Genetics and Molecular Research

Development and characterization of novel microsatellite markers for *Ginkgo biloba* using 454 pyrosequencing

Q. Zhou[1]*, X.X. Wang[1]*, M. Xu[1], F.L. Cao[1], F.X. Yu[2] and L.A. Xu[1]

[1]Co-Innovation Center for Sustainable Forestry in Southern China, Nanjing Forestry University, Nanjing, China
[2]Jiangxi Provincial Key Laboratory of Ornamental Plant Genetic Improvement, Research Institute for Bioresources, Jiangxi Academy of Sciences, Nanchang, China

*These authors contributed equally to this study.
Corresponding author: L.A. Xu
E-mail: laxu@njfu.edu.cn

Genet. Mol. Res. 15 (1): gmr.15017385
Received August 6, 2015
Accepted November 13, 2015
Published March 28, 2016
DOI http://dx.doi.org/10.4238/gmr.15017385

ABSTRACT. As a “living fossil” that is used to understand the evolutionary history of seed plants, *Ginkgo biloba* is a well-known multipurpose tree with edible seeds, medicinal properties, and ornamental value, but little is known about its genetic diversity. Microsatellite, or simple sequence repeat (SSR), markers have proven to be powerful tools for genetic studies of plants. In this study, we isolated 30 novel polymorphic microsatellite loci in *G. biloba* using 454 pyrosequencing. The characteristics of these loci were tested with 48 cultivars. The number of alleles (N_A) per locus ranged from two to seven. The observed (H_O) and expected (H_E) heterozygosities ranged from 0.000 to 0.750 and from 0.021 to 0.792, with an average of 0.326 and 0.443, respectively. In terms of genetic diversity in the *Ginkgo* population, N_A was 3.300, N_E was 2.090, I was 0.782, H_O was 0.326, and H_E

was 0.443. These polymorphic SSRs will be useful for the assessment of population genetic diversity and resource conservation of *G. biloba*.

Key words: *Ginkgo biloba*; 454 Pyrosequencing; Microsatellites; Genetic diversity

INTRODUCTION

Ginkgo biloba, also known as the maidenhair tree, is one of the oldest living tree species on the planet. This tree is considered to be a “living fossil”, meaning that it has continued to survive even after major extinction events. Thus, understanding the origin and phylogeography of *G. biloba* will help to elucidate the evolutionary history of seed plants, and the climate and geological changes that have occurred over time in the Northern Hemisphere. Its natural habitat is restricted to small areas of China (Shen et al., 2005; Gong et al., 2008; Tang et al., 2012), but this species is now widely cultivated around the world due to its edible seeds, medicinal properties, and ornamental value. Despite considerable field surveys and extensive fossil records for the genus *Ginkgo* (Zhou and Zheng, 2003; Zhou, 2009; Zhou et al., 2012), little is known about population genetics, molecular ecology, and genetic resources of this species. Microsatellites, or simple sequence repeats (SSRs), consist of tandem arrays of short nucleotide motifs, which are randomly dispersed throughout eukaryotic genomes. SSR markers have distinguishing features such as reproducibility, abundant polymorphism, and co-dominant inheritance. These make SSRs powerful high-resolution tools for the study of population genetics, molecular ecology, and marker-assisted selection (MAS) in plants.

Although a few SSR markers are available, genomic resources for *G. biloba* remain limited (Yan et al., 2006; Li et al., 2009; Xu et al., 2015). Owing to the urgent need for germplasm conservation in this species, and its important role in phylogenetic studies, the generation of transcriptome data will contribute to the conservation of these trees and to genetic research in gymnosperms. In this study, 30 novel microsatellite markers were isolated and characterized from *G. biloba* based on 454 pyrosequencing.

MATERIAL AND METHODS

Sample collection and DNA extraction

Young leaves of 48 *Ginkgo* cultivars from the major planting areas of the species were collected from the Pizhou Ginkgo Germplasm Garden in Jiangsu Province. These samples were individually ground to powder in liquid nitrogen, and DNA extraction was performed using a DNeasy Plant Mini Kit (Qiagen).

SSR development and genotyping

Total RNA was extracted from the female cones and leaves of a single individual using an RNeasy Plant Mini Kit (Qiagen). Methods for cDNA library construction and 454 pyrosequencing have been previously described (Chen et al., 2011). Subsequently, SSRs were detected using the MIcroSAtellite (MISA) program search module. The parameters were set to detect perfect di-, tri-, tetra-, penta-, and hexanucleotide motifs, with a minimum of nine, six, five, five, and four repeats,

respectively. SSR primer pairs were designed using the Primer Premier5.0 software.

PCRs were performed under the following conditions: an initial denaturation at 94°C for 5 min, followed by 30 cycles of 30 s at 94°C, 30 s at the locus-specific annealing temperature, and 40 s at 72°C, followed by a final extension of 1 min at 72°C. A typical 10-μL reaction contained: 1X buffer, 2.5 mM $MgCl_2$, 0.2 mM each dNTPs, 0.25 μM each primer, 0.25 U Taq DNA polymerase (Takara), and 25 ng genomic DNA. Polymorphic PCR products were analyzed on 8% silver-stained polyacrylamide gels.

Data analysis

Genetic diversity indices, including the number of alleles at each locus (N_A), the number of effective alleles per locus (N_E), Shannon's information index (I), observed heterozygosity (H_O), and expected heterozygosity (H_E) were estimated using GenAlEx version 6.5 (Peakall and Smouse, 2012).

RESULTS AND DISCUSSION

Two cDNA libraries were constructed from female cones and leaves of a single individual, and sequenced on the Roche 454 GS-FLX platform. In total, 251,636 raw reads with an average length of 345 bp were generated from female cones, and 223,261 raw reads with an average length of 346 bp were generated from leaves. After adaptor removal and assembly by Newbler version 2.3 (Roche Applied Science, USA) using default parameters, 19,128 contigs and 24,940 singletons were generated from female cones, and 14,671 contigs and 15,108 singletons were generated from leaves.

The MISA program was used for SSR detection. Among 44,068 sequences derived from female clones, a total of 605 perfect SSRs were identified, including 274 dinucleotide repeats, 206 trinucleotide repeats, 42 tetranucleotide repeats, 27 pentanucleotide repeats, and 56 hexanucleotide repeats. Of 19,779 sequences from leaves, a total of 365 perfect SSRs were identified, including 179 dinucleotide repeats, 116 trinucleotide repeats, 24 tetranucleotide repeats, 6 pentanucleotide repeats, and 40 hexanucleotide repeats (Table 1).

Table 1. Frequency, type, and distribution of simple sequence repeats (SSRs) in *Ginkgo biloba*.

	Clones	Leaves
No. of sequences examined	44,068	19,779
No. of SSR-containing sequences	562	336
Total number of perfect SSRs	605	365
No. of dinucleotide motifs	274	179
No. of trinucleotide motifs	206	116
No. of tetranucleotide motifs	42	24
No. of pentanucleotide motifs	27	6
No. of hexanucleotide motifs	56	40
No. of compound SSRs	23	22

A total of 556 non-redundant SSR-containing sequences were utilized for primer design. Of 556 SSRs, 176 primer pairs produced polymorphic products in six *Ginkgo* individuals. The characteristics of 30 SSR loci were then investigated using *Ginkgo* cultivars. Among the 30 SSR loci

amplified, 15 were dinucleotide repeats; seven were trinucleotide repeats; two were tetracleotide repeats; and six were hexanucleotide repeats (Table 2). In total, 99 alleles were detected in all 48 *G. biloba* individuals. The N_A per locus ranged from two to seven. The H_O and H_E ranged from 0.000 to 0.750 and from 0.021 to 0.792, with an average of 0.326 and 0.443, respectively (Table 1).

Table 2. Characterization of 30 polymorphic microsatellite markers for *Ginkgo biloba*, including locus name, repeat motif, primer sequences, expected size of alleles (*S*), number of alleles (N_A), number of effective alleles (N_E), observed heterozygosity (H_O), and expected heterozygosity (H_E).

Locus	Repeat motif	*S* (bp)	Forward primer (5' to 3')	Reverse primer (5' to 3')	N_A	N_E	H_O	H_E
GbeSSR16	$(AG)_{14}$	236	GATCCATTTTCTGGTTCT	GTGTGATTTCTTCCTTGT	4	2.3	0.313	0.558
GbeSSR25	$(AT)_{11}$	288	CATTACAGCGACTGAAACA	AGAGTGGGCTTAGCTTGAT	4	2.6	0.396	0.612
GbeSSR32	$(TA)_{11}$	196	TTCGCTGTAGCATTTGTG	GCAGGTTGTATTTCGGAG	3	2.3	0.458	0.571
GbeSSR78	$(CTT)_7$	314	CTGAAGACGGAAACCACCT	GCCGAAAACAAGAACAATG	3	1.7	0.021	0.401
GbeSSR79	$(AT)_{10}$	192	GGCAATCAGAATACCTATC	TAAAGCCTACATCACATCC	2	1.2	0.208	0.187
GbeSSR91	$(AAC)_8$	297	ACCCTCCCAGAAAAAGTC	AGGTTGGCAATGTTAGCA	5	3.1	0.458	0.675
GbeSSR94	$(TTTC)_6$	201	AGTCCCATTGACCTTTTG	GATCGGCATATTTACTATTC	2	1.1	0.125	0.117
GbeSSR120	$(CA)_{11}$	234	AAGTCATAAGCGACAGTG	CCGTCTTTCAGATCAATA	7	4.8	0.750	0.792
GbeSSR137	$(AGC)_6$	384	ATTCTCCCCACTCTCATC	CTGTAACTTCTGCACTAGC	2	1.1	0.083	0.080
GbeSSR150	$(TC)_9$	160	ACATAGTGAGAGTCAGCAT	AGAGATACAATACAGAAAGG	3	2.4	0.271	0.587
GbeSSR156	$(GA)_{18}$	281	CTGTAACTAACTAATGAGA	GGTAGTGTATTTTTGAAC	2	1.0	0.042	0.041
GbeSSR202	$(AG)_{14}$	267	CCCTTGTTTCTCCATAAT	TGCTCATATAGGTGCTCT	5	1.8	0.333	0.430
GbeSSR215	$(AAGCAG)_5$	157	AGGAACTGATGATGACGAT	CACTTTCACTTTAATAACGG	4	2.1	0.292	0.534
GbeSSR237	$(TTG)_7$	252	ATCTTCAACCCCCTCAACT	CACGATCACATGCAAATATAC	4	2.5	0.354	0.597
GbeSSR245	$(TCCCAC)_6$	320	TCCGAGACCGTCGCAATA	TGGTTCCCTCCCAATGAT	3	1.2	0.000	0.192
GbeSSR367	$(CATGGA)_5$	170	ACTGGGTGGAATACTGAT	CAACTACAAAGTGGAAGA	3	1.9	0.292	0.467
GbeSSR383	$(AT)_{14}$	299	CCGATGTAAGCAGGTCAG	GCATTCGTGTCATTGTTG	4	2.6	0.458	0.615
GbeSSR400	$(GTAT)_5$	244	GTCTTATATGTGCTTCAGC	CAACTCGAACTATATTACCT	2	1.9	0.313	0.482
GbeSSR430	$(TAGGAT)_5$	325	TATCGTGTCGTGGAACCG	CCGAATACAAAGCAGCCT	2	1.5	0.396	0.342
GbeSSR432	$(AGGCGG)_4$	287	GGAGACAAATAGCGGTAA	CACAAGCGTCATCATTCT	2	1.9	0.375	0.486
GbeSSR463	$(TA)_{15}$	350	CAGCAATGGAGACTTCTT	GCCAATACTCTTTAACGG	4	2.8	0.417	0.644
GbeSSR495	$(AT)_{15}$	357	GAGAAACATCAAGGAGAGT	ATAATAAGGGCATTGTGAG	4	3.0	0.646	0.663
GbeSSR496	$(TTA)_7$	317	GATTGTGAAAAGAGAAGG	ATTGTAGATTGCTCCAAC	5	3.5	0.750	0.717
GbeSSR502	$(AT)_{10}$	188	GAATAGAAGAGATGTGCG	CATATGTTAGTTTGTGGG	3	2.1	0.688	0.523
GbeSSR503	$(TC)_{14}$	229	AACTTATTAGCAATCCTCG	CACACATTTACTGAACCTAT	2	1.3	0.188	0.234
GbeSSR511	$(TTC)_6$	424	TCCCATTTCACCAGTCTC	CTCGTCCTCTCCATCACT	3	2.4	0.438	0.578
GbeSSR536	$(GCA)_6$	176	AAACAACCACAACCGCCA	CCTTCGCTCCATTCGTCTC	2	1.2	0.063	0.135
GbeSSR538	$(CT)_{11}$	103	AGAGATTTTGCGACAGAGC	GGTAGCAGTTGAACCGTTA	5	2.7	0.500	0.627
GbeSSR549	$(AATGGT)_4$	274	ATGGCTGCCCTCAACTTG	TGGACTGCTTGGCCTTAG	2	1.0	0.021	0.021
GbeSSR550	$(TG)_{12}$	154	GTTACAGGCAGATTATACC	ATGGCACTAACCACACAG	3	1.6	0.125	0.389

The assessment of population genetic diversity performed using these 30 novel SSR markers will be helpful for the effective management and sustainable utilization of *G. biloba* resources. In terms of genetic diversity in the 48 *Ginkgo* cultivars, N_A was 3.300, N_E was 2.090, *I* was 0.782, H_O was 0.326, and H_E was 0.443.

The once diverse *Ginkgo* group is today represented by a single species, *G. biloba*, which is now widely cultivated around the world. The novel polymorphic SSR markers developed in this study will be available for studies of conservation genetics for *G. biloba*.

Conflicts of interest

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

Research supported by the Special Fund for Forest Scientific Research in the Public Welfare (#201204403-3), the Opening Fund for Jiangxi Provincial Key Laboratory of Ornamental Plant Genetic Improvement (#2013-KLB-01), and the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD).

REFERENCES

Chen Y, Jiang XM, Zhang L, Zhang XY, et al. (2011). Transcriptome characterization of *Camellia* Sect. Oleifera based on the 592 499 ESTs. *Sci. Silvae Sin.* 47: 161-163.

Gong W, Zeng Z, Chen YY, Chen C, et al. (2008). Glacial refugia of *Ginkgo biloba* and human impact on its genetic diversity: evidence from chloroplast DNA. *J. Integr. Plant Biol.* 50: 368-374.http://dx.doi.org/10.1111/j.1744-7909.2007.00375.x

Li YY, Zang LP and Chen XY (2009). Development of polymorphic microsatellite markers for *Ginkgo biloba* L. by database mining. *Conserv. Genet. Resour.* 1: 81-83. http://dx.doi.org/10.1007/s12686-009-9019-6

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.http://dx.doi.org/10.1093/bioinformatics/bts460

Shen L, Chen XY, Zhang X, Li YY, et al. (2005). Genetic variation of *Ginkgo biloba* L. (Ginkgoaceae) based on cpDNA PCR-RFLPs: inference of glacial refugia. *Heredity (Edinb)* 94: 396-401.http://dx.doi.org/10.1038/sj.hdy.6800616

Tang CQ, Yang Y, Ohsawa M, Yi SR, et al. (2012). Evidence for the persistence of wild *Ginkgo biloba* (Ginkgoaceae) populations in the Dalou Mountains, southwestern China. *Am. J. Bot.* 99: 1408-1414.http://dx.doi.org/10.3732/ajb.1200168

Xu M, Xu LA, Cao FL, Zhang HJ, et al. (2015). Development of novel chloroplast microsatellite markers for *Ginkgo biloba*. *Genet. Mol. Res.* 14: 7715-7720.http://dx.doi.org/10.4238/2015.July.13.17

Yan XF, Lian CL and Hogetsu T (2006). Development of microsatellite markers in ginkgo (*Ginkgo biloba* L.). *Mol. Ecol. Notes* 6: 301-302. http://dx.doi.org/10.1111/j.1471-8286.2006.01134.x

Zhou Z and Zheng S (2003). The missing link in Ginkgo evolution. *Nature* 423: 821-822.http://dx.doi.org/10.1038/423821a

Zhou ZY (2009). An overview of fossil Ginkgoales. *Palaeoworld* 18: 1-22. http://dx.doi.org/10.1016/j.palwor.2009.01.001

Zhou ZY, Quan C and Liu YS (2012). Tertiary Ginkgo ovulate organs with associated leaves from North Dakota, U.S.A., and their evolutionary significance. *Int. J. Plant Sci.* 173: 67-80. http://dx.doi.org/10.1086/662651