Genetics and Molecular Research

Isolation and characterization of microsatellite markers for *Eucommia ulmoides* (Eucommiaceae), an endangered tree, using next-generation sequencing

W.R. Zhang[1]*, Y. Li[2]*, J. Zhao[1], C.H. Wu[1], S. Ye[1] and W.J. Yuan[1]

[1]Institute of Chinese Materia Medica, Henan University, Kaifeng, China
[2]College of Forestry, Henan Agricultural University, Zhengzhou, China

*These authors contributed equally to this study.
Corresponding author: W.J. Yuan
E-mail: lyhenau@yeah.net

Genet. Mol. Res. 15 (2): gmr.15027789
Received October 6, 2015
Accepted December 7, 2015
Published April 25, 2016
DOI http://dx.doi.org/10.4238/gmr.15027789

ABSTRACT. *Eucommia ulmoides* Oliver, a single extant species of Eucommiaceae, is an endemic dioecious tree in China. The natural resources of *E. ulmoides* have rapidly declined in recent years because of the over-collection of its cortex. To design a suitable protection strategy, it is necessary to develop a set of molecular markers to investigate genetic diversity and population structure of *E. ulmoides*. Pyrosequencing of an enriched microsatellite library by Roche 454 FLX+ platform was used to isolate simple sequence repeats (SSRs) for *E. ulmoides*. A total of 1568 SSRs that contained enough flanking sequences for primer pair design were identified from 45,236 raw sequence reads. One hundred SSRs were randomly selected to design primer pairs and polymerase chain reaction was performed. Among these 100 tested primer pairs, 16 were polymorphic across 18 individuals from three *E. ulmoides* populations. The number of alleles ranged from 3 to 8, with an average of 5.1. The expected heterozygosity ranged from 0.110 to 0.830, with

an average of 0.648, and the observed heterozygosity ranged from 0.111 to 0.833, with an average of 0.524. The inbreeding coefficient ranged from -0.349 to 0.547. This set of microsatellite markers could be valuable for landscape genetic structure assessment and molecular marker-assisted breeding in *E. ulmoides*.

Key words: Simple sequence repeat; Genetic diversity; Population structure; *Eucommia ulmoides*

INTRODUCTION

Eucommia ulmoides Oliver, a single extant species of Eucommiaceae, is an endemic dioecious tree in China. Its cortex has been used in traditional Chinese medicine for over 2000 years. Previous studies revealed that the pharmacological properties of its leaves are similar to those of its cortex (Hussain et al., 2003; Chen et al., 2010). It is pharmacologically used to treat lumbago and hypertension, prevent bone senescence, and enhance liver and kidney functions. The cortex is also used as a raw material in the chemical industry for the gutta-percha production. For these reasons, the cortex has been over-harvested for its high medical and industrial value. *E. ulmoides* is currently on the red list of endangered plant species of China (Yao et al., 2012). To design a suitable protection strategy, it is necessary to develop a set of molecular markers to investigate genetic diversity and population structure of *E. ulmoides*. Although, molecular markers such as random amplified polymorphic DNA (RAPD; Wang et al., 2006) and amplified fragment length polymorphism (AFLP; Yao et al., 2012) have been used to investigate the genetic diversity of *E. ulmoides*, simple sequence repeats (SSRs) have been considered more reliable markers owing to their high reproducibility (Li et al., 2002; Wei et al., 2013). Deng et al. (2006) first reported on microsatellite markers of *E. ulmoides* using the traditional fast isolation by AFLP sequences containing repeats protocol. However, a limited number of markers do not meet the requirements of a survey of landscape genetic structure (Hall and Beissinger, 2014). Meanwhile, because of its endangered status, artificial cultivation of *E. ulmoides* has been started in China. This requires higher molecular marker density to help future variety breeding.

With the development of technology, next-generation sequencing technique has widely been used for SSR isolation (Fu et al., 2014; Li et al., 2015). Next-generation sequencing can be used to obtain a large number of markers in a short time. In the present study, we report a large number of SSRs for *E. ulmoides*, obtained using a Roche 454 FLX+ platform.

MATERIAL AND METHODS

Plant material and DNA extraction

Fresh leaves of *E. ulmoides* were collected from 18 individuals belonging to three natural populations (Xinyang: 31.911°N, 114.110°E; Nanyang: 33.619°N, 111.737°E; Luoyang: 33.773°N, 111.627°E) from Henan Province of China. Vouchers were deposited in the herbarium of the Institute of Chinese Materia Medica at Henan University (Herbarium accession No. HCMM-101-118). Total genomic DNA was extracted from fresh leaves using a plant genomic DNA extraction kit DP305 (Tiangen Biotech, Beijing, China) according to the manufacturer protocol.

DNA sequencing and microsatellite locus searching

Approximately 2 μg genomic DNA was used to construct a DNA library according to the Roche 454 GS-FLX+ library preparation protocol. Hybridization of eight biotinylated oligonucleotides [$(AG)_{10}$, $(AC)_{10}$, $(AAC)_8$, $(ACG)_8$, $(AAG)_8$, $(ACAT)_6$, $(ATCT)_6$, and $(AGG)_8$] with the DNA library was performed to enrich the repetitive motifs. The enriched products were then sequenced using Roche 454 FLX+ platform by Personalbio (Shanghai, China). One-twelfth run was performed on the sequencing platform. The program, MISA (Thiel et al., 2003), was used to search for SSR loci with the default settings and Primer3 (Rozen and Skaletsky, 2000) was used to design primer pairs keeping the default settings.

Polymerase chain reaction (PCR) and genotyping

PCRs were performed in a 20-μL reactions that contained 50 ng genomic DNA, 0.5 μM each primer, and 10 μL 2X Taq PCR MasterMix (0.1 U Taq polymerase/μL, 0.5 mM dNTPs, 20 mM Tris-HCl, pH 8.3) using an S1000 Thermal cycler (Bio-Rad, USA). PCR conditions were as follows: pre-incubation at 94°C for 5 min; followed by 35 cycles of denaturation at 94°C for 50 s; annealing at a locus-specific temperature for 45 s; elongation at 72°C for 50 s; a final extension step at 72°C for 8 min. PCR products were tested for polymorphism using 8% native polyacrylamide gels and visualized by silver nitrate staining.

Data analysis

Electropherograms for the fragment analysis were analyzed using the GeneMarker software v1.95 (SoftGenetics, State College, PA, USA). The number of alleles per locus (N_A), observed heterozygosity (H_O), expected heterozygosity (H_E), and deviations from Hardy-Weinberg equilibrium (HWE) were calculated with Popgene v1.32 (Yeh et al., 2000). Linkage disequilibrium (LD) between the pairs of SSR loci were calculated using GENEPOP v4.2 (Rousset, 2008). Inbreeding coefficients (F_{IS}) were calculated by FSTAT v2.9.3.2 (Goudet, 1995).

RESULTS AND DISCUSSION

One-twelfth run of Roche 454 pyrosequencing yielded 45,236 raw sequence reads, ranging from 355 to 986 bp with an average length of 573 bp. A total of 1568 SSRs were identified that contained enough flanking sequences for primer pair design. These sequences were deposited in the NCBI GenBank (GenBank accession Nos. KP248123-KP249690). One hundred SSRs were randomly selected to design primer pairs and PCR amplifications were performed. Because of the amplification of multiple bands or unsuccessful amplification of the target fragments, seventy-eight primer pairs were discarded. The remaining pairs were tested for polymorphisms using 18 individuals from three *E. ulmoides* populations; polymorphic amplification products were obtained in 16 individuals (Table 1). For polymorphic primer pairs, N_A per locus ranged from 3 to 8, with an average of 5.1 (Table 1). H_E and H_O ranged from 0.110 to 0.830 (with an average of 0.648) and from 0.111 to 0.833 (with an average of 0.524) at the species level, respectively (Table 1). F_{IS} varied from -0.349 to 0.547 (Table 1). Genetic diversity parameters in each population are presented in Table 2. Six SSR loci

(*DZ4*, *DZ7*, *DZ8*, *DZ10*, *DZ11*, and *DZ14*) showed significant deviation from the HWE ($P < 0.05$) due to heterozygote excess or deficiency. Four pairs of loci, namely, *DZ1* and *DZ4*, *DZ6* and *DZ8*, *DZ8* and *DZ14*, and *DZ6* and *DZ16*, showed significant LD, indicating significant allelic associations between these loci. This set of microsatellite markers could be valuable for landscape genetic structure assessment and molecular marker-assisted breeding in *E. ulmoides*.

Table 1. Primer sequences and characterization for sixteen microsatellite loci isolated from *Eucommia ulmoides*.

Primer	Primer sequence (5'-3')	Repeat motif	Ta (°C)	Allele size (bp)	N_A	H_E	H_O	F_{IS}	GenBank Accession No.
DZ1	F: TTACCAAAACGACATGCCAA R: TTGCTGTGGGAACTGTGTGT	$(AAC)_{11}$	58	271-283	5	0.732	0.611	0.169	KP248123
DZ2	F: AGCAAGAACCTGTTGCCATT R: CAGTCGATAAGTTCGCCCTC	$(AAC)_{13}$	58	316-345	8	0.830	0.833	-0.004	KP248129
DZ3	F: GCACAAAATTTAAGTGCGTAATC R: TTCGAGAGATCATTCGTGAGA	$(AAC)_{13}(AAT)_5$	56	316-328	7	0.816	0.722	0.118	KP248130
DZ4	F: TGGATTGGGGATAATTCAATG R: GAGCCAGGCTGTGATCTAGG	$(AAT)_5(AAC)_{21}$	58	267-282	7	0.786	0.444*	0.441	KP248144
DZ5	F: ACGAAACATCGATCACACCA R: GGGAGCGTTTCTGAAGAAGAT	$(AC)_{10}$	58	322-332	5	0.713	0.444	0.383	KP248146
DZ6	F: TTGATCCCTTGCCACAATTT R: GGAACCCGTTGGATTCTACA	$(AC)_{10}(AT)_7(AG)_8$	58	347-353	4	0.644	0.444	0.317	KP248149
DZ7	F: GCGTTACAAGATTTCATCGCT R: AGCACCCATCGAAGAGAAAA	$(ACA)_{12}$	58	370-379	4	0.729	0.500*	0.320	KP248167
DZ8	F: CGTACCGATTTCGTATCGTG R: AACATGCATGGGGTTATGGT	$(AG)_{11}$	58	363-371	5	0.711	0.333*	0.538	KP248170
DZ9	F: TTTGGAAGTTGGGAACAAAAA R: AATTGATGGTTTCGGTTTCG	$(AAACAC)_6$	58	135-153	4	0.513	0.444	0.137	KP248175
DZ10	F: GCGTGATTTGATTTGGCTCT R: TGATGGCTCATTTTGGCATA	$(AAC)_{12}$	58	207-216	5	0.724	0.333*	0.547	KP248179
DZ11	F: TCACTCCTAGACCACCCACC R: GGGTGTCGTTGTGTGTATGC	$(AC)_{13}$	58	270-278	5	0.700	0.556*	0.211	KP248183
DZ12	F: GGATATTCGTAGCCGTTGGA R: AAAACACTCGAATTGTGGGC	$(AC)_{19}(AT)_{10}$	58	220-228	5	0.479	0.556	-0.164	KP248189
DZ13	F: TGTACGCTAAATTTTCGGGC R: TCTTCTCTCGCTCTTCCTCG	$(AG)_{18}$	58	316-328	6	0.792	0.778	0.019	KP248201
DZ14	F: TCACACACATACACGCATGG R: TTTTGGGGTCCAAATTGTGT	$(CT)_{13}$	58	264-272	3	0.624	0.833*	-0.349	KP248213
DZ15	F: TACGCCTCTCATGTCCTTCA R: AACTGCCCCTGACATTCTTG	$(GA)_{15}$	60	261-273	5	0.470	0.444	0.056	KP248216
DZ16	F: TCAGAACATCAGCTCATCCAA R: AAGGCTCGATGCCAGATAGA	$(GA)_{15}$	60	304-310	3	0.110	0.111	-0.015	KP248217

Ta = PCR annealing temperature; N_A = number of alleles; H_E = expected heterozygosity; H_O = observed heterozygosity; F_{IS} = inbreeding coefficient; *Significant deviation from Hardy-Weinberg equilibrium.

Table 2. Genetic diversity parameters for three populations of *Eucommia ulmoides*.

Primer	Xinyang (N = 6)				Nanyang (N = 6)				Luoyang (N = 6)			
	N_A	H_E	H_O	F_{IS}	N_A	H_E	H_O	F_{IS}	N_A	H_E	H_O	F_{IS}
DZ1	5	0.818	0.833	-0.020	5	0.742	0.667	0.111	4	0.697	0.333	0.545
DZ2	6	0.818	1.000	-0.250	5	0.742	0.667	0.111	4	0.803	0.833	-0.042
DZ3	4	0.803	0.833	-0.042	4	0.697	0.667	0.048	6	0.803	0.667	0.184
DZ4	4	0.727	0.000*	1.000	5	0.849	0.667	0.231	4	0.712	0.667	0.070
DZ5	4	0.636	0.333	0.500	4	0.682	0.500	0.286	4	0.712	0.500	0.318
DZ6	3	0.530	0.667	-0.290	2	0.485	0.333	0.333	3	0.621	0.333	0.487
DZ7	3	0.682	0.833	-0.250	2	0.303	0.000*	1.000	4	0.561	0.667	-0.212
DZ8	4	0.742	0.833	-0.136	3	0.682	0.167*	0.773	2	0.546	0.000*	1.000
DZ9	4	0.636	0.833	-0.351	2	0.303	0.333	-0.111	3	0.621	0.167*	0.750
DZ10	5	0.788	0.333	0.600	3	0.621	0.667	-0.081	3	0.667	0.000*	1.000
DZ11	5	0.758	0.500	0.362	4	0.727	0.667	0.091	4	0.636	0.500	0.231
DZ12	2	0.485	0.667	-0.429	5	0.727	0.833	-0.163	2	0.167	0.167	0.000
DZ13	5	0.803	0.833	-0.042	3	0.667	0.667	0.000	5	0.789	0.833	-0.064
DZ14	2	0.530	0.500	0.063	3	0.667	1.000	-0.579	2	0.546	1.000*	-1.000
DZ15	3	0.439	0.333	0.259	3	0.439	0.500	-0.154	4	0.561	0.500	0.118
DZ16	2	0.167	0.167	0.000	2	0.167	0.167	0.000	1	0.000	0.000	-

N = number of individuals tested; N_A = number of alleles; $H_{E\ =}$ expected heterozygosity; H_O = observed heterozygosity; F_{IS} = inbreeding coefficient; *Significant deviation from Hardy-Weinberg equilibrium; - = not surveyed.

ACKNOWLEDGMENTS

Research supported by the Postdoctoral Science Foundation of China (#13A180057) and the National Natural Science Foundation of China (#31100272).

REFERENCES

Chen X, Sang X, Li S, Zhang S, et al. (2010). Studies on a chlorogenic acid-producing endophytic fungi isolated from *Eucommia ulmoides* Oliver. *J. Ind. Microbiol. Biotechnol.* 37: 447-454.http://dx.doi.org/10.1007/s10295-010-0690-0

Deng J, Liu Y and Huang H (2006). Development and characterization of microsatellite markers in *Eucommia ulmoides* Oliver (Eucommiaceae). *Mol. Ecol. Notes* 6: 496-498. http://dx.doi.org/10.1111/j.1471-8286.2006.01287.x

Fu ZZ, Li YH and Li Y (2014). Isolation and characterization of microsatellite markers for *Forsythia suspensa* (Thunb.) Vahl (Oleaceae) using 454 sequencing technology. *Biochem. Syst. Ecol.* 57: 328-331. http://dx.doi.org/10.1016/j.bse.2014.09.012

Goudet J (1995). FSTAT (version 1.2): A computer software to calculate F-statistics. *J. Hered.* 86: 485-486.

Hall LA and Beissinger SR (2014). A practical toolbox for design and analysis of landscape genetics studies. *Landscape Ecol.* 29: 1487-1504. http://dx.doi.org/10.1007/s10980-014-0082-3

Hussain M, Hussain MT, Rama NH, Hameed S, et al. (2003). Synthesis and antimicrobial activities of some isocoumarin and dihydroisocoumarin derivatives. *Nat. Prod. Res.* 17: 207-214.http://dx.doi.org/10.1080/1057563021000040826

Li Y, Zhang W, Fu ZZ, Wang W, et al. (2015). Isolation and characterization of microsatellite markers for *Achyranthesbidentata* (Amaranthaceae) using next-generation sequencing platform. *Biochem. Syst. Ecol.* 61: 437-440. http://dx.doi.org/10.1016/j.bse.2015.06.032

Li YC, Korol AB, Fahima T, Beiles A, et al. (2002). Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. *Mol. Ecol.* 11: 2453-2465.http://dx.doi.org/10.1046/j.1365-294X.2002.01643.x

Rousset F (2008). genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Mol. Ecol. Resour.* 8: 103-106.http://dx.doi.org/10.1111/j.1471-8286.2007.01931.x

Rozen S and Skaletsky HJ (2000). Primer3 on the WWW for general users and for biologist programmers. In: Bioinformatics methods and protocols: methods in molecular biology (Krawetz S and Misener S, eds.) Humana Press, Totowa, USA, 365-386.

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.). *Theor. Appl. Genet.* 106: 411-422.

Wang AQ, Huang LQ, Shao AJ, Cui GH, et al. (2006). [Genetic diversity of *Eucommia ulmoides* by RAPD analysis]. *Zhongguo Zhong Yao Za Zhi* 31: 1583-1586.

Wei N, Dick CW, Lowe AJ and Gardner MG (2013). Polymorphic microsatellite loci for *Virola sebifera* (Myristicaceae) derived from shotgun 454 pyrosequencing. *Appl. Plant Sci.* 1: 1200295.http://dx.doi.org/10.3732/apps.1200295

Yao XH, Deng JY and Huang HW (2012). Genetic diversity in *Eucommia ulmoides* (Eucommiaceae), an endangered traditional Chinese medicinal plant. *Conserv. Genet.* 13: 1499-1507. http://dx.doi.org/10.1007/s10592-012-0393-3

Yeh FC, Yang R, Boyle TJ, Ye Z, et al. (2000). PopGene 32. Microsoft Window-based freeware for population genetic analysis, version 1.32. Molecular Biology and Biotechnology Centre, University of Alberta, Edmonton, Canada.