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

Development of novel chloroplast microsatellite markers for *Ginkgo biloba*

M. Xu[1], L.A. Xu[1], F.L. Cao[1], H.J. Zhang[1] and F.X. Yu[2]

[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

Corresponding author: L.A. Xu
E-mail: laxu@njfu.edu.cn

Genet. Mol. Res. 14 (3): 7715-7720 (2015)
Received October 15, 2014
Accepted February 27, 2015
Published July 13, 2015
DOI http://dx.doi.org/10.4238/2015.July.13.17

ABSTRACT. *Ginkgo biloba* is considered to be a living fossil that can be used to understand the ancient evolutionary history of gymnosperms, but little attention has been given to the study of its population genetics, molecular phylogeography, and genetic resources assessment. Chloroplast simple sequence repeat (cpSSR) markers are powerful tools for genetic studies of plants. In this study, a total of 30 perfect cpSSRs of *Ginkgo* were identified and characterized, including di-, tri, tetra-, penta-, and hexanucleotide repeats. Fifteen of 21 designed primer pairs were successfully amplified to yield specific polymerase chain reaction products from 16 *Ginkgo* cultivars. Polymorphic cpSSRs were further applied to determine the genetic variation of 116 individuals in 5 populations of *G. biloba*. The results showed that 24 and 76% genetic variation existed within and among populations of this species, respectively. These polymorphic and monomorphic cpSSR markers can be used to trace the origin and evolutionary history of *Ginkgo*.

Key words: Chloroplast genome; Genetic diversity; *Ginkgo biloba*; Microsatellites

INTRODUCTION

Microsatellites, or simple sequence repeats (SSRs), consist of tandem arrays of short nucleotide motifs and are widely distributed throughout the nuclear and cytoplasmic genome of eukaryotes. Compared with the nuclear genome, the chloroplast genome is typically non-recombinant, uniparentally inherited, and has a slower evolutionary rate; therefore, it is more likely to retain ancient genetic patterns and to provide unique insight into fundamental evolutionary processes, such as genetic drift, population subdivision, and differentiation (Ebert and Peakall, 2009). Chloroplast SSR (cpSSR) markers are powerful tools that can be used in genetic studies of plants.

The well-known relic *Ginkgo biloba* is one of the oldest living seed plants, and its natural habitat is restricted to small areas of China (Shen et al., 2005; Gong et al., 2008; Tang et al., 2012). Because of its edible seeds, medicinal efficacy, and ornamental value, this species is now widely cultivated worldwide. Despite the considerable field surveys and extensive fossil records for the species, little attention has been given to the study of its population genetics, molecular phylogeography, and genetic resources assessment.

The chloroplast genome sequence of *G. biloba* has recently been assembled, annotated, and published (Lin et al., 2012), and is a valuable and unique resource for cpSSR marker development. Although few nuclear SSR markers are available for this species (Yan et al., 2006; Li et al., 2009), these loci are insufficient to capture all genetic variation within the population. The objectives of this study were to analyze the distribution of perfect SSRs in the chloroplast genome of *G. biloba*, isolate a set of cpSSR markers, and test their applicability in population genetics.

MATERIAL AND METHODS

Sample collection and DNA extraction

Leaf materials from 16 *Ginkgo* cultivars from the major planting areas of the species were collected from the Pizhou Ginkgo Germplasm Garden in Jiangsu Province. Materials from 116 wild or semiwild *Ginkgo* trees were collected from 5 different localities, including Wuchuan, Fenggang, Panxian, Duyun (Guizhou Province), and Tianmu Mountain (Zhejiang Province). Immediately after collection, the leaf materials for each individual were dried on silica gel in Ziplock plastic bags until DNA extraction. The leaf samples were individually ground into powder in liquid nitrogen and DNA extraction was performed using the DNeasy Plant Mini Kit (Qiagen, Hilden, Germany) according to manufacturer instructions.

SSR development

The complete chloroplast genome sequence of *G. biloba* (GenBank Accession No.: AB684440) was downloaded from GenBank (http://www.ncbi.nlm.nih.gov/) and saved in FASTA format. Perl scripts were developed to search for SSRs in the chloroplast genome of *G. biloba* using the MIcroSAtellite search module (http://pgrc.ipk-gatersleben.de/misa/). The parameters were set to detect perfect di-, tri-, tetra-, penta-, and hexanucleotide motifs with a minimum of 5, 4, 3, 3, and 3 repeats, respectively. Repeat motifs, number of repeats, and start and end positions of the SSRs were extracted from the MIcroSAtellite output. Primers were

designed using online software (http://www.genscript.com/cgi-bin/tools/primer_genscript.cgi?op=advanced). The parameters for designing the primers were set as follows: primer length from 18-22 base pairs, with 20 as the optimum value. Polymerase chain reaction (PCR) product size ranged from 100 to 500 base pairs; the optimum annealing temperature was 58°C.

PCR amplification and genotyping

PCRs were performed in a 20-μL reaction mixture containing 30-50 ng genomic DNA, 1X PCR buffer (Takara, Shiga, Japan), 0.2 mM dNTP mix (Takara), 1.7 mM $MgCl_2$ (Takara), 0.5 U *Taq* polymerase (Takara), and 0.5 μM of each primer. The amplification reaction conditions were as follows: 3 min denaturation at 94°C, followed by 30 cycles of 30 s at 94°C, 30 s at 56°C, and 30 s at 72°C, with a final extension at 72°C for 3 min. Primer pairs were initially tested for amplification success using genomic DNA samples of the 16 *Ginkgo* cultivars, and the PCR products were resolved on a 6% polyacrylamide denaturing gel and visualized by silver staining. The forward primers of the selected primer pairs were labeled with 6-carboxyfluorescein fluorescent dye. The characteristics of these labeled markers were further tested with 116 individuals from 5 populations of *G. biloba.* PCR amplification followed the conditions described above. Polymorphic PCR products were detected by capillary electrophoresis using the ABI 3730 DNA Analyzer with the GeneScanTM600 LIZ size standard (Applied Biosystems, Foster City, CA, USA) and the sizes were determined using the GeneMapper software version 4.0.

Data analysis

The number of different alleles (N_A), number of effective alleles (N_E), Shannon's Information index (*I*), and unbiased haploid diversity index (*uh*) per polymorphic locus within the different populations were estimated using GenAIEx version 6.5 (Peakall and Smouse, 2006, 2012).

RESULTS AND DISCUSSION

To identify cpSSRs, di-, tri, tetra-, penta-, and hexanucleotide motifs were identified from the 156,945-bp chloroplast genome sequence using MIcroSAtellite. A total of 30 perfect SSRs were identified, including 16 dinucleotide repeats, 3 trinucleotide repeats, 7 tetranucleotide repeats, 1 pentanucleotide repeat, and 3 hexanucleotide repeats (Table 1). An average of 1 SSR locus occurred every 5.23 kb of the chloroplast genome. Finally, 21 locus-specific primer pairs were designed from the sequences flanking the SSR loci and were tested on 16 *Ginkgo* cultivars. Fifteen of these 21 primer pairs were successfully amplified to yield specific PCR products with a single band, of which only the *GbcpSSR1* and *GbcpSSR4* loci were polymorphic; the other 6 primer pairs were amplified to produce nonspecific PCR products with 2 bands visualized by silver staining (Table 2).

For the polymorphic *GbcpSSR1* and *GbcpSSR4* loci, the N_A, N_E, *I*, and *uh* within populations were further estimated among 116 individuals from 5 populations of *G. biloba* (Table 3). Similar levels of genetic variation were revealed by the *GbcpSSR1* and *GbcpSSR4* loci. N_A, N_E, *I*, and *uh* at the *GbcpSSR1* locus were 2.000, 1.890, 0.664, and 0.475, respectively, while the values of these variables at the *GbcpSSR4* locus were 2.000, 1.842, 0.650, and 0.461, respectively. Allele frequencies by population and locus are shown in Figure 1.

Table 1. Frequency, type, and distribution of SSRs in the *Ginkgo biloba* chloroplast genome.

Repeats	3	4	5	6	7	8	9	Total
AT			2			1	1	4
CT						1		1
TA			1	1	5	3		10
TC			1					1
TTA		1						1
TTC		2						2
AAAC	1							1
AGGT	1							1
CTAC	1							1
GATA		1						1
TAGA	1							1
TATC	1	1						2
AATGA	1							1
GAATAA	1							1
GGAGTG			1				1	
TCTTAT	1							1
Total	8	5	4	2	5	5	1	30

Table 2. Isolation and characterization of cpSSR markers for *Ginkgo biloba*.

Locus	Repeat motif	Ta (°C)	S (bp)	Forward primer (5' to 3')	Reverse primer (5' to 3')
GbcpSSR1^{P}	(GGAGTG)6	58	291	TCGAAATCCTTCCGAGCTAA	GACCGTCGCAATAAATCGTT
GbcpSSR2^{M}	(AAAC)3	58	284	TCCTGGACGTAATCCTGGAC	CACTCAAACTGGGTTTTCCAA
GbcpSSR3^{M}	(TTC)4	58	362	CCGACCAAAAAGGATTTTCA	GGCTGAACTGCTCGAAGTCT
GbcpSSR4^{P}	(TTA)4	58	381	ATCACCGAATGAGGAAAACG	CAGGTTCACGAATGGGAGTT
GbcpSSR5^{M}	(AATGA)3	58	142	CACGCGAAATGGAGATATGA	CTAATGTCAACCGCGAAGGA
GbcpSSR6^{N}	(TA)7	58	418	CAACAGCTACGTGGGTTGAG	GCCTTACCATGGCGTTACTC
GbcpSSR7^{M}	(TA)8	58	465	GTTCCGTCGAATTCATCGTT	CATTTCCGTTCGGTACGAGT
GbcpSSR8^{M}	(TA)6	58	339	TTGTATTTTCGGGTGCATCA	CATACCCGCATTTTCTTCGT
GbcpSSR9^{M}	(TC)5	58	489	TCCGGACAGACCAAAATCTC	GCCAGTAGCTCTTTCCATCG
GbcpSSR10^{M}	(AT)9	58	346	ACAACGAATTCTTGCCGAAG	AATGCTACGCCTTGAACCAC
GbcpSSR11^{N}	(TA)7...(TA)7	58	393	GTGCCTTTAGCGGGCTTATT	TATCGGGTCAGTGTGATGGA
GbcpSSR12^{N}	(CT)8	58	284	ATCCGACAACGATTGAGTGG	TGGAAATTGATCGAGTCATCC
GbcpSSR13^{N}	(AT)8	58	343	CCACTTGTCCCCGTACATCT	AAAATCCGTGCTTCTGTTCG
GbcpSSR14^{M}	(TTC)4	58	240	CCACAGATGAAATCTGGGATG	TCAATCGATGACTGGGAAGA
GbcpSSR15^{N}	(TATC)4A(TA)7	58	159	TCCCTCCAAACCGTACAGAC	AAAGGAAGGAAACGGATGCT
GbcpSSR16^{M}	(AGGT)3	58	266	AGCCAATGTCCGAGTACCAG	CGGAGACCTGTGTTTTTGGT
GbcpSSR17^{M}	(TCTTAT)3	58	242	CTGCGGAAAAATAGCTCGAC	TGGATCTGTTCCCAGGAGTC
GbcpSSR18^{N}	(TA)7(TAGA)3	58	252	AGAGGAGGAATGGACGGTTT	CACTACCCCCTTCCTTCGTT
GbcpSSR19^{M}	(AT)5	58	414	GGCTCATCGAAACGAAAATG	TGCAAGAGGTGTGAATCTGC
GbcpSSR20^{M}	(AT)5	58	175	TGGATCGAATCACATTCTTCA	TCGTCCCGGATCGATTGTA
GbcpSSR21^{M}	(TA)5	58	351	CCAATGTATCTCGCAGCTTG	CAAGATGGACCTTGCCTAGAA

Ta, annealing temperature (°C); S, expected size (bp); ^{M}monomorphic locus with specific amplification; ^{P}polymorphic locus with specific amplification; ^{N}nonspecific amplification locus.

In terms of genetic diversity in the *G. biloba* populations, N_A was 1.700, N_E was 1.153, I was 0.216, and *uh* was 0.126 at the species level. At the population level, N_A ranged from 1.500 to 2.000, N_E ranged from 1.059 to 1.385, I ranged from 0.107 to 0.451 with an average value of 0.241, and *uh* ranged from 0.056 to 0.290 with an average value of 0.167. Of the 5 populations tested, the Duyun population showed the highest levels of genetic diversity (N_A = 2.000, N_E = 1.385, I = 0.451, *uh* = 0.290), exceeding the levels of the Wuchuan population (N_A = 2.000, N_E = 1.137, I = 0.237, *uh* = 0.123) and Tianmu Mountain (N_A = 1.500, N_E = 1.123, I = 0.174, *uh* = 0.105), while the Panxian population showed the lowest levels of genetic variability (N_A = 1.500, N_E = 1.059, I = 0.107, *uh* = 0.056). Analysis of molecular variance showed that 24 and 76% genetic variation existed within and among the populations of *G. biloba*, respectively.

Table 3. Characteristics of the polymorphic cpSSR loci in 5 populations of *Ginkgo biloba*.

Locus PopID	GbcpSSR1				GbcpSSR4			
	N_A	N_E	*I*	*uh*	N_A	N_E	*I*	*uh*
DY (N = 24)	2	1.385	0.451	0.290	2	1.385	0.451	0.290
FG (N = 17)	1	1.000	0.000	0.000	2	1.125	0.224	0.118
PX (N = 18)	2	1.117	0.215	0.111	1	1.000	0.215	0.111
TM (N = 18)	2	1.246	0.349	0.209	1	1.000	0.000	0.000
WC (N = 19)	2	1.108	0.202	0.100	2	1.166	0.271	0.146
(N = 116)	2	1.890	0.664	0.475	2	1.842	0.650	0.461

N_A, number of different alleles per locus; N_E, number of effective alleles per locus; *I*, Shannon's Information index; *uh*, unbiased haploid diversity; N, population size; PopID, population identifier.

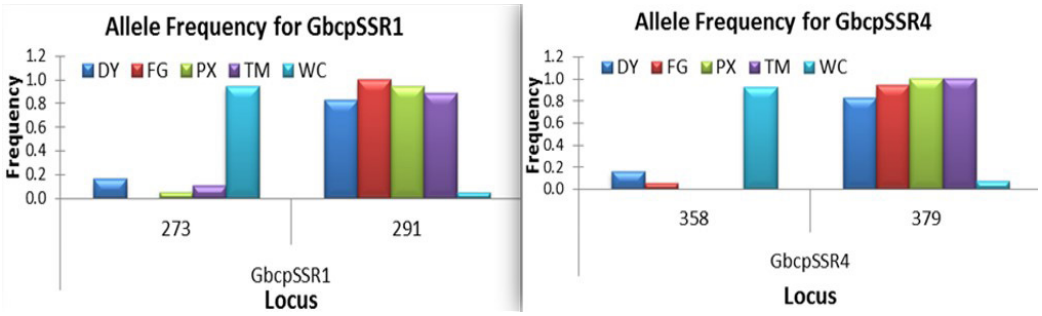

Figure 1. Allele frequencies with graphs by population and locus for the haploid data.

The *Ginkgo* genus represents an ancient evolutionary lineage within the gymnosperms (Wu et al., 2013), as demonstrated by extensive fossil findings from the Jurassic to the Tertiary periods in both the Northern and the Southern hemispheres (Zhou and Zheng, 2003; Zhou, 2009; Zhou et al., 2012). The novel polymorphic and monomorphic cpSSR markers developed in this study are useful for genetic studies of *G. biloba*, but they may also be important for determining the origin and evolutionary history of *Ginkgo*.

ACKNOWLEDGMENTS

Research supported by the Special Fund for Forest Scientific Research in the Public Welfare (#201204403-3), the Jiangsu Provincial Platform for Conservation and Utilization of Agricultural Germplasm, the Opening Fund for Jiangxi Provincial Key Laboratory of Ornamental Plant Genetic Improvement (#2013-KLB-01), the China Postdoctoral Science Foundation Funded Project (#2013M541682), the Jiangsu Planned Projects for Postdoctoral Research Funds (#1302023B), the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD), and the Collaborative Innovation Plan of Jiangsu Higher Education (CIP).

REFERENCES

Ebert D and Peakall R (2009). Chloroplast simple sequence repeats (cpSSRs): technical resources and recommendations for expanding cpSSR discovery and applications to a wide array of plant species. *Mol. Ecol. Resour.* 9: 673-690.

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.

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.

Lin CP, Wu CS, Huang YY and Chaw SM (2012). The complete chloroplast genome of *Ginkgo biloba* reveals the mechanism of inverted repeat contraction. *Genome Biol. Evol*. 4: 374-381.

Peakall R and Smouse PE (2006). GENALEX6: genetic analysis in Excel. Population genetic software for teaching and research. *Mol. Ecol*. 6: 288-295.

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

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* 94: 396-401.

Tang CQ, Yang YC, 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.

Wu CS, Chaw SM and Huang YY (2013). Chloroplast phylogenomics indicates that *Ginkgo biloba* is sister to Cycads. *Genome Biol. Evol*. 5: 243-254.

Yan XF, Lian CL and Hogetsu T (2006). Development of microsatellite markers in ginkgo (*Ginkgo biloba* L.). *Mol. Ecol*. 6: 301-302.

Zhou ZY (2009). An overview of fossil Ginkgoales. *Palaeoworld* 18: 1-22.

Zhou ZY and Zheng SL (2003). The missing link in *Ginkgo* evolution. *Nature* 423; 821-822.

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.