	$(ATC)_{27}$	283	NP_187692.2	Zinc finger family protein
CoUg5820	CTGGTTGTTTTGAGTGAGAG	50	TGGTATGGAGAAGAGGAATG	52.2	$(GAA)_{13}$	180	ACF06508.1	Senescence-associated protein
CoUg9871	TGCCTACTGCTCACCTGCGT	61.5	CAATGGGCCATCATCTACT	52.7	$(TCT)_{8}$	214	-	-
CoUg10008	CTTTTGCTTCTGGCTCTGT	52.7	ATGTCGCTGGTGTGTCCTT	55.4	$(CAT)_{10}$	199	-	-
CoUg14786	CAACCCTACCCTTCTCTGC	54.4	CCTTAACTCTTCGTTCCCG	54.9	$(GGA)_{7}$	223	-	-
CoUg15419	AACCAACCGATTGCTGCGAT	62.7	CCTTCCTCTGCTTCTGTCCA	57.5	$(GAA)_{8}$	127	NP_973532.1	Protein kinase family protein
CoUg15473	CATTGAAGAGAGCAAAAACG	53.7	GGAGGCCCTTACAAGAAACT	55.4	$(CCT)_{9}$	199	NP_174563.2	Glycosyl hydrolase family 17 protein
CoUg15609	GATTGAGAGGAGATACATTG	46.2	GAAGAAGAAGAAGCAGAGG	48.6	$(CTT)_{9}$	152	XP_002329920.1	Chromatin remodeling complex subunit
CoUg16919	TTACTGGAGGAACTGATTTT	49	AACCTTATTGAGGACTATGT	45.4	$(TGA)_{13}$	176	-	-
CoUg65596	TATTCAAATGGGTCAAGTCC	52.3	CTTCAGCGATTAGAGATTCC	54.8	$(CTC)_{7}$	134	CAJ00011.1	Coiled-coil-helix-coiled-coil-helix domain-containing protein
CoUg5643	TCCCCAATTCTCTCTCTAC	48.3	CATCAACCATATCTCTATGC	47.5	$(ATAG)_{6}$	107	XP_002281109.1	UBX domain-containing protein
CoUg6766	TCACTGAGGTTTTAGGGCC	54.9	AAATCTTCCGAGGGAGGAG	55.6	$(ACAG)_{6}$	180	XP_002298244.1	PAF1 complex component
CoUg12112	GCCGCTACTTCTAACTTCA	50.6	ACTCCGACCCCATTTCTCT	55.6	$(CATA)_{7}$	137	AAX37335.1	UDP-glucuronic acid decarboxylase 2
CoUg12183	AACTTTCCACAACACCACC	52.3	TCTAAGCTAATCGAGCTGC	50.9	$(GCAG)_{6}$	139	AAT39308.2	V-type ATPase subunit family protein

Continued on next page

Table 1. Continued.

Unigene ID	Forward primer		Reverse primer		Repeat	Size (bp)	GenBank reference No.	Putative function
	Sequence (5'-3')	Ta (°C)	Sequence (5'-3')	Ta (°C)				
CoUg59126	ATTAGTGGGCAGGCATTCC	57.3	AGGCACAGTGGCACAGATT	55.6	$(AAAT)_5$	197	NP_191903.3	Cytokinin dehydrogenase
CoUg59697	CGGACCTCTTTTAGAGATC	48.5	GGCTCCTAATAGTCCAGTT	48.3	$(AGGA)_5$	164	NP_181884.1	FKBP-type peptidyl-prolyl cis-trans isomerase family protein
CoUg59762	CATCATCCTCTCTCTTCTT	45.1	ACGTGCATCTATGTCACAT	47.8	$(AACA)_5$	146	XP_002523659.1	Phospholipase A21
CoUg60001	ATCCCCCATATTTCTCAGC	53.8	TCCACCACCACTCCTCTTC	54.6	$(ATTG)_5$	139	NP_200782.1	UDP-galactose/UDP-glucose transporter-related protein
CoUg62529	CTTCCATTGAGGACTTCTA	46.7	GATATGTTGTCTTGGTTGA	44.8	$(TCAT)_5$	210	NP_172576.1	Hydrolase, alpha/beta fold family protein
CoUg68309	AGCAATAAACCCACCCCAT	56.7	AGGCAAAGCAAAAGCATCC	57.9	$(CTCA)_5$	137	-	-
CoUg68886	CTATCGGAAATCGACATGC	53.6	CGTTACTGACCAGTGAGCG	54.4	$(ACAA)_5$	134	NP_192945.2	DNA-binding family protein
CoUg3005	GATTCTTCTTCTTCCCCATT	52.6	GAGTCACTTCTTCATCCATA	46.4	$(TC)_7(TA)_8$	153	NP_565446.1	Metaxin-related
CoUg3184	CATAAATTCCGAACCTCCG	55.6	TAGAAATCGACTCCATAGC	47.3	$(CT)_7(AT)_7$	148	-	-
CoUg3222	AACAAAACCCACTCACACT	49.2	AGACGAGCTGACAAGCTAC	49.6	$(TC)_8(TA)_8$	131	-	-
CoUg3397	TAGGGTTTTCACGATGATGC	55.9	CTGTTGCTGGTGTTGCGGTT	61.7	$(CAG)_7(CAA)_7$	184	XP_001527553.1	Suppressor protein SEF1
CoUg6065	TCTTCCATTTCATTACCACC	52.3	CCCTCCTTCTCCTTCTTCT	52.3	$(AC)_6(AT)_8$	171	-	-
CoUg6948	GGTGGTTCTCTAGTGAGTTT	47.9	GCCCCTGATTTTTAGGTTAT	53.4	$(GC)_6(AC)_{12}(TC)_{11}$	187	-	-
CoUg7148	CAACTTCTCCAAACCGTTA	50.9	CTTCTTCAAATCTTCGTCAC	49.5	$(AG)_7(TG)_7$	132	AAQ82703.1	Potyviral capsid protein interacting protein 2b

Ta = annealing temperature.

Statistical analysis

Due to the putatively polyploid character of *C. oleifera*, statistical analysis entailed treating the SSRs as dominant markers. The data were scored in a binary format (presence as '1' and absence as '0'). The binary data were used to calculate genetic parameters using PopGene v1.32. The polymorphism information content (PIC) value for each SSR marker was calculated as described previously (Huang, 2013; Singh et al., 2013). For *C. chekangoleosa* (diploid) and *C. japonica* (diploid), statistical analysis treated the SSRs as codominant markers.

RESULTS

Characterization of microsatellites in the *C. oleifera* unigenes

Transcriptome sequencing of developing seeds of *C. oleifera* produced 69,798 unigenes, which were used to identify microsatellites with the MISA tool. A total of 6949 putative SSR motifs from 6042 SSR-containing unigenes were identified; that is, 9.9% of the unigenes contained at least one of the considered microsatellites. Dinucleotide repeats were the most common types of SSR, accounting for 64.8%, followed by trinucleotide (26.1%), hexanucleotide (5%), pentanucleotide (2.6%), and tetranucleotide (1.5%) repeats (Figure 1).

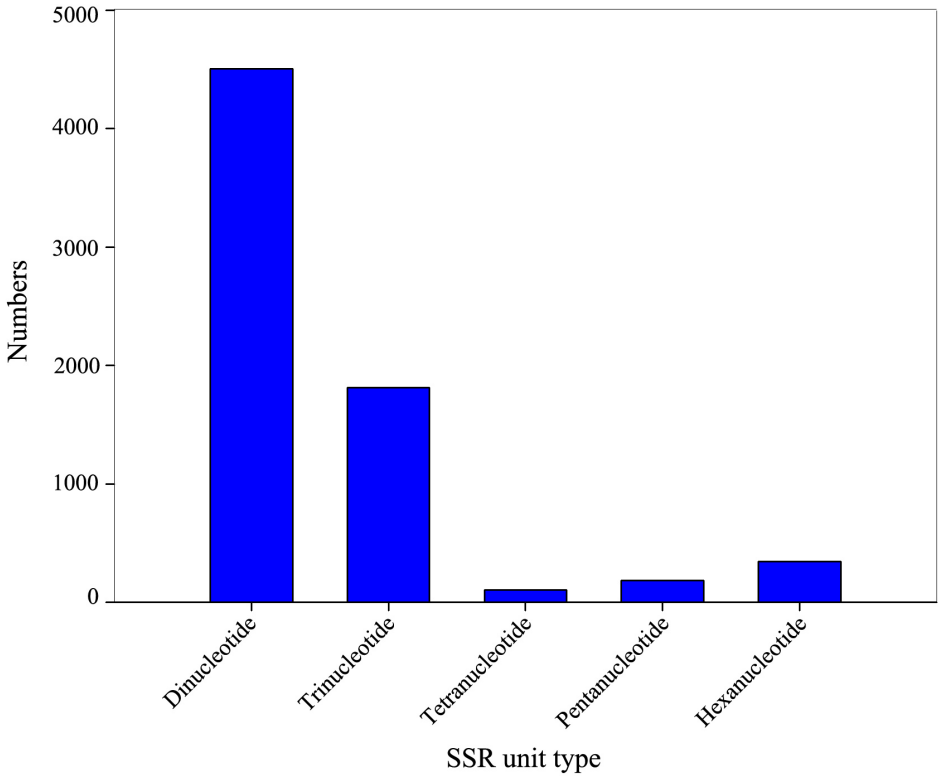

Figure 1. Distribution of different SSR unit types in *Camellia oleifera.*

The most abundant SSR motif was AG/CT, followed by AAG/CTT, AT/AT, AC/GT, ACC/GGT, ATC/ATG, AGG/CCT, AAC/GTT, and AGC/CTG (Figure 2). Among the dinucleotide repeats, the motif AG/CT was the most common (84.5%), followed by AT/TA (8.5%) and AC/GT (6.9%). Among the trinucleotide repeats, the motif AAG/CTT was the most common, accounting for 30.2%, followed by ACC/GGT (16.3%) and ATC/ATG (15.5%). Other motifs were identified in insignificant numbers.

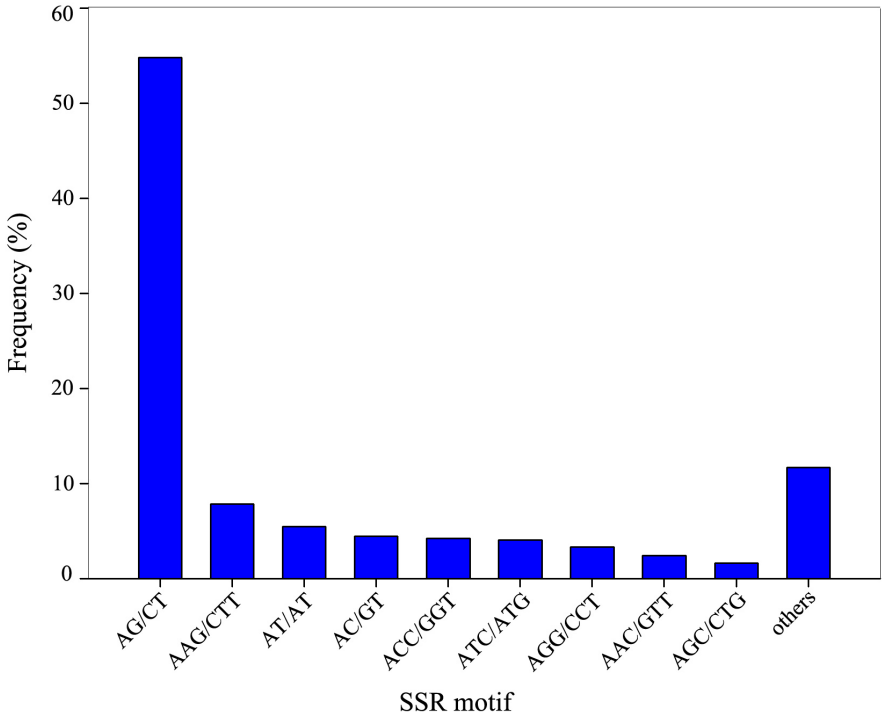

Figure 2. Frequency of different SSR motifs in *Camellia oleifera.*

Development of *C. oleifera* polymorphic SSRs

One hundred and fifty primer pairs, corresponding to 150 unigene loci, were designed for the PCR test. Initial amplification was conducted on three *C. oleifera* varieties with unlabeled primers. The amplified products were analyzed using agarose gel electrophoresis. Thirty-two loci generated nonspecific products and 16 loci yielded no products. The remaining 102 loci yielded products of expected size and were then assessed for polymorphism in 20 *C. oleifera* varieties with the M13 (-21) (5'-TGTAAAACGACGGCCAGT-3') sequence-tag method.

As shown by capillary electrophoresis, PCR products appearing as one band on an agarose gel were clearly separated into four distinct fragments by capillary electrophoresis (Figure 3A). Based on capillary electrophoresis data, 52 loci among the *C. oleifera* varieties displayed polymorphism tested (Table 1). The number of alleles (N_A) per locus ranged from 2 to 15, with an average of 7.096. The expected heterozygosity (H_E) value ranged from 0.32 to 0.897, with an average of 0.707. The PIC ranged from 0.498 to 0.887, with an average

of 0.742. Of the 52 loci, 49 were high polymorphic (PIC > 0.5) and only three (CoUg3898, CoUg12183, and CoUg14786) were moderately polymorphic (0.25 < PIC < 0.5) (Table 2).

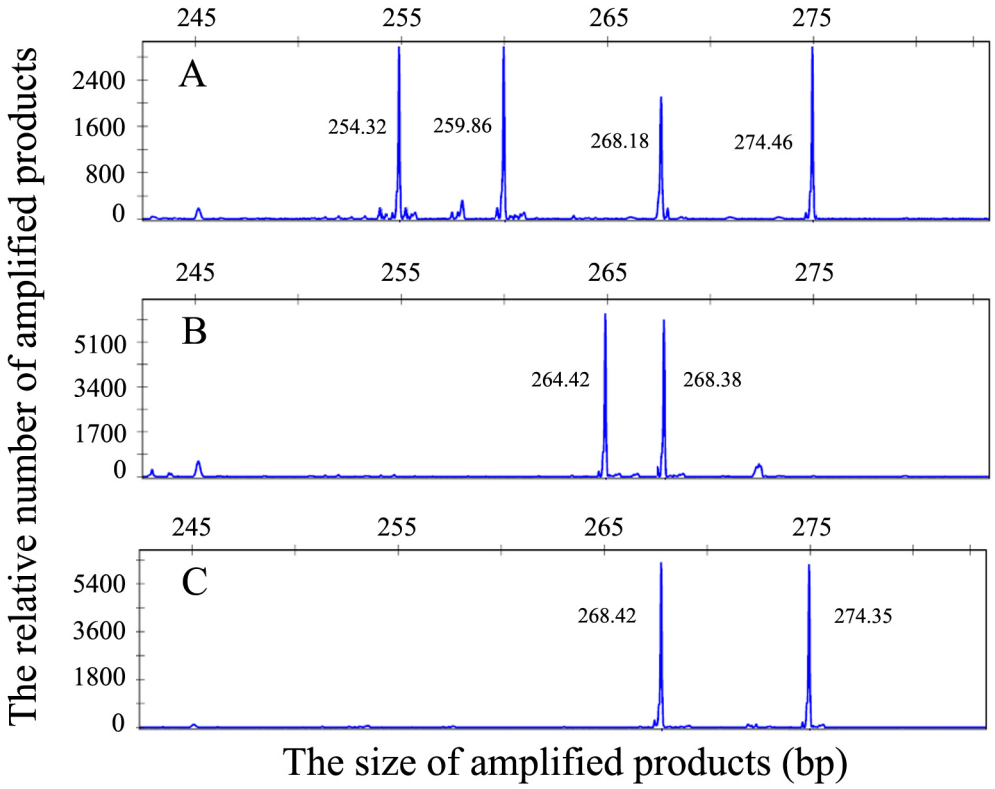

Figure 3. Analysis of fluorescently labeled PCR products by capillary electrophoresis.

Cross-species transferability of the *C. oleifera* SSRs

Cross-amplification tests were performed in 18 elite germplasm clones of *C. chekangoleosa* and 15 varieties of *C. japonica*, both of which are diploid species in *Camellia*. The SSR markers from *C. oleifera* produced at most two alleles in each sample of the two species (Figure 3B and C). For *C. chekangoleosa*, 47 SSR markers exhibited polymorphisms and the cross-species transferability rate was 90.4%. The N_A ranged from 2 to 10, with an average of 5.319 per locus. The H_E ranged from 0.056 to 0.873, with an average of 0.672. The PIC value ranged from 0.1 to 0.862, with an average of 0.67. Of the 47 markers, 42 were highly polymorphic, one was moderately polymorphic, and four had low polymorphism (PIC < 0.25) (Table 2).

For *C. japonica*, 41 markers exhibited polymorphisms and the cross-species transferability rate was 78.8%. The N_A ranged from 2 to 7, with an average of 4.61 per locus. The H_E ranged from 0.067 to 0.86, with an average of 0.652. The PIC value ranged from 0.117 to 0.831, with an average of 0.649. Of the 41 markers, 36 were highly polymorphic, one was moderately polymorphic, and four had low polymorphism (Table 2). These results indicated that the 52 SSR markers developed from *C. oleifera* harbored rich polymorphisms in the other *Camellia* plants.

Table 2. Genetic parameters generated among *Camellia oleifera* and its allied species using 52 unigene-derived SSR markers.

Unigene ID	*C. oleifera* (N = 20)				*C. chekangoleosa* (N = 18)				*C. japonica* (N = 15)			
	N_A	H_O	H_E	PIC	N_A	H_O	H_E	PIC	N_A	H_O	H_E	PIC
CoUg65	8	1.000	0.834	0.852	7	0.824	0.857	0.835	7	0.857	0.860	0.831
CoUg641	9	0.900	0.850	0.865	4	0.222	0.671	0.661	4	0.200	0.660	0.648
CoUg3680	6	0.611	0.657	0.702	6	0.500	0.765	0.765	6	0.467	0.766	0.760
CoUg4931	10	0.684	0.870	0.882	7	0.706	0.752	0.756	6	0.643	0.725	0.707
CoUg5179	14	0.950	0.881	0.890	10	0.833	0.851	0.826	-	-	-	-
CoUg5452	10	0.737	0.853	0.867	7	0.389	0.821	0.829	6	0.333	0.816	0.800
CoUg5585	10	1.000	0.839	0.856	8	0.722	0.803	0.803	-	-	-	-
CoUg6123	15	0.800	0.888	0.897	-	-	-	-	-	-	-	-
CoUg7330	8	0.700	0.796	0.819	4	0.111	0.686	0.695	4	0.133	0.745	0.734
CoUg7967	6	0.650	0.768	0.798	4	0.278	0.618	0.631	3	0.267	0.559	0.598
CoUg8529	8	0.900	0.843	0.859	8	0.941	0.863	0.837	-	-	-	-
CoUg9308	10	0.842	0.837	0.853	6	0.412	0.733	0.747	6	0.500	0.706	0.730
CoUg10320	8	0.611	0.827	0.846	5	0.333	0.684	0.719	5	0.400	0.743	0.748
CoUg11003	5	0.278	0.657	0.712	5	0.177	0.786	0.775	5	0.214	0.802	0.782
CoUg12783	12	0.900	0.852	0.866	7	0.444	0.829	0.825	7	0.467	0.825	0.826
CoUg13050	3	0.067	0.559	0.539	-	-	-	-	-	-	-	-
CoUg13386	5	0.842	0.669	0.714	4	0.333	0.722	0.691	4	0.267	0.747	0.709
CoUg14253	8	0.450	0.766	0.793	5	0.059	0.711	0.710	5	0.071	0.638	0.649
CoUg17053	7	0.500	0.761	0.791	5	0.177	0.640	0.665	4	0.143	0.564	0.602
CoUg17210	5	0.667	0.668	0.714	7	0.222	0.833	0.826	7	0.200	0.832	0.821
CoUg66187	3	0.150	0.492	0.552	6	0.529	0.763	0.754	5	0.357	0.738	0.737
CoUg2348	8	0.750	0.745	0.776	6	0.000	0.775	0.753	4	0.000	0.726	0.702
CoUg3009	5	0.900	0.705	0.746	7	0.833	0.754	0.749	7	0.800	0.759	0.752
CoUg3402	7	0.850	0.623	0.678	2	0.059	0.166	0.198	-	-	-	-
CoUg4364	11	1.000	0.859	0.872	4	0.059	0.668	0.570	4	0.071	0.664	0.658
CoUg5820	9	1.000	0.813	0.833	-	-	-	-	-	-	-	-
CoUg9871	7	0.900	0.732	0.767	8	0.889	0.813	0.792	-	-	-	-
CoUg10008	9	0.900	0.780	0.801	5	0.111	0.752	0.740	5	0.067	0.738	0.719
CoUg14786	2	0.053	0.269	0.320	3	0.000	0.585	0.567	3	0.000	0.540	0.520
CoUg15419	3	0.706	0.418	0.523	2	0.167	0.157	0.245	2	0.133	0.129	0.208
CoUg15473	4	0.895	0.522	0.603	3	0.118	0.116	0.194	2	0.143	0.138	0.219
CoUg15609	4	0.800	0.585	0.651	2	0.056	0.056	0.100	2	0.067	0.067	0.117
CoUg16919	8	0.737	0.791	0.815	4	0.235	0.683	0.653	4	0.214	0.691	0.657
CoUg65596	9	0.789	0.853	0.867	6	0.529	0.784	0.763	6	0.643	0.794	0.764
CoUg5643	6	1.000	0.750	0.785	4	0.000	0.734	0.713	4	0.000	0.751	0.724
CoUg3898	3	0.177	0.415	0.445	-	-	-	-	-	-	-	-
CoUg12112	7	0.800	0.744	0.778	5	0.118	0.562	0.565	4	0.000	0.561	0.541
CoUg12183	3	0.111	0.298	0.34	-	-	-	-	-	-	-	-
CoUg59126	6	0.947	0.626	0.683	4	0.278	0.614	0.631	4	0.333	0.641	0.650
CoUg59697	10	1.000	0.811	0.832	4	0.389	0.467	0.509	3	0.286	0.442	0.475
CoUg59762	5	0.923	0.607	0.667	5	0.722	0.759	0.747	5	0.714	0.757	0.747
CoUg60001	4	0.900	0.472	0.566	4	0.056	0.427	0.432	3	0.067	0.301	0.320
CoUg62529	7	1.000	0.745	0.779	4	0.294	0.629	0.628	3	0.143	0.601	0.586
CoUg68309	3	0.353	0.549	0.628	4	0.765	0.758	0.711	4	0.714	0.773	0.726
CoUg68886	4	0.579	0.607	0.654	3	0.111	0.560	0.580	3	0.067	0.549	0.555
CoUg3005	12	0.722	0.882	0.891	10	0.556	0.865	0.862	-	-	-	-
CoUg3184	5	0.474	0.739	0.775	5	0.222	0.751	0.731	5	0.286	0.759	0.728
CoUg3222	11	0.944	0.878	0.888	7	0.375	0.802	0.818	7	0.385	0.791	0.809
CoUg3397	11	0.750	0.856	0.870	8	1.000	0.873	0.849	7	0.867	0.851	0.821
CoUg6065	3	0.158	0.580	0.653	4	0.000	0.535	0.519	3	0.000	0.519	0.500
CoUg6948	7	0.368	0.691	0.728	7	0.529	0.795	0.757	6	0.500	0.712	0.689
CoUg7148	6	0.526	0.680	0.718	5	0.353	0.779	0.767	5	0.357	0.775	0.759
Mean	7.096	0.697	0.707	0.742	5.319	0.362	0.672	0.670	4.610	0.302	0.652	0.649

N = number of individuals tested; N_A = number of alleles; H_O = observed heterozygosity; H_E = expected heterozygosity. PIC = polymorphism information content.

Putative function of the SSR-containing unigenes

To determine the function of polymorphic SSR-associated unigenes, the newly developed SSRs were evaluated for associations with genes of known function. The 52 sequences were BLASTed against the GenBank nonredundant database using BLASTx with *E* value $< 1 \times 10^{-5}$. Of the 52 unigenes, 35 showed significant similarities to known genes, including those for an alpha/beta hydrolase fold family protein (CoUg65), an F-box family protein (CoUg3680), a pectate lyase family protein (CoUg7967), a zinc finger family protein (CoUg4364), UDP-galactose/UDP-glucose transporter-related protein (CoUg60001), and others listed in Table 1.

DISCUSSION

In this study we discovered 6949 SSR-motifs with 2-6 nucleotide repeats from 6042 SSR-containing unique putative transcripts among the 69,798 unigenes in *C. oleifera*. We found that the most abundant SSR motifs in this tree species identified in this study were AG/CT and AAG/CTT. A similar bias towards AG and AAG, and against CG repeats, has been reported in EST-SSRs of other plant species (Blanca et al., 2011; Xu et al., 2012; Zhang et al., 2012). According to Gonzalez-Ibeas et al. (2007), this may have resulted from the tendency of CpG sequences to be methylated, which may potentially inhibit transcription.

We developed 52 polymorphic SSR markers in which all SSR motifs contained 20 or more nucleotides. Of these markers, 47 and 41 can be transferable to two allied species of *C. oleifera*, *C. chekangoleosa* and *C. japonica*, respectively. In addition, the 52 SSR markers potentially encoded functional genes since they were developed from unigenes. Genbank database search identified 35 of the 52 loci putatively coding for functional proteins, therefore these genes may be correlated with seed development in *C. oleifera*. In comparison with genomic SSR markers, these unigene-derived SSR markers have special features because they are associated with functional genes and may increase the efficiency of marker-assisted selection (Gupta and Rustgi, 2004). These 52 informative unigene-derived SSR markers will be valuable for analyses of genetic variation and marker-assisted selection in breeding programs for *C. oleifera*, *C. chekangoleosa*, and *C. japonica*.

ACKNOWLEDGMENTS

Research supported by the National Natural Science Foundation of China (Grant #31100497) and the Key Project of the Chinese Ministry of Education (Grant #212126). We thank Dr. Xiaolin Lei of the Jiangxi Academy of Forestry for plant sample collection.

REFERENCES

Blanca J, Cañizares J, Roig C, Ziarsolo P, et al. (2011). Transcriptome characterization and high throughput SSRs and SNPs discovery in *Cucurbita pepo* (Cucurbitaceae). *BMC Genomics* 12: 104.

Dutta S, Kumawat G, Singh BP, Gupta DK, et al. (2011). Development of genic-SSR markers by deep transcriptome sequencing in pigeon pea [*Cajanus cajan* (L.) Millspaugh]. *BMC Plant Biol.* 11: 17.

Gonzalez-Ibeas D, Blanca J, Roig C, González-To M, et al. (2007). MELOGEN: an EST database for melon functional genomics. *BMC Genomics* 8: 306.

Gupta PK and Rustgi S (2004). Molecular markers from the transcribed/expressed region of the genome in higher plants. *Funct. Integr. Genomics* 4: 139-162.

Gupta SK and Gopalakrishna T (2010). Development of unigene-derived SSR markers in cowpea (*Vigna unguiculata*) and their transferability to other *Vigna* species. *Genome* 53: 508-523.

Huang Y (2013). Population genetic structure and interspecific introgressive hybridization between *Camellia meiocarpa* and *C. oleifera. J. Appl. Ecol.* 24: 2345-2352.

Ma JL, Ye H, Rui YK, Chen GC, et al. (2011). Fatty acid composition of *Camellia oleifera* oil. *J. Consum. Prot. Food Safety* 6: 9-12.

Marioni JC, Mason CE, Mane SM, Stephens M, et al. (2008). RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res.* 18: 1509-1517.

Mortazavi A, Williams BA, Mccue K, Schaeffer L, et al. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* 5: 621-628.

Nagalakshmi U, Wang Z, Waern K, Shou C, et al. (2008). The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* 320: 1344-1349.

Schuelke M (2000). An economic method for the fluorescent labeling of PCR fragments. *Nat. Biotechnol* 18: 233-234.

Shao FG (2011). Research on Transcriptome and Gene Expression Profiles involved in Lipid Biosynthesis and Regulation in the Developing Seeds of *Camellia oleifera*. Doctoral thesis, Central South University of Forestry and Technology, Changsha.

Shi J, Dai XG, Chen YG, Chen JH, et al. (2013). Discovery and experimental analysis of microsatellites in an oil woody plant *Camellia chekiangoleosa*. *Plant Syst. Evol.* 299: 1387-1393.

Singh RK, Jena SN, Khan S, Yadav S, et al. (2013). Development, cross-species/genera transferability of novel EST-SSR markers and their utility in revealing population structure and genetic diversity in sugarcane. *Gene* 524: 309-329.

Wen Q, Xu LC, Gu YC, Huang MR, et al. (2012). Development of polymorphic microsatellite markers in *Camellia chekiangoleosa* (Theaceae) using 454-ESTs. *Am. J. Bot.* 99: e203-205.

Xu W, Yang Q, Huai HY and Liu AZ (2012). Development of EST-SSR markers and investigation of genetic relatedness in tung tree. *Tree Genet. Genomes* 8: 933-940.

Zhang JN, Liang S, Duan JL, Wang J, et al. (2012). *De novo* assembly and characterisation of the transcriptome during seed development, and generation of genic-SSR markers in peanut (*Arachis hypogaea* L.). *BMC Genomics* 13: 90.