ISSN 1479-2621
Plant Genetic Resources: Characterization and Utilization (2019) **17**(4); 371–374
doi:10.1017/S1479262119000066

Short Communication

Characterization and cross-species transferability of a novel set of microsatellites derived from root transcriptomes of *Camellia oleifera*

Shuangcheng Wu[1], Hang Ye[2], Yuansong Chen[3], Jiemei Deng[1], Jiexia Su[1], Yayu Xie[1], Qinru Xie[1], Zhaoyuan Zhang[2], Zihai Qin[2], Yufei Xiao[2], Xiaoyun Wang[4]* and Pengliang Wang[1,2]*

[1]*Guangxi Key Laboratory of Beibu Gulf Marine Biodiversity Conservation, Qinzhou University, Qinzhou Guangxi, China,* [2]*Guangxi Key Laboratory of Special Non-wood Forest Cultivation and Utilization, Guangxi Forestry Research Institute, Nanning Guangxi, China,* [3]*Guangxi State-owned Sanmenjiang Forestry Farm, Liuzhou Guangxi, China and* [4]*Collaborative Innovation Center for the Modern Technology and Industrial Development of Jiangxi Minority Traditional Medicine, Jiangxi University of Traditional Chinese Medicine, Nanchang Jiangxi, China*

Received 2 September 2018; Accepted 11 January 2019 – First published online 21 February 2019

Abstract

Camellia oleifera is an important woody plant producing healthy edible oils. People need a large number of molecular markers, especially microsatellite, in breeding of *C. oleifera*. In this study, we sequenced the root transcriptomes of *C. oleifera*, and then designed a novel set of microsatellite markers based on the root-expressed genes. We assembled a total of 57,121 unigenes with a length of 42.63 Mb, which harboured 15,902 microsatellites. Among these microsatellites, di-nucleotide repeat motifs were the most abundant group (56.45%), then followed by tri- (25.20%), mono- (12.12%), hexa- (3.21%), penta- (2.18%) and quad-nucleotide ones (0.84%). In total, 6738 primer pairs were designed successfully to amplify the microsatellite loci. To test these microsatellite markers, 48 primer pairs were randomly selected and synthesized and validated in *C. oleifera* and its eight relatives. Up to 75% of the primer pairs amplified in *C. oleifera* and its relatives, and 62.5% displayed polymorphism. The transferability and diverse alleles across its eight relatives were detected for each polymorphic primer pair. The novel set of microsatellites derived from the root transcriptomes here provided a useful resource for future molecular genetics improvement of *C. oleifera* and its relatives.

Keywords: *Camellia oleifera*, microsatellite, root transcriptome, transferability

Introduction

Camellia oleifera is one of the most important healthy oil-bearing tree species. It has the highest total yields and acreage in China (Zhuang, 2008). It is very important to increase the yield of *C. oleifera* to meet Chinese people's increasing requirement for edible oils. Variety improvement through molecular breeding, a promising method to breed elite varieties with precision and efficiency (Philips, 2006; Xu, 2010), is the key factor to increase yields (Chen *et al.*, 2005; Zhuang, 2010; Shi *et al.*, 2011; Tan *et al.*, 2012).

*Corresponding author. E-mail: pengliang_wang@163.com; wxy20052002@aliyun.com

Microsatellite is a kind of codominant, multi-allelic, even covering and high genetic informational marker (Powell *et al.*, 1996; Varshney *et al.*, 2005a), therefore is very useful in molecular breeding (Chen *et al.*, 2016; Wang *et al.*, 2017). Microsatellite markers of *C. oleifera* have been developed from the transcriptomes of several tissues, e.g. shoots, leaves, buds and flowers, through 454 GS-FLX platform (Xia *et al.*, 2014) or Illumina HiSeq 2000 platform (Chen *et al.*, 2017). However, these microsatellite sites were not validated by experiments. Microsatellites in the transcriptomes of leaves and seeds were also detected, then used in genetic diversity (Li *et al.*, 2017) and transferability of these microsatellites across two relatives were evaluated (Jia *et al.*, 2014, 2015).

The formation of different tissues was determined by differential expression and interaction of large numbers of genes (Slovak *et al.*, 2016; Drapek *et al.*, 2017). The deficiency of markers from root transcriptomes may limit the molecular breeding of root-related traits. Hence, here we developed a novel set of microsatellites from root transcriptomes of *C. oleifera*, and then reported the transferability of these markers to its relatives in genus *Camellia*.

Experimental

The root samples were collected from tissue cultured plants ('Cenruan No.3') at the three key time points, i.e. the second day after shoot induced in rooting medium (T1), the 14th day after induced (root point appearance) (T2) and the 50th day after induced (root growth) (T3). The RNA extraction, cDNA library construction and Illumina paired-end sequencing were carried out. After removing low-quality reads, adaptors, possible contaminations and ribosomal RNA sequences, the clean reads were assembled into unigenes using Trinity (Grabherr *et al.*, 2011). Microsatellites were identified in the assembled unigenes using MISA script (Thiel *et al.*, 2003) against the criteria as follows: a minimum of 12, 6, 5, 5, 4 and 4 times of repeats for 1–6 nucleotide motifs, respectively. Primer pairs covering the microsatellites were designed by Primer 3. Forty-eight primer pairs (online Supplementary Table S1) were randomly selected and synthesized for validation and application.

A total of 18 accessions (Table 1) from *C. oleifera* (three accessions) and its eight relatives (15 accessions) were sampled for validation and transferability. Their genomic DNAs were extracted respectively with a modified CTAB method. SSR-PCRs were performed in a volume of 10 µL containing 20 ng of DNA, 0.25 µmol/L of each primer, 5 µL 2 × Taq PCR MasterMix and 2 µL ddH_2O. The SSR-PCRs were 94°C for 4 min; 32 cycles of 94°C for 45 s, 60°C for 45 s and a 72°C extension for 45 s; followed by a final extension at 72°C for 7 min. The PCR products were separated by 6.0% non-denatured polyacrylamide gels, visualized via silver staining according to the previous reports (Wang *et al.*, 2014, 2017).

Table 1. The accessions of *Camellia* species used in this study

Species	Accessions	Number of accessions
Camellia oleifera	PTYC, CL23, CL04	3
Camellia chrysantha	JHC03, JHC02	2
Camellia gigantocarpa	BBDGYC	1
Camellia japonica	RBHSC	1
Camellia meiocarpa	XGYC	1
Camellia nanyongensis	NRYC	1
Camellia osmantha	XHYC, YA02	2
Camellia polyodonta	WTHHYC	1
Camellia vietnamensis	NZSS, NZ01, GS01, HG01, ND06, LCDGYC	6

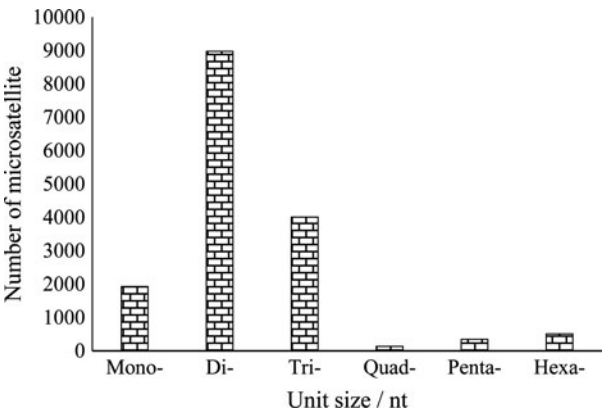

Fig. 1. Frequencies of microsatellites in *Camellia oleifera* root transcriptomes according to unit size.

The genetic diversity parameters including the number of alleles (Na), the effective number of alleles (Ne), observed heterozygosity (Ho), expected heterozygosity (He) and Shannon's index (*I*) were calculated via GenAlEx version 6.5 (Peakall and Smouse, 2012).

Discussion

A total of 58,656,354 raw reads were generated from the root transcriptomes, and 55,072,968 clean reads were

https://doi.org/10.1017/S1479262119000066 Published online by Cambridge University Press

Table 2. Genetic diversity parameters of *Camellia* samples based on microsatellites

Locus	*N*	Na	Ne	*I*	Ho	He
CoSSR1	15	5	3.81	1.43	1.00	0.74
CoSSR2	12	5	2.80	1.20	0.58	0.64
CoSSR3	15	7	4.21	1.65	0.60	0.76
CoSSR4	9	7	5.79	1.83	0.44	0.83
CoSSR5	17	3	1.61	0.68	0.47	0.38
CoSSR6	10	6	3.45	1.47	0.40	0.71
CoSSR10	18	2	1.80	0.64	0.67	0.44
CoSSR11	16	4	2.26	0.92	0.75	0.56
CoSSR12	17	3	2.25	0.91	0.76	0.56
CoSSR14	15	3	2.41	0.95	0.67	0.58
CoSSR16	15	2	1.14	0.24	0.00	0.12
CoSSR21	13	3	2.89	1.08	0.46	0.65
CoSSR23	14	4	2.47	1.08	0.57	0.59
CoSSR25	14	3	2.90	1.08	0.71	0.66
CoSSR26	14	4	2.14	0.99	0.50	0.53
CoSSR27	14	3	2.90	1.08	0.79	0.66
CoSSR28	7	5	4.26	1.51	0.14	0.77
CoSSR29	6	6	4.80	1.68	0.33	0.79
CoSSR30	14	4	3.19	1.27	0.71	0.69
CoSSR32	14	3	2.56	1.00	0.79	0.61
CoSSR34	14	3	2.67	1.04	0.50	0.63
CoSSR37	9	5	3.12	1.35	0.67	0.68
CoSSR38	7	6	4.08	1.57	0.71	0.76
CoSSR39	10	5	4.35	1.54	0.20	0.77
CoSSR40	9	6	2.84	1.38	0.67	0.65
CoSSR41	9	6	3.95	1.54	0.56	0.75
CoSSR42	10	3	2.38	0.94	0.40	0.58
CoSSR45	10	3	2.25	0.94	0.40	0.56
CoSSR46	10	2	1.22	0.33	0.20	0.18
CoSSR50	4	2	1.28	0.38	0.25	0.22

N, number of samples; Na, number of alleles; Ne, number of effective alleles; Ho, observed heterozygosity; He, expected heterozygosity; *I*, Shannon's index.

assembled into 57,121 unigenes with a total length of 42,633,606 nt, and the N50 of 1174 nt.

A total of 12787 sequences with 15,902 microsatellites were identified from 57,121 unigenes. Di-nucleotide repeats were the most abundant group (56.45%), then followed tri- (25.20%), mono- (12.12%), hexa- (3.21%), penta- (2.18%) and quad-nucleotide (0.84%) (Fig. 1). The microsatellite-containing sequence percentage in this study was higher than that (9–10%) of transcriptomes from many other tissues reported previously (Xia *et al.*, 2014; Jia *et al.*, 2015; Chen *et al.*, 2017) and slightly lower than that (26.75%) of SSR developed from leaf transcriptomes (Li *et al.*, 2017). Although the tissue and identification criteria were the same, the proportions of microsatellite-containing sequences in transcriptomes were different. As to repeat motif frequencies, our results agreed to previous reports that the most frequent motif was di-nucleotides, followed by tri-nucleotides, and other motif frequencies were lower than 5% (Xia *et al.*, 2014; Jia *et al.*, 2015; Li *et al.*, 2017). However, percentages of motif frequencies and microsatellite-containing sequences were different, perhaps because of the factors such as different materials, dataset size and detection criteria (Varshney *et al.*, 2005b).

We then designed 6738 primer pairs. Forty-eight primer pairs were randomly selected and synthesized. Of the synthesized primers, 12 (25%) could not produce any amplicons in all species, and six (12.5%) were monomorphic. The remaining (62.5%) were codominant and polymorphic.

Each primer had different transferring ability across *Camellia* relatives (online Supplementary Table S1). There were 15 primers with transferring abilities of more than 75%, and only two pairs were <50%. We found that only one pair (CoSSR10) amplified across all relative species.

The relationship among the species in this study was also suggested by the transferability (online Supplementary Table S1). These primers had the highest proportion amplified in *Camellia vietnamensis* (96.67%), and the lowest one in *Camellia nanyongensis* and *Camellia polyodonta* (10%). This suggested that *C. vietnamensis* was the closest relative to *C. oleifera*, and *C. nanyongensis* and *C. polyodonta* seemed to be the most distant species from *C. oleifera*.

Genetic diversity of all samples was revealed (Table 2). The average of Na and Ne was 4.10 and 2.93. The average of Ho and He was 0.53 and 0.60. *I* ranged from 0.24 to 1.83, with an average of 1.12.

The novel set of microsatellites here provided a useful tool for molecular genetic improvement in *C. oleifera* and its relatives.

Supplementary material

The supplementary material for this article can be found at https://doi.org/10.1017/S1479262119000066

Acknowledgements

This work was funded by the National Natural Science Foundation of China (31460208), Project of Science and Technology of Guangxi (GuiKeAB16380143), Guangxi Innovation-driven Project (GuiKeAA17204058) and the

Project of Science and Technology of Qinzhou (201616804, 20137003).

References

Chen Y, Yang X, Peng S, Li D, Wang X and Duan W (2005) Research status and developmental strategy of selection and breeding of elitle in *Camellia* in China. *Forestry Science and Technology* 19: 1–4.

Chen Y, Dai X, Hou J, Guan H, Wang Y, Li Y and Yin T (2016) DNA fingerprinting of oil camellia cultivars with SSR markers. *Tree Genetics & Genomes* 12: 1–8.

Chen J, Yang X, Huang X, Duan S, Long C and Rong J (2017) Leaf transcriptome analysis of a subtropical evergreen broadleaf plant, wild oil-tea camellia (*Camellia oleifera*), revealing candidate genes for cold acclimation. *BMC Genomics* 18: 211.

Drapek C, Sparks EE and Benfey PN (2017) Uncovering gene regulatory networks controlling plant cell differentiation. *Trends in Genetics* 33: 529–539.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N and Regev A (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* 29: 644–652.

Jia B, Lin Q, Zhang L, Tan X, Lei X, Hu X and Shao F (2014) Development of 15 Genic-SSR markers in oil-tea tree (*Camellia oleifera*) based on transcriptome sequencing. *Genetika* 46: 789–797.

Jia BG, Lin Q, Feng YZ, Hu XY, Tan XF, Shao FG and Zhang L (2015) Development and cross-species transferability of unigene-derived microsatellite markers in an edible oil woody plant, *Camellia oleifera* (Theaceae). *Genetics and Molecular Research* 14: 6906–6916.

Li H, Wang S, Ding H, Chen Y, Hu C, Li N and Wei H (2017) Development of EST-SSR molecular markers based on transcriptome sequencing of *Camellia oleifera*. *Plant Physiology Journal* 53: 1267–1278.

Peakall R and Smouse PE (2012) Genalex 6.5: genetic analysis in Excel. Population genetic software for teaching and research – an update. *Bioinformatics* 28: 2537–2539.

Philips RL (2006) Genetic tools from nature and the nature of genetic tools. *Crop Science* 46: 2245–2252.

Powell W, Machray GC and Provan J (1996) Polymorphism revealed by simple sequence repeats. *Trends Plant Science* 1: 215–222.

Shi J, Yin T and Shi J (2011) Review on development of *Camellia* spp. Related industry in China. *Journal of Southwest Forestry University* 31: 82–87.

Slovak R, Ogura T, Satbhai SB, Ristova D and Busch W (2016) Genetic control of root growth: from genes to networks. *Annals of Botany* 117: 9–24.

Tan X, Ma L, Li F, Yao X, Dong P, Su S, Mao Y, Li J and Yuan D (2012) Industrialization development strategy of woody grain and oil in China. *Nonwood Forest Research* 30: 1–5.

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

Varshney RK, Graner A and Sorrells ME (2005a) Genic microsatellite markers in plants: features and applications. *Trends in Biotechnology* 23: 48–55.

Varshney RK, Sigmund R, Börner A, Korzun V, Stein N, Sorrells ME, Langridge P and Graner A (2005b) Interspecific transferability and comparative mapping of barley EST-SSR markers in wheat, rye and rice. *Plant Sci* 168: 195–202.

Wang P, Wang H, Zheng Y, Lv Z, Liu J and Wang X (2014) Genetic mapping and QTL analysis for seed yield, vegetative characters and cold tolerance in centipedegrass. *Scientia Horticulturae* 176: 97–104.

Wang P, Yang L, Zhang E, Qin Z, Wang H, Liao Y, Wang X and Gao L (2017) Characterization and development of EST-SSR markers from a cold-stressed transcriptome of centipedegrass by Illumina paired-end sequencing. *Plant Molecular Biology Reporter* 35: 215–223.

Xia EH, Jiang JJ, Huang H, Zhang LP, Zhang HB and Gao LZ (2014) Transcriptome analysis of the oil-rich tea plant, *Camellia oleifera*, reveals candidate genes related to lipid metabolism. *PLoS ONE* 9: e104150.

Xu Y (2010) *Molecular Plant Breeding*. Wallingford: CAB International.

Zhuang R (2008) *Oil-Tea Camellia in China*, 2nd edn. Beijing: China Forestry Publishing House.

Zhuang R (2010) Historical perspective and prospect of selection and breeding of elitle in Camellia in China. *Forestry ScIence and Technology* 24: 1–5.